Statistical methods for the analysis of tooth shape

David L. Robinson

VOL. 2 (of 2)

Submitted for the degree of Doctor of Philosophy

November 2005

Dept. of Probability and Statistics, School of Mathematics and Statistics
& Dept. of Oral Health and Development, School of Clinical Dentistry

University of Sheffield

Chapter 7

Implementation of the new semi-landmark methods

7.1 Introduction

Having identified the most promising modifications of the semi-landmark technique from those proposed in the previous chapter (in terms of filtering out unwanted patterns of variation along pre-specified chord directions, while restricting the possibility of large unrealistic changes in shape being produced), the remainder of this thesis consists of a detailed assessment and comparison of these techniques. In chapter 8, we utilise a simulation study to evaluate the success of the new methods (compared with the original), in addressing the problem of unwanted variation due to the position of a patient's gum on buccal tooth surfaces. In chapters 9 and 10 we consider their use in other applications; for addressing other reliability problems associated with the analysis of tooth shape and in removing different patterns of simulated variation on simple geometric shapes.

Firstly however, we consider the implementation issues arising with the use of the new semi-landmark methods, as we did for Bookstein's original method in section 5.5. In section 7.2 we consider the importance of the GPA method used and how differences in the sizes of the Procrustes mean and fits, produced by the different GPA methods, affect the performance of the nearest point criterion. Convergence assessment is discussed in sections 7.3 and 7.4 and in 7.5 we describe the purposely written S-plus routine used for investigating and performing the analyses in the following chapters. Details are given of the various arguments for specifying how the new positions of semi-landmarks are determined, how the GPA steps are performed and for displaying graphically the initial, final and individual steps of the different methods.

7.2 Convergence assessment

As in section 5.5, let $X_{i(r)}^{new}$, i=1,…,n, denote the new configurations produced on iteration r after allowing the semi-landmarks of each Procrustes fit $X_{i(r-1)}^{P}$, to move to new positions with respect to mean shape $\hat{\mu}_{(r-1)}$, where each set of semi-landmarks move distances given by vector $\lambda_{i(r)}$ along unit chord directions defined by $U_{i(r)}$. In addition, let $X_{i(r)}^{P}$, i=1,…,n, and $\hat{\mu}_{(r)}$ denote the new Procrustes fits and corresponding mean shape of the $X_{i(r)}^{new}$, obtained by GPA at the end of the rth iteration.

Again, the semi-landmark process should be stopped at the end of the rth complete iteration, i.e. following GPA, to ensure that the configurations are registered to their mean shape, in preparation for subsequent analysis. As in section 5.5, for a pre-specified tolerance level ε, there are several possibilities by which we could assess convergence (that the landmarks have stopped moving), as given by equations (5.65) to (5.67), which in practice will essentially give identical results. Here we again use:

$$\sum_{i=1}^{n} d_F^2(X_{i(r)}^P, \hat{\mu}_{(r)}) - \sum_{i=1}^{n} d_F^2(X_{i(r-1)}^P, \hat{\mu}_{(r-1)}) = \sum_{i=1}^{n} \left\| X_{i(r)}^P - \hat{\mu}_{(r)} \right\|^2 - \sum_{i=1}^{n} \left\| X_{i(r-1)}^P - \hat{\mu}_{(r-1)} \right\|^2$$
$$= RSS(\hat{\mu})_{(r)} - RSS(\hat{\mu})_{(r-1)} = \Delta RSS(\hat{\mu})_{(r)} < \varepsilon \qquad (7.1)$$

which takes the same value regardless of the method of GPA used. However, if iterative GPA method (ii) is used, then this will be the same as considering:

$$\sum_{i=1}^{n} \left\| X_{i(r)}^P - \overline{X}_{i(r)}^P \right\|^2 - \sum_{i=1}^{n} \left\| X_{i(r-1)}^P - \overline{X}_{i(r-1)}^P \right\|^2 = RSS_{(r)} - RSS_{(r-1)} < \varepsilon \,, \qquad (7.2)$$

since then $\overline{X}^P = \hat{\mu}$ with $RSS = RSS(\hat{\mu})$. As described in 5.5.2, assessing convergence on the Procrustes fits is also in keeping with the methods of convergence assessment used in the iterative GPA methods, which are based on the same measure of sample variation in shape. (This is also the quantity we are specifically aiming to minimise, by allowing the semi-landmarks to move iteratively along their chords).

As with the original semi-landmark method, an alternative would be to assess convergence on the quantities being optimised at each movement stage, here in terms of the sums of squared Euclidean distances between the landmarks in $\hat{\mu}_{(r-1)}$ and $X_{i(r)}^{new} = vec^{-1}(vec(X_i^P) - U_i\lambda_i)$ or following an OLS superimposition of $X_{i(r)}^{new}$ to $\hat{\mu}_{(r-1)}$ as the semi-landmarks move along their chords. That is, stop the process if:

$$\sum_{i=1}^{n}\left\|X_{i(r)}^{new} - \hat{\mu}_{(r-1)}\right\|^2 - \sum_{i=1}^{n}\left\|X_{i(r-1)}^{new} - \hat{\mu}_{(r-2)}\right\|^2 < \varepsilon = \varepsilon_1 \qquad (7.3)$$

or if:

$$\sum_{i=1}^{n}\left\|SG_{i(r)}(X_{i(r)}^{new}) - \hat{\mu}_{(r-1)}\right\|^2 - \sum_{i=1}^{n}\left\|SG_{i(r-1)}(X_{i(r-1)}^{new}) - \hat{\mu}_{(r-2)}\right\|^2 < \varepsilon = \varepsilon_2 . \qquad (7.4)$$

However, again this does not always tell us whether or not the GPA registration has converged, as the $X_{i(r)}^{new}$ optimising (7.3) or (7.4) with respect to the previous mean shape, may not always be those which optimise the GPA registration to the updated mean, obtained at the end of the iteration. With any of the semi-landmark methods, including the original, it makes no sense to assess convergence on quantities such as those above, since the registration of the $X_{i(r)}^{new} = vec^{-1}(X_i^P) - U_i\lambda_i$ changes in relation to $\hat{\mu}_{(r-1)}$, which is also no longer the mean of the sample of the X_i^{new}. (The registration of the $X_{i(r)}^{new}$ needs to be standardised at the end of each iteration by GPA, producing configurations $X_{i(r)}^P$ and mean $\hat{\mu}_{(r)}$, on which convergence is then assessed).

7.3 GPA and scaling options for mean and fits

7.3.1 Full Procrustes criterion

As noted in 6.3.2, for the full Procrustes (FP) criterion, the prior registration of $Y^0 = X_{(r-1)}^P$ and $T = \hat{\mu}_{(r-1)}$ is irrelevant to the shape of $Y^{new} = vec^{-1}(vec(X_{(r-1)}^P) - U\lambda)$ $= X_{(r)}^{new}$ that is produced, since these differences are filtered out by the penalty function. Therefore, just as with the bending energy (BE) criterion, at each movement step, it does not matter which method of GPA is used to obtain the $X_{(r-1)}^P$ and $\hat{\mu}_{(r-1)}$ or if we use $T = \bar{X}_{(r-1)}^P$ as the reference shape, having used GPA method (i) to obtain the X_i^P and

mean $\hat{\mu}$ (although $\sum_{i=1}^{n}\left\|X_i^P-\overline{X}^P\right\|^2 \neq \sum_{i=1}^{n} d_F^2(X_i^P,\hat{\mu})=RSS(\hat{\mu})$ and so we would not assess convergence on this). Differences in the scales of $\hat{\mu}_{(r-1)}$ and the $X_{(r-1)}^P$, as described in 3.2.2.5, will produce different values of (6.77), but have no effect on the resulting shape of the $X_{(r)}^{new}$.

7.3.2 Nearest point criterion

For the nearest point (NP) method, however, different shaped configurations for $Y^{new}=X_{(r)}^{new}$ will result depending on the prior alignment of $Y^0=X_{(r-1)}^P$ and $T=\hat{\mu}_{(r-1)}$, and therefore depending on the method of GPA used. At each movement step, each new semi-landmark position in $X_{(r)}^{new}=vec^{-1}(vec(X_{(r-1)}^P)-U\lambda)$ is determined 'locally' as the point at least Euclidean distance to the corresponding landmark in $\hat{\mu}_{(r-1)}$, without allowing either $X_{(r)}^{new}$ or $\hat{\mu}_{(r-1)}$ to be rotated, translated or re-scaled.

Recall from section 3.2.2.5 that depending on the GPA method used, differences in the sizes of the estimated mean shape $\hat{\mu}$ and Procrustes fits X_i^P will result. These are recapped and summarised in table 7.1 below for GPA methods (i) and (ii) (the latter equivalent to method (iii) in section 3.2.2, with X_i pre-scaled to unit size). Recall also that, for both methods:

$$RSS(\hat{\mu})=\sum_{i=1}^{n}\left\|X_i^P-\hat{\mu}\right\|^2=\sum_{i=1}^{n} d_F^2(X_i^P,\hat{\mu}) \tag{7.5}$$

with $RSS=\sum_{i=1}^{n}\left\|X_i^P-\overline{X}_i^P\right\|^2=RSS(\hat{\mu})$, if iterative GPA method (ii) is used.

The following example compares the results of using the NP method on a small (n=10 configurations) test sample of upper central incisors with both GPA methods (i), using the eigenvector solution (since m=2) and method (ii), using the iterative algorithm described in section 3.2.2.2.1. We ignore method (iii) since the only difference from method (ii) when $\|X_i\|\neq 1$ is that the resulting mean shape and fits are on the scale of the data. Semi-landmarks and chord directions were as in fig. 5.12, representing unwanted patterns of variation due to differences in the positions of patient's gingival tissue.

	GPA method (i). Eigenvector solution (or method (ii) with $\hat{\mu}$ and X_i^P scaled to sizes below). Constraint: $\|\mu\|=1$	**GPA method (ii) (or (iii) with X_i pre-scaled to unit size). Iterative algorithms. Constraint: $\sum_{i=1}^{n}\|X_i^P\|^2 = n$**
$\hat{\mu}$	$\|\hat{\mu}\|=1$, $\hat{\mu}=c\overline{X}^P$, c some scalar	$\hat{\mu}=\overline{X}^P$
	Orientation and centres same	
X_i^P	$\|X_i^P\|=\cos\rho_i \leq 1$ minimise $\|SG_i(X_i)-\hat{\mu}\|^2$	$\|X_i^P\|>$ or <1
	Orientation and centres same	

Table 7.1: Properties of Procrustes fits and mean shape obtained by different GPA methods.
ρ_i=closest great circle distance between rotations of $\hat{\mu}$ and X_i^P on the pre-shape sphere (see fig. 3.3).

Fig. 7.1(first two columns) shows the original and final Procrustes fits after running the semi-landmark procedure with NP criterion, using both GPA methods, until convergence at $\Delta RSS(\hat{\mu})<0.0001$. (A convergence of 0.0001 for n=10 configurations follows the recommendation of $\Delta RSS(\hat{\mu})/n<0.00001$, suggested in section 5.5.6). Fig. 7.1(final column) displays the final estimated mean shapes, $\hat{\mu}_{SL}$, overlayed with original estimate $\hat{\mu}_0$, obtained with all landmarks fixed (there is no visible difference). The value of $d_F(\hat{\mu}_{SL},\hat{\mu}_0)$, the full Procrustes distance between $\hat{\mu}_{SL}$ and $\hat{\mu}_0$ representing the change in shape from the original estimate is indicated on each plot. If there is no sizeable bias entering the procedure, there will be no tendency for the mean shape to change over the course of the iterations.

Clearly, there is very little difference in the results using either GPA method. The two scatters of final fits and plots of the final estimated means are virtually identical, indicating that in practice it will not matter which approach is used. The subtle differences that were found are described below.

Using the iterative GPA method (ii), convergence took fewer iterations and the change in mean shape over the course of the iterations was slightly smaller, as noted in fig. 7.1. The occurrence of a larger systematic change in mean shape using GPA method (i), could be explained by the fact that $\hat{\mu}$ is not the arithmetic mean of the Procrustes fits and is always larger in centroid size than the X_i^P (see table 7.1). Intuitively, it perhaps therefore makes more sense with the NP method, that the Procrustes mean shape should

have positions which are the arithmetic means of each of the landmark positions of the Procrustes fits, which will be the case when using the GPA method (ii) throughout the procedure. If we were to use the arithmetic mean of the Procrustes fits from GPA method (i) as our mean shape when determining the movement of each set of semi-landmarks, but not convergence (as noted in the previous section), then the results are more similar to those obtained using the iterative GPA method (ii). A plot of the Procrustes fits revealed practically identical results to those in fig. 7.1. Convergence at $\Delta RSS(\hat{\mu})$<0.0001 was achieved in 5 iterations and a value of $d_F(\hat{\mu}_{SL}, \hat{\mu}_0)$=0.0075 is slightly smaller than that found using GPA method (i) above.

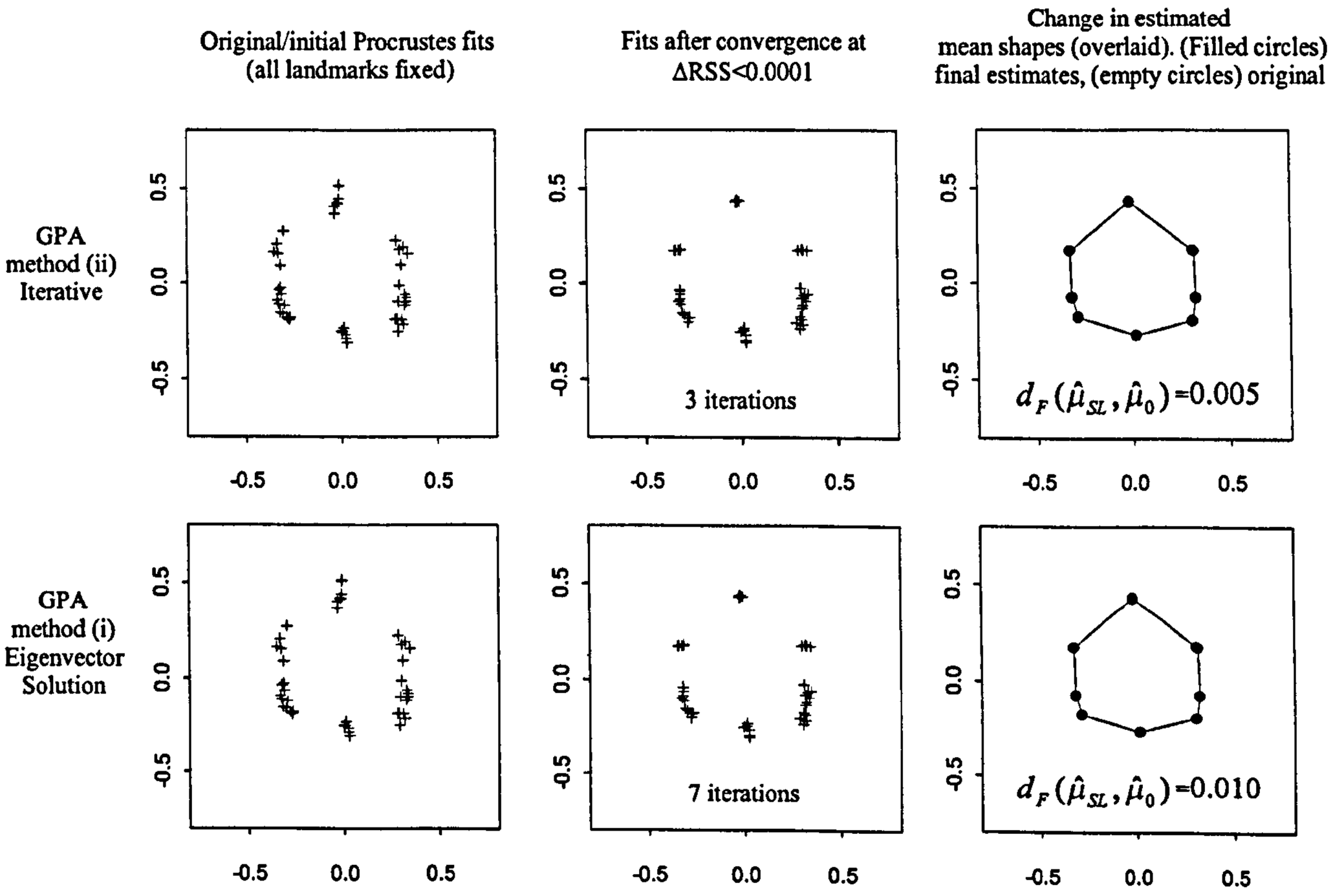


Fig. 7.1: Results of nearest point semi-landmark method using different GPA methods.

7.3.3 Use of partial Procrustes fits

Note that one could also obtain partial Procrustes fits, following GPA with method (i), by calculating configurations $X_i^{PP} = X_i^P / \|X_i^P\|$ with $\|X_i^{PP}\|$=1 and $\|\hat{\mu}\|$=1, and use these for the semi-landmark steps and when assessing convergence. For the FP (or BE) criteria, the new shapes produced on each iteration will be exactly the same as those

produced using full Procrustes fits obtained by either GPA method. As explained in 7.3.1, where the semi-landmarks move to is unaffected by the size of configurations. With the NP technique, there will be some differences in the resulting shapes, compared to when full Procrustes fits are used, as was seen with the different sized fits and means from each of the GPA methods in the previous section. However, after performing a similar investigation to that in section 7.3.2, it was found that any differences in the end results of using X_i^{PP}, compared to X_i^P, with $\|\hat{\mu}\|=1$, were extremely small. Assessing convergence using $\sum_{i=1}^{n}\|X_i^{PP}-\hat{\mu}\|^2$ rather than $\sum_{i=1}^{n}\|X_i^{P}-\hat{\mu}\|^2$ (before computing the X_i^{PP}), was also found to make very little difference with both methods.

Having obtained X_i^{PP}, one could also obtain $\overline{X}_i^{PP}$ and use this as the mean shape when determining new semi-landmark positions and assessing convergence. (An analysis using partial Procrustes fits as approximate tangent coordinates, would use $\overline{X}_i^{PP}$ as the mean shape.) However, $\overline{X}_i^{PP}$ is not the same shape as $\hat{\mu}$ and so slightly different semi-landmark positions will again result. Moreover, $\sum_{i=1}^{n}\|X_i^{PP}-\overline{X}^{PP}\|^2 \neq \sum_{i=1}^{n}\|X_i^{PP}-\hat{\mu}\|^2$. The first term is not optimised at the GPA step, whereas when GPA method (i) is used to obtain the X_i^P and then X_i^{PP}, relative to $\hat{\mu}$, the second term is. Consequently, the former would not be as good a measure for assessing convergence (in the same way as $\sum_{i=1}^{n}\|X_i^{P}-\overline{X}^{P}\|^2$ is not when GPA method (i) is used to obtain the X_i^P).

Options for using partial fits (following GPA by method (i)), when determining new semi-landmark positions, with $\overline{X}_i^{PP}$ or $\hat{\mu}$ as the mean shape, are included in the S-plus routine described in section 7.5. Convergence is then assessed using $\sum_{i=1}^{n}\|X_i^{PP}-\hat{\mu}\|^2$.

For the rest of this thesis however, we continue to concentrate on the use of full Procrustes fits, as in the preceding chapters.

7.4 Convergence issues

With both the data in 7.3.2 and various other sets of test samples, comprising shapes with a variety of combinations of landmarks, semi-landmarks and chord directions, convergence at a tolerance of $\Delta RSS(\hat{\mu}) < \varepsilon = 0.001 \times n$ was always found to be achieved in fewer than 10-15 iterations, when using either the FP or NP criterion to determine the new positions of semi-landmarks along their chords. However, as with original minimum bending energy (BE) approach, problems were again encountered with smaller values of ε, with the routine frequently failing to converge with both new methods. Again, for any dataset, it appears that there is always a tolerance level $\varepsilon_B < 0.001 \times n$, below which the process will not converge (at least within the first few hundred iterations), with $\Delta RSS(\hat{\mu})$ remaining at around the same magnitude ($> \varepsilon_B$) once this level has been reached.

What happens with both methods, as with the original BE approach, is that once we reach a certain level of convergence, on each iteration the configurations are still systematically changing (although only very slightly), moving in the same directions on each iteration by very small, but roughly consistent amounts. This results in a systematic change in the mean $\hat{\mu}$, obtained at the end of the rth iteration. On the following iteration, since $\hat{\mu}$ has changed, then so too have the optimal semi-landmark positions of each configuration, and so these must change in the same way again (and so on). On each iteration the change in $\hat{\mu}$ is very small, but once a certain level of $\Delta RSS(\hat{\mu})$ has been reached this becomes large in comparison to the change in shape of some of the X_i^P (relative to $\hat{\mu}$), resulting in a new set of configurations with new mean shape $\hat{\mu}_{(r)}$ and Procrustes fits $X_{i_{(r)}}^P$, such that,

$$\Delta RSS(\hat{\mu})_{(r)} = \sum_{i=1}^{n} \left\| X_{i(r)}^P - \hat{\mu}_{(r)} \right\| - \sum_{i=1}^{n} \left\| X_{i(r-1)}^P - \hat{\mu}_{(r-1)} \right\| > \varepsilon_B .$$

Specific details for each new method are described in the in sections 7.4.1 and 7.4.2 below.

7.4.1 Full Procrustes method

For the FP method, the process always converges at $\varepsilon = \varepsilon_A = 0.001 \times n$, usually after only 2, sometimes 3 iterations, but in general, with $\varepsilon = \varepsilon_B$, $\varepsilon_B < \varepsilon_A$, it does not. Once some minimum value has been reached, $\Delta RSS(\hat{\mu})$ is then typically seen to be always increasing on each iteration, as shown in fig. 7.2.

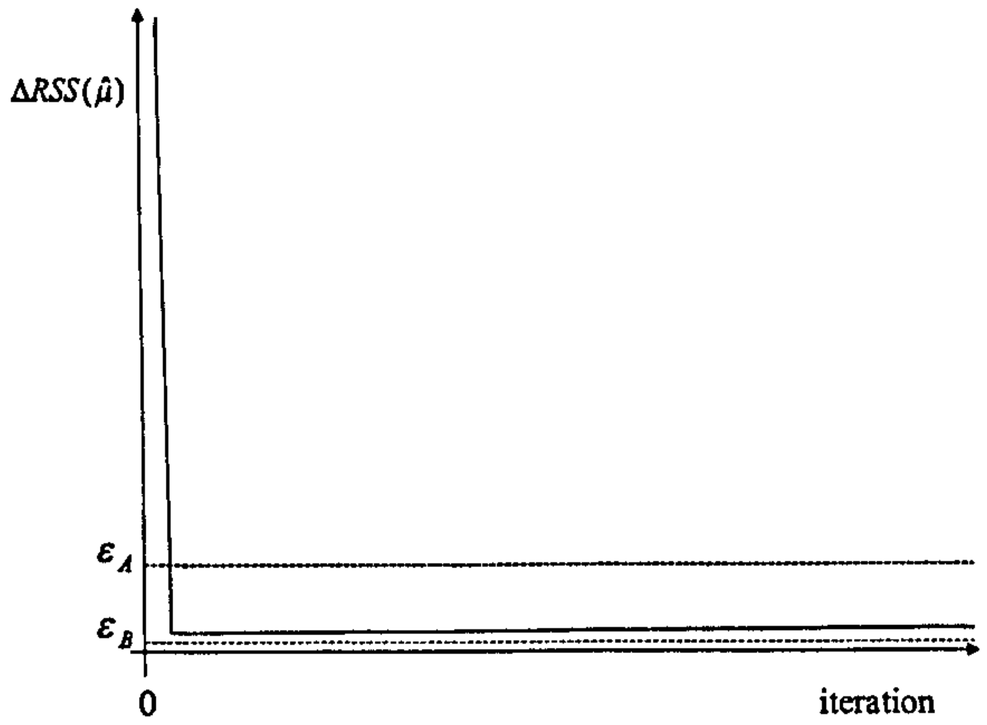


Fig. 7.2: Typical iteration history of $\Delta RSS(\hat{\mu})$ using Full Procrustes criterion to determine new positions of semi-landmarks on each iteration. (In comparison to fig. 5.9, the initial sharp decrease in $\Delta RSS(\hat{\mu})$ is quicker here than with the BE method.)

The result has no dependency on the method of GPA used throughout the iterations. As with the original BE criterion, this no effect on the shape of the new configurations produced at the end of each iteration, or on the summary measure $RSS(\hat{\mu})$, obtained after the GPA step, on which we assess convergence, as this takes the same value for any set of configurations comprising the same shapes.

For each configuration, on each iteration of the FP method, we have:

$$\min_{SG} \left\| SG(X_{(r)}^{new}) - \hat{\mu}_{(r-1)} \right\|^2 = \min_{SG,\lambda} \left\| SG(vec^{-1}(vec(X_{(r-1)}^{P}) - U\lambda)) - \hat{\mu}_{(r-1)} \right\|^2$$

$$\leq \min_{SG} \left\| SG(X_{(r-1)}^{P}) - \hat{\mu}_{(r-1)} \right\|^2 \approx \left\| X_{(r-1)}^{P} - \hat{\mu}_{(r-1)} \right\|^2, \qquad (7.6)$$

with '=' instead of '≈' if GPA method (i) is used and where the quantity on the left is that which is optimised to determine the new positions of the semi-landmarks. For all configurations, at the end of each iteration, $\hat{\mu}_{(r)} \approx \hat{\mu}_{(r-1)}$ and so

$$\left\| X_{(r)}^{P} - \hat{\mu}_{(r)} \right\|^2 \leq \left\| X_{(r-1)}^{P} - \hat{\mu}_{(r-1)} \right\|^2. \qquad (7.7)$$

Fig 7.3(a) (left) shows how on the first few iterations, the change in the $X^P_{(r-1)}$ to $X^{new}_{(r)}$ and $\hat{\mu}_{(r-1)}$ to $\hat{\mu}_{(r)}$ results in new Procrustes fits $X^P_{(r)}$ and mean shape such that (7.7) is true, usually by way of the situation shown in fig 7.3(a)(bottom left), since the optimal positions of the semi-landmarks are determined relative to $\hat{\mu}_{(r-1)}$, not $\hat{\mu}_{(r)}$ and so we expect that

$$\left\|X^P_{(r)} - \hat{\mu}_{(r)}\right\|^2 \approx \left\|SG(X^P_{(r)}) - \hat{\mu}_{(r)}\right\|^2 \geq \left\|SG(X^{new}_{(r)}) - \hat{\mu}_{(r-1)}\right\|^2$$

(although instances where the inequality is reversed, as in fig 7.3(a)(top left) are not uncommon). With either situation, (7.7) is true and so the contribution to $RSS(\hat{\mu})$ by each configuration is reduced and $\Delta RSS(\hat{\mu})$ decreases.

Although on the first few iterations, the size of the change in $\hat{\mu}$ is small compared to the changes in the X^P (relative to $\hat{\mu}_{(r-1)}$ and $\hat{\mu}_{(r)}$), this becomes comparatively larger when the process fails to converge from the second, third or fourth iteration onwards. At and beyond the point at which the process enters the flatter part of the plot in fig 7.2, systematic changes in $\hat{\mu}$, produced by systematic changes in the X^{new} result in $\hat{\mu}_{(r)}$ changing to be more different in shape from some of the $X^{new}_{(r)}$ (in terms of the difference between $X^P_{(r)}$ and $\hat{\mu}_{(r)}$) than $\hat{\mu}_{(r-1)}$ was from $X^{new}_{(r)}$, as depicted in fig. 7.3(a) (top or bottom right). That is, there are always configurations where the change in $\hat{\mu}$ and X^P always results in

$$\left\|X^P_{(r)} - \hat{\mu}_{(r)}\right\|^2 \geq \left\|X^P_{(r-1)} - \hat{\mu}_{(r-1)}\right\|^2, \tag{7.8}$$

leading to an increase in the contribution to $\Delta RSS(\hat{\mu})$ for these configurations. (An alternative representation of these Euclidean distances between points in tangent spaces to the pre-shape sphere, is shown in fig 7.3(b)). Convergence is therefore very slow with the small but continual changes in $\hat{\mu}$ and X_i stopping the value of $\Delta RSS(\hat{\mu})$ (obtained by either GPA method) falling below a certain level. Since there are always configurations for which (7.8) is true, this usually results in $\Delta RSS(\hat{\mu})$ increasing between iterations, as seen in fig 7.2.

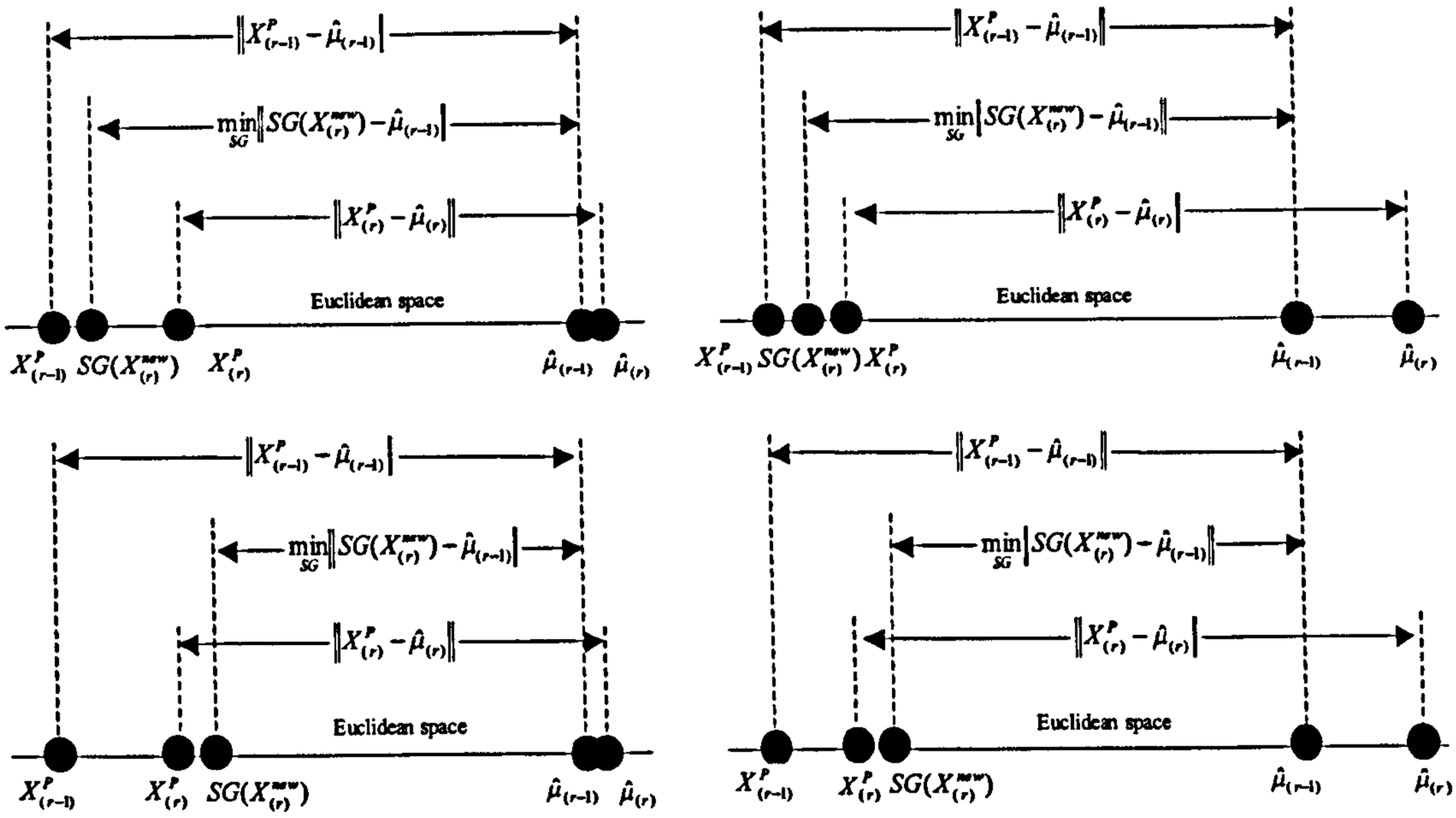


Fig 7.3(a): Euclidean distances following semi-landmark movement and GPA registration to previous and update means $\hat{\mu}_{(r-1)}$ and $\hat{\mu}_{(r)}$. (Top) Typical with NP method, (bottom) with FP method. (Left) while $\left\|X^P_{(r)} - \hat{\mu}_{(r)}\right\|^2 \leq \left\|X^P_{(r-1)} - \hat{\mu}_{(r-1)}\right\|^2$. (Right) while $\left\|X^P_{(r)} - \hat{\mu}_{(r)}\right\|^2 \geq \left\|X^P_{(r-1)} - \hat{\mu}_{(r-1)}\right\|^2$.

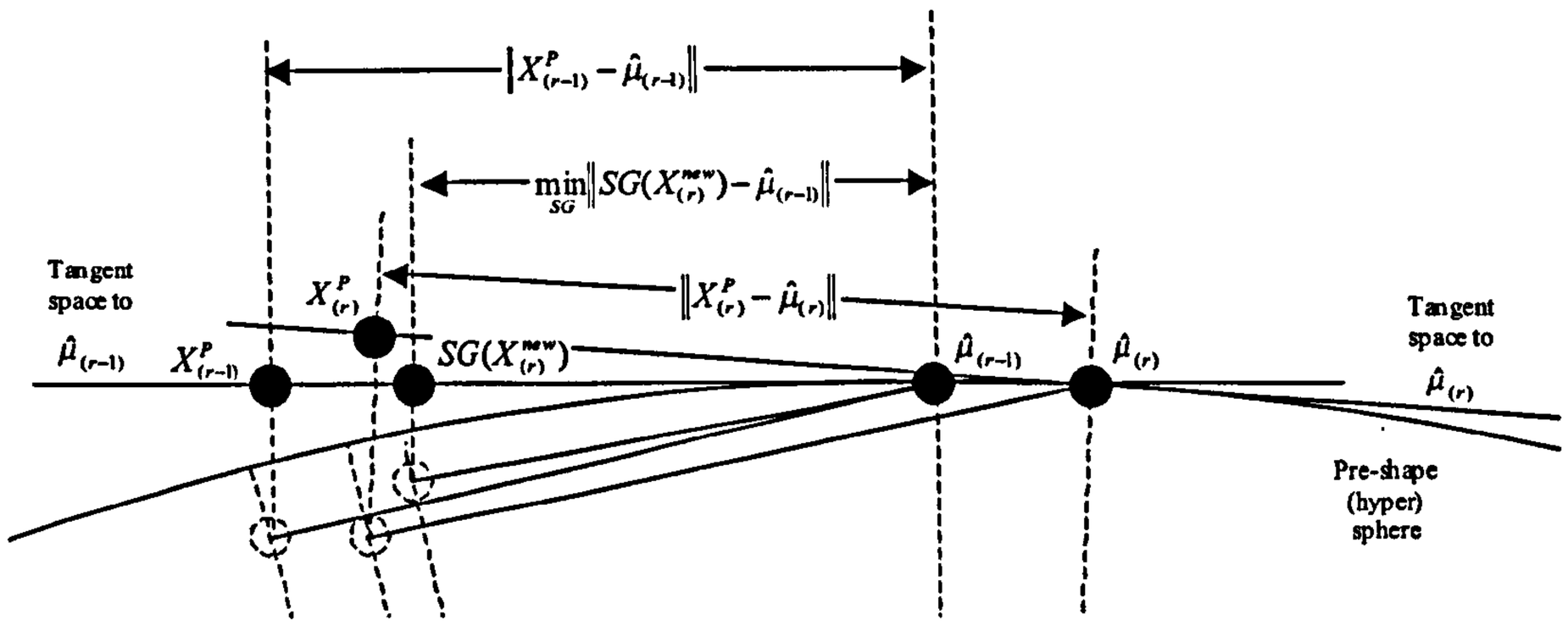


Fig 7.3 (b): Representation of Euclidean distances in fig. 7.4 (bottom left) as points in the tangent spaces to means $\hat{\mu}_{(r-1)}$ and $\hat{\mu}_{(r)}$, on the pre-shape sphere. (Dashed empty circles) Full Procrustes fits to $\hat{\mu}_{(r-1)}$ or $\hat{\mu}_{(r)}$ in pre-shape space. (Filled circles) Full Procrustes fits in tangent space. (Solid straight lines touching top of circle) tangent spaces. (Solid straight lines within circle) Full Procrustes distances.

7.4.2 Nearest point method

For the NP method, the process also always converges at $\varepsilon = \varepsilon_A = 0.001 \times n$, usually after several more iterations than when using the FP method, but in general, with $\varepsilon = \varepsilon_B$, $\varepsilon_B < \varepsilon_A$, again it does not. Regardless of the GPA method used, once the initial rapid

decrease in $\Delta RSS(\hat{\mu})$ has occurred, we still find $\Delta RSS(\hat{\mu})$ typically decreasing on each iteration, but by vary small amounts, as shown in fig. 7.4.

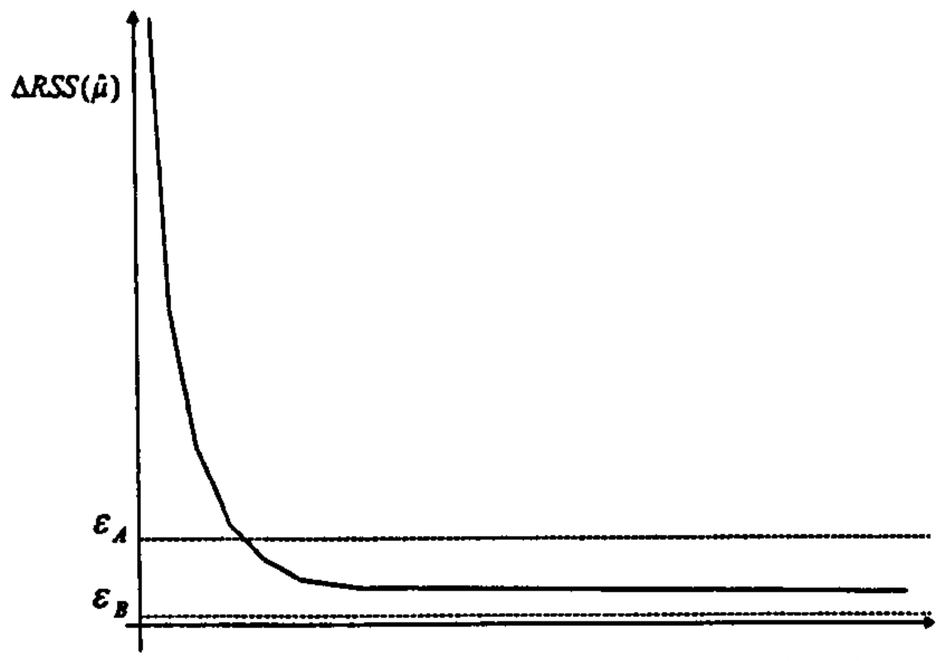


Fig. 7.4: Typical iteration history of $\Delta RSS(\hat{\mu})$. Nearest point criterion used to determine new positions of semi-landmarks on each iteration.

For each configuration, on each iteration, we have:

$$\left\|X_{(r)}^{new}-\hat{\mu}_{(r-1)}\right\|^2=\min_{\lambda}\left\|vec^{-1}(vec(X_{(r-1)}^{P})-U\lambda)-\hat{\mu}_{(r-1)}\right\|^2\leq\left\|X_{(r-1)}^{P}-\hat{\mu}_{(r-1)}\right\|^2 \quad (7.9)$$

where the quantity on the left is that which is optimised to determine the new positions of the semi-landmarks. At the end of any iteration r, $X_{(r)}^{new}$ is re-registered by GPA to a new estimate of mean shape $\hat{\mu}_{(r)}$ (with $\hat{\mu}_{(r)}\approx\hat{\mu}_{(r-1)}$) to obtain $X_{(r)}^{P}$. Because the ordinary least squares (OLS) registration gives equal weight to all pairs of corresponding landmarks as $X_{(r)}^{P}$ is registered to $\hat{\mu}_{(r)}$, the reduction in sum of squared distances between the semi-landmarks in $X_{(r)}^{new}$ and corresponding landmarks in $\hat{\mu}_{(r-1)}$ ($\approx\hat{\mu}_{(r)}$) in (7.9) is distributed around the residuals of the rest of the configuration. Since $\hat{\mu}_{(r)}$ hardly changes we expect that for each configuration:

$$\left\|X_{(r)}^{P}-\hat{\mu}_{(r)}\right\|^2\leq\left\|X_{(r-1)}^{P}-\hat{\mu}_{(r-1)}\right\|^2, \quad (7.10)$$

since $$\left\|X_{(r)}^{P}-\hat{\mu}_{(r)}\right\|^2\approx\min_{SG}\left\|SG(X_{(r)}^{new})-\hat{\mu}_{(r)}\right\|^2\leq\left\|X_{(r)}^{new}-\hat{\mu}_{(r-1)}\right\|^2\leq\left\|X_{(r-1)}^{P}-\hat{\mu}_{(r-1)}\right\|^2$$

by (7.9) (with equality on the left side if GPA method (i) is used). In other words, the contribution to $RSS(\hat{\mu})$ should always be reduced between iterations. For the NP method, fig 7.3(a) (top left) also illustrates how the change in the $X_{(r-1)}^{P}$ to $X_{(r)}^{new}$ and $\hat{\mu}_{(r-1)}$ to $\hat{\mu}_{(r)}$ results in new Procrustes fits $X_{(r)}^{P}$ and mean shape such that (7.10) is true

and so the contribution to $RSS(\hat{\mu})$ by each configuration is reduced and $\Delta RSS(\hat{\mu})$ decreases. We may also have configurations where the situation in fig 7.3(a) (bottom left) results, where

$$\left\|X^{P}_{(r)} - \hat{\mu}_{(r)}\right\|^2 \geq \left\|X^{new}_{(r)} - \hat{\mu}_{(r-1)}\right\|^2 \geq \min_{SG}\left\|SG(X^{new}_{(r)}) - \hat{\mu}_{(r-1)}\right\|^2,$$

although these are less frequent and still result in (7.10) being true. However, as we enter the flat part of the plot of $\Delta RSS(\hat{\mu})$ versus r (fig 7.4), new configurations eventually begin to be produced where the change in X^{P} and $\hat{\mu}$ results in:

$$\left\|X^{P}_{(r)} - \hat{\mu}_{(r)}\right\|^2 \approx \min_{SG}\left\|SG(X^{new}_{(r)}) - \hat{\mu}_{(r)}\right\|^2 \geq \left\|X^{P}_{(r-1)} - \hat{\mu}_{(r-1)}\right\|^2. \qquad (7.11)$$

Again, the systematic change in $\hat{\mu}$, produced by the systematic changes in the X^{new}, result in $\hat{\mu}_{(r)}$ being more different in shape to $X^{new}_{(r)}$ than $\hat{\mu}_{(r-1)}$ was to $X^{P}_{(r-1)}$, as depicted in fig. 7.3(a) (top or bottom right), resulting in an increase in the contribution to $RSS(\hat{\mu})$ for some configurations. The effect of occurrences of (7.11) is to offset any decreases in $RSS(\hat{\mu})$ which are produced, slowing the convergence and stopping the value of $\Delta RSS(\hat{\mu})$, dropping below a certain level.

7.4.3 Practical importance of convergence issue

Fig. 7.5 shows the original Procrustes fits (all landmarks fixed), and final fits after running the new semi-landmark methods, using (top row) the FP criterion with GPA method (ii) or (middle and bottom rows) NP criterion using GPA method (i) or (ii), for 100 iterations. Recall that the method of GPA makes no difference to the FP criterion. For both methods, an examination of the results after 100 iterations of failing to converge shows how if the semi-landmarks are systematically moving in the same directions for so many iterations, the shapes of the configurations can become notably distorted.

For the NP method, the slightly larger change in mean shape observed when using GPA method (i) in 7.3.2, becomes more evident as the procedure is run for more iterations.

As shown in fig. 7.5, at the 100th iteration, $d_F(\hat{\mu}_{SL}, \hat{\mu})$ is roughly twice as large using GPA method (i), compared to method (ii). From the resulting mean shapes it can be seen that over time, there is more of a tendency for the gingival landmark at the top of the LACC to move gingivally (upwards) and the papilla landmarks to move incisally (downwards), using GPA method (i), compared to (ii), perhaps explained by the fact that $\hat{\mu}$ is not the arithmetic mean of the Procrustes fits and is always larger in centroid size than the X_i^P, when using method (i), as suggested in 7.3.2 (see also table 7.1).

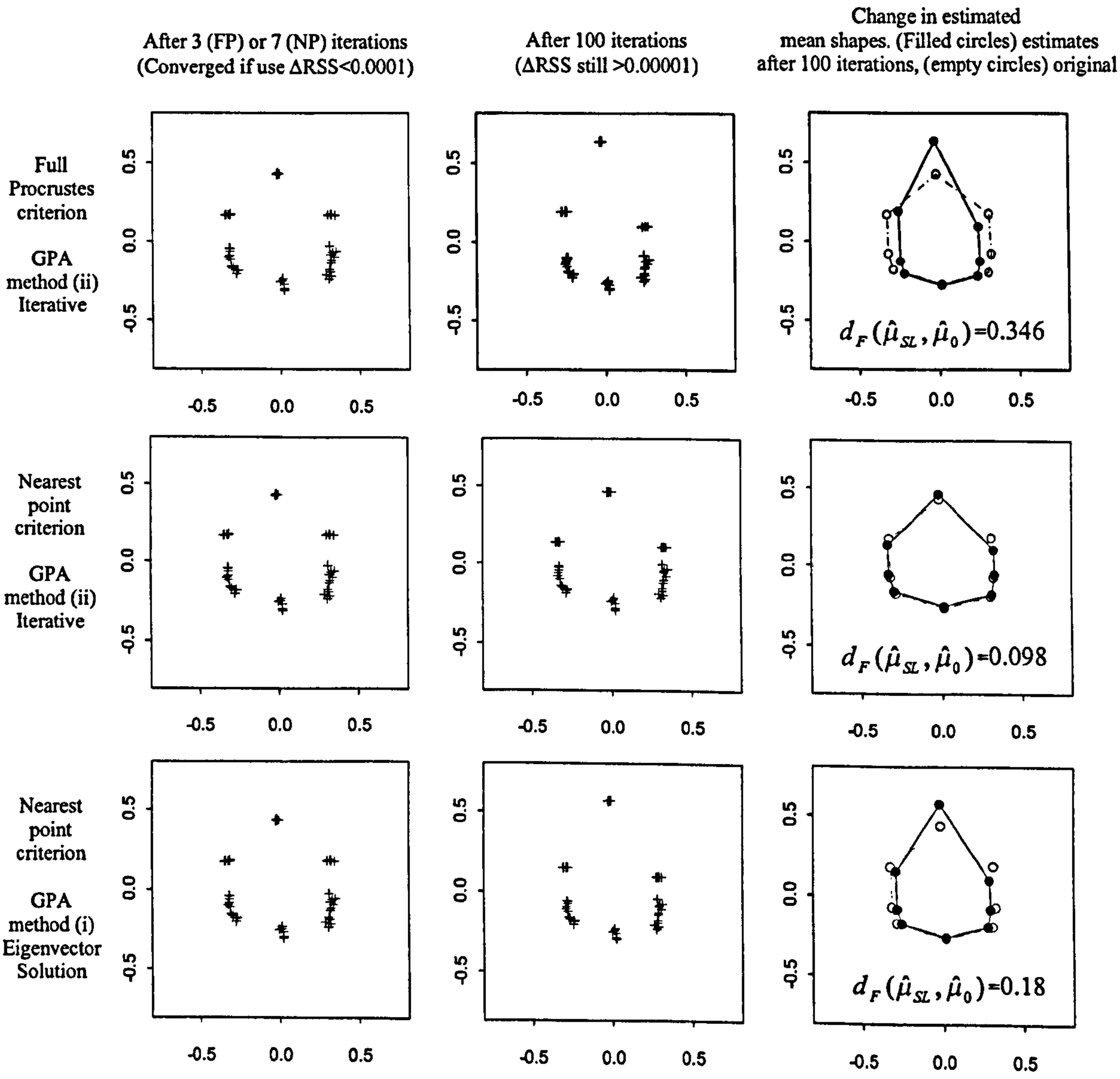


Fig. 7.5: (First two columns) Procrustes fits from semi-landmark method after convergence at $\Delta RSS(\hat{\mu})$ <0.0001 and after 100 iterations, using (top row) Full Procrustes criterion, (middle row) Nearest point criterion with iterative GPA method (ii), (bottom row) Nearest point criterion with GPA method (i). (Last column) Estimated mean shapes after 100 iterations, compared to initial estimates, when all landmarks fixed.

One possible workaround to the convergence issue, that was investigated, was the idea of 'damping' the movement of the semi-landmarks determined for each configuration, on every iteration. For $Y^0 = X^P$, $T = \hat{\mu}$ and $Y^{new} = vec^{-1}(vec(X^P) - U\lambda) = X^{new}$, having determined λ minimising the quantity:

$$\left\| vec^{-1}(vec(Y^0) - U\lambda) - T \right\|^2 \text{ or } \left\| SG(vec^{-1}(vec(Y^0) - U\lambda) - T \right\|^2,$$

the new version of the configuration was instead calculated as:

$$Y^{new} = vec^{-1}(vec(X^P) - U\delta\lambda) = X^{new}$$

where values of δ between 0.5 and 1.5 were investigated. However, with both methods it was found that for the data in section 7.3.2 and fig. 7.1 (as well as other datasets), this did little more than result in lower or higher (respectively) values of $\Delta RSS(\hat{\mu})$ on each iteration, producing an upward or downward shift in the appearance of the plots of $\Delta RSS(\hat{\mu})$ in figs. 7.2 and 7.4.

With both new semi-landmark methods, the same convergence issue (as described at the start of section 7.4) was also found to occur when using partial Procrustes fits for determining the new semi-landmark positions and assessing convergence. Again, once the initial rapid decrease in $\sum_{i=1}^{n} \left\| X^{PP}_{i(r)} - \hat{\mu}_{(r)} \right\| - \sum_{i=1}^{n} \left\| X^{PP}_{i(r-1)} - \hat{\mu}_{(r-1)} \right\|$ has taken place there is always a level of tolerance, $\varepsilon_B < 0.001 \times n$, below which the process will not converge.

Despite these findings however, in practice the issue is unlikely to be important. For all of our test datasets the change in X_i^P and hence $\hat{\mu}$, are very small, with $\Delta RSS(\hat{\mu})$ (or the analogous measure for partial Procrustes fits) less than $\varepsilon = 0.001 \times n$ with both new semi-landmark methods, so in practice this issue does not matter, especially when ε is set above the guidelines recommended in section 5.6.6. As noted in 5.5, this also then gives a level of convergence consistent with recommendations of Gower (1975) and Rohlf & Slice (1990), that with samples of size less than 100, a change in $RSS(\hat{\mu})$ of 0.001 or 0.01 is adequate when assessing convergence of the iterative GPA method. As before, the solution is to stop the each process before it enters the "uninformative

regime", and check whether or not a suitable level of convergence for the number of configurations involved, has been achieved, based on the recommendations above.

7.5 S-plus routine

The S-plus routine, 'semi.it.new' (2000-02) used for the investigations of the nearest point and full Procrustes semi-landmark methods in chapters 8 to 10, can be found in the appendix. The arguments which may be supplied, along with their default values are described below:

7.5.1 User arguments

`configs` $k\times2\times n$ real array of raw landmark configuration data. The j=1,...,k landmarks should be in the same order for each of the n configurations

`l1, l2` Mean shape and Procrustes fits always rotated so that landmarks l1 and l2 (1,...,k) in $\hat{\mu}$ lie horizontally (l1=left, l2=right). (Default: rotate $\hat{\mu}$ so major axis vertical).

`joinline` Vector sequence of landmark numbers 1,...,k in order to be joined by a continuous straight line edge whenever shapes plotted. (Default: no line).

`proc.opt` GPA method to use for obtaining Procrustes mean $\hat{\mu}$ and corresponding fits X_i^P.

1=Complex-eigenvector method (calls on Dryden's 'procrustes.2d' routine),

2=Iterative procedure of Gower (1975), Rohlf & Slice (1990). (Default: 2)

`GPA.crit` Convergence criteria for iterative GPA method. (Default: 0.0001)

`arith.mean` Ignored unless proc.opt=1. 1=Use arithmetic mean $\overline{X}^P$ as reference configuration during movement steps. 0=Don't. (Default: 0). If proc.opt=2, then $\hat{\mu} = \overline{X}^P$ anyway.

`Scale.opt` Ignored unless proc.opt=1. 1=Scale Procrustes fits to size 1 to give partial fits.

0=Leave as cos ρ. (Default: 0)

`DM` A $k\times2$ matrix in which the jth row is either a pair of zeroes, indicating that landmark j is not a semi-landmark, or a pair of landmark numbers, indicating that it is (i.e. j is always one of the sublist $j_{(l)}$). Each pair of landmark number is used to specify the unit chord directions $u_{j_{(l)}}\ (u_{jx_{(l)}}, u_{jy_{(l)}})^T$ for each semi-landmark, by calculating the unit direction vector between each pair of specified landmarks in each $X^P_{i(r-1)}$.

(Default: matrix of 0's, i.e. no-semi-landmarks).

`method` For determination of new semi-landmark positions in each $X^{new}_{i(r)}$.

0=Min. bending energy, 1=Full Procrustes, 2=Nearest point, 9=Don't slide. (Default: 0)

`conv.crit` Convergence criterion ε for $\Delta RSS(\hat{\mu}_r) = RSS(\hat{\mu}_r) - RSS(\hat{\mu}_{(r-1)}) < \varepsilon$.

(Default: 0.001 – Same as suggested by Gower (1975), Rohlf & Slice (1990) for GPA).

`SSinfo` "on" - Prints contribution to $RSS(\hat{\mu}_r)$ of semi and fixed landmarks in each $X^P_{i(r)}$, on each iteration. (Default: 0=no info.)

`max.it` Number of iterations to stop after, if still hasn't converged. (Default: 20).

`final.GPA` Option to do a different GPA on final X^P_i returned by $new configs (see below), having already reached convergence using method specified by proc.opt. (Default: return X^P_i as left by proc.opt). 1=Complex-eigenvector method, 2=Iterative method.

`ind.pics` Plots for each configuration at each iteration. (Default: no pics).

"on"- Plot of $X^{new}_{i(r)}$ (filled circles) and $X^P_{i(r-1)}$ (empty circles) on same figure, with chord directions shown as dotted lines through the nominal semi-landmark positions in $X^P_{i(r-1)}$. Lines extended in either direction to length specified by 'mag' (see below).

"on.super" – As above with plot of full Procrustes superimposition of $X^{new}_{i(r)}$ to $\hat{\mu}_{(r-1)}$.

"tps" – Pair of deformation grids showing PTPS mapping from $\hat{\mu}_{(r-1)}$ to $X^P_{i(r-1)}$ and $\hat{\mu}_{(r-1)}$ to $X^{new}_{i(r)}$, as in centre and right panels of fig. 5.8(a).

"tpscomps" - Plots of affine and non-affine components of each PTPS transformation above (as described in section 5.3.3 and in fig. 5.5, left and right panels).

"affinefits" - Plots of GLS affine fit of $\hat{\mu}_{(r-1)}$ to $X^P_{i(r-1)}$ and $\hat{\mu}_{(r-1)}$ to $X^{new}_{i(r)}$, as described in 5.3.5.4 and fig. 5.11. Displayed on a deformation grid.

`mag` See above. (Default=$\|\hat{\mu}\|/4$).

`final.pics` "on" - Plots of initial and final scatters of Procrustes fits about mean shape, ignoring "joinline" argument. (Default: no pics).

"B&A" – n individual plots of each original and final X^P_i on same figure.

The output values are the same as those from 'semi.it', described in 5.5.3.

`$newconfigs` $k \times 2 \times n$ real array of the final Procrustes fits.

`$mshape` Final Procrustes mean shape ($k \times 2$ matrix).

`$totalrss` Vector of values of $RSS(\hat{\mu}_r)$ obtained after initial and each GPA and at the end of each iteration.

`$changerss` Vector of values of $\Delta RSS(\hat{\mu}_r) = RSS(\hat{\mu}_r) - RSS(\hat{\mu}_{r-1})$.

`$iter` Number of iterations performed.

7.5.2 Optimisation algorithm ('nlmin')

For the FP criterion, new positions of the semi-landmarks in each $X_i^{new} = vec^{-1}(vec(X_i^P) - U_i\lambda_i$ are determined by seeking:

$$\min_{SG_i,\lambda_i} \left\| SG_i(vec^{-1}(vec(X_i^P) - U_i\lambda_i) - \hat{\mu} \right\|^2 \text{ or } \min_{SG_i,\lambda_i} \left\| SG_i(\hat{\mu}) - vec^{-1}(vec(X_i^P) - U_i\lambda_i) \right\|^2$$

$$= \min_{\beta_i,\gamma_i,\theta_i,\lambda_i} \left\| \beta_i vec^{-1}(vec(X_i^P) - U_i\lambda_i)\Gamma(\theta_i) + 1_k\gamma_i^T - \hat{\mu} \right\|^2$$

which, like the objective functions for the PTPS mappings with affine penalties, given by equations (6.65), (6.71) and (6.72) in section 6.2.3, have no straightforward analytical solution. Instead we must make use of an optimisation algorithm, such as the 'nlmin' function built into S-plus, for finding the minimum of a non-linear function. The optimiser is a standard, well-documented routine, based on a general quasi-Newton method. For further details see Dennis *et al.* (1981) and Dennis & Mei (1979).

Use of 'nlmin', within a routine such as 'semi.it.new' described above, requires some re-expression of the function to be minimised so that its only argument is a real vector, here λ_i, and the specification of a vector of the same length to be used as a starting point for the optimiser. For the FP objective function and those considered in section 6.2.3, a vector of zero's of length L was always used (L=number of semi-landmarks).

The function could also be used in determining λ_i when using the original BE or NP criteria, although the use of explicit solutions when available, allows much faster computation.

7.5.3 Flowchart representation

Finally, fig. 7.7 displays the different steps of the iterative semi-landmark procedure, using the 'semi.it.new' routine as a flowchart. Each grey box results in one of two or more options (the arrows leading out of the box), depending on the arguments specified to the function, which were described in 7.5.1.

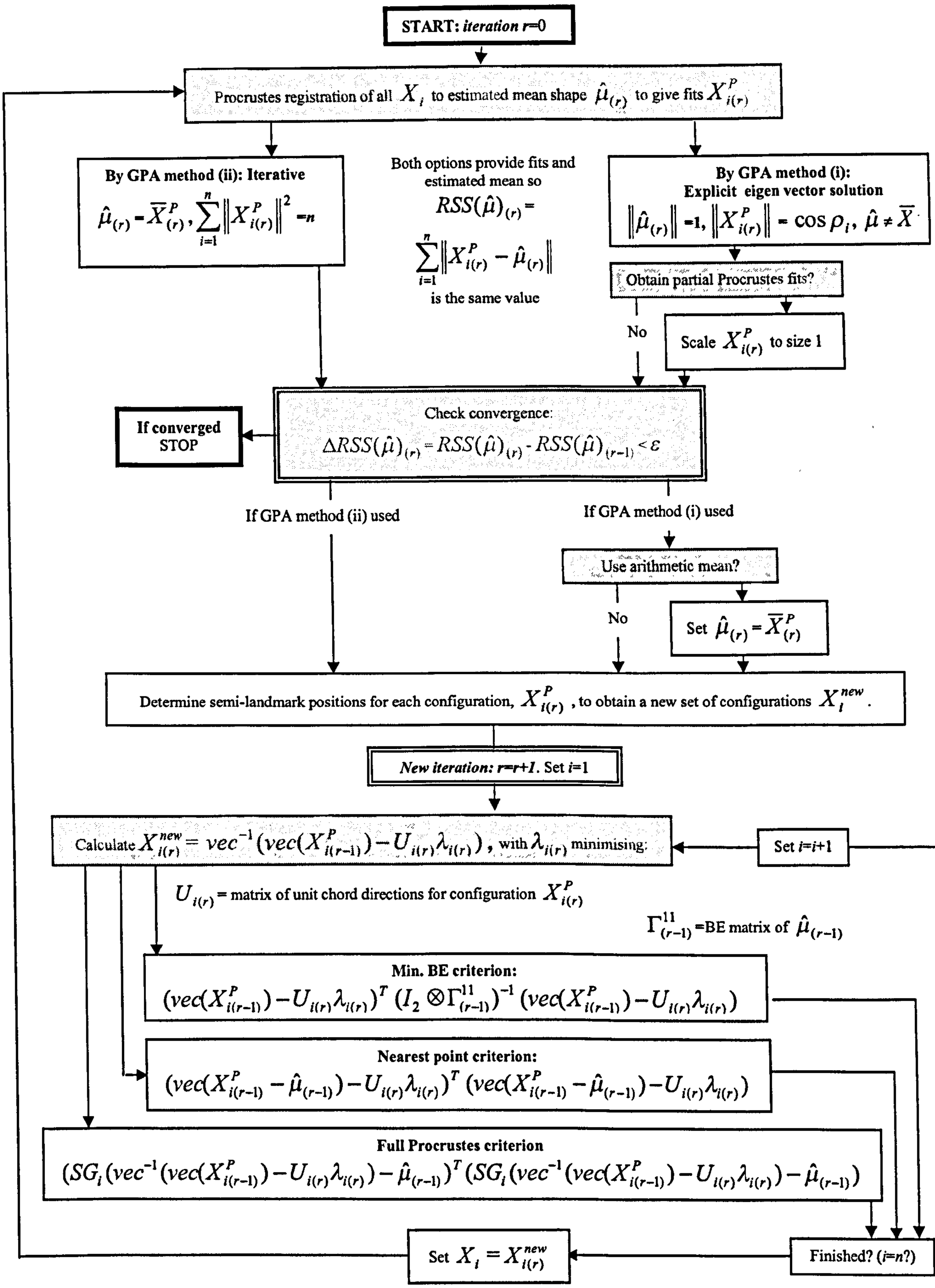


Fig 7.7: Flowchart indicating the different steps of the iterative semi-landmark procedure, using the 'semi.it.new' routine.

Chapter 8

Gingival simulation study

8.1 Introduction

In this chapter we aim to establish how well each of the newly proposed methods works in addressing the problem of unwanted variation in the gingival landmarks (those located on the gingival margin and interdental papillae).

The key aim with each of the semi-landmark methods is that variations in buccal tooth shape, due to differences in the curvature and extent of coverage of a patient's gingival tissue, be removed in the same way as we filter out differences due to size, location and orientation. In fact, we could consider such unwanted shape differences as a fourth equivalence class.

As well as the reduction or removal of such variation between patients, a successful method must also be able to identify that configurations obtained from the same individual tooth, but with the gingival margin and/or interdental papillae in different positions, are in fact still the same shape. For this reason we consider simulating tooth configurations for our test data, where different positions of the gingival landmarks have been generated according to some chosen probability distribution. The parameters of this distribution are chosen to reflect what is known about the variation in these positions in both the dental literature and from the experience of periodontologists.

The investigation centres on the buccal surface of the upper central incisor, with semi-landmarks and chord directions to be used as originally described in 5.6.2. However, the simulation and analysis described here is intended to be suitable for buccal surfaces of all tooth types.

Before any analysis was performed however, careful consideration had to be given to how the simulations were to be generated and used and to the criteria for comparing the methods being investigated.

8.2 Materials and methods

8.2.1 Images to be used

A first important decision was whether to simulate gum positions from images of actual extracted teeth (or equivalently teeth where the entire crown was visible, such as in cases of gingival recession) or to use images of unextracted teeth, where the gum tissues are already in one possible position.

8.2.1.1 Teeth where entire crown is visible

As noted in 1.2.1, by the 'shape' of a tooth we refer to the shape of the 'anatomical crown', the portion of tooth covered by enamel (Wheeler, 1974), which ends at the cemento-enamel junction (CEJ); the border between the crown and root of a tooth. So far we have only considered the shape of 'clinical crown'; what is visible of the anatomical crown in the presence of gingival tissue. What is observed in practice is some realisation of fig. 8.1, with the three gingival landmark positions located below (incisally) to the CEJ on the long axis of the crown and on either side of the tooth. The arrowed lines indicate approximate ranges of these positions.

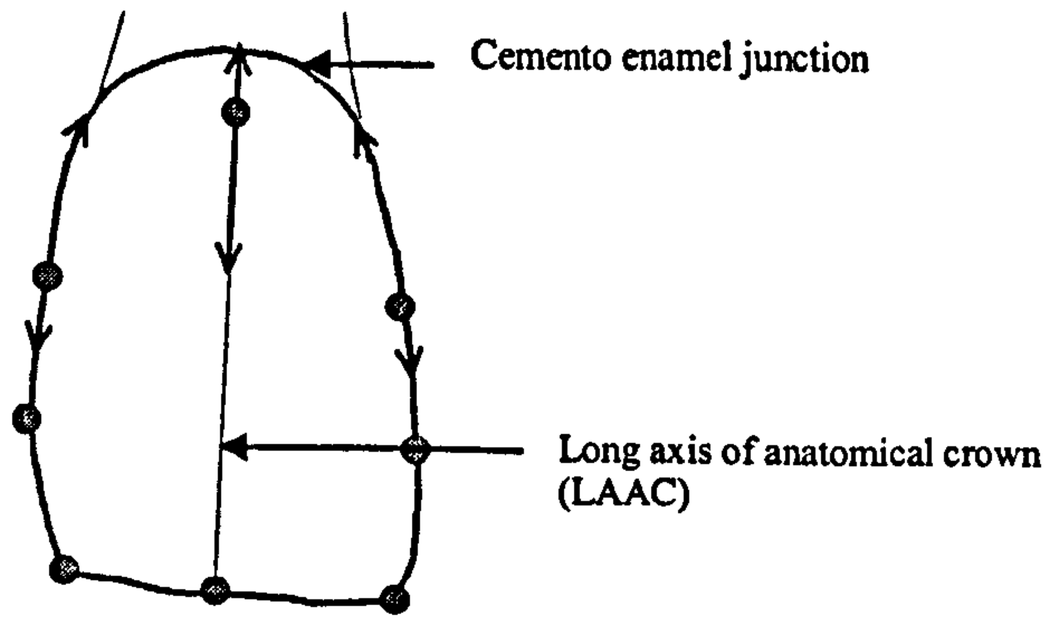


Fig. 8.1: Possible gingival landmark positions on full 'anatomical' crown of upper central incisor

The use of images with a fully visible crown would therefore allow us to simulate landmarks at different possible positions of the gingival tissue following this idea, using appropriately chosen probability distributions and ranges of variation. Dissimilarities in shape between two teeth from different patients or two representations of the same tooth that can be attributed to differences along the long axis of the anatomical crown (LAAC) or sides of the teeth, beneath the cemento-enamel junction, are of no interest.

Note however, the definitions used to identify the positions of the gingival landmarks so far have always assumed the presence of interdental papilla and that the cemento-enamel junction has always been obscured by the gingival margin. For the upper central incisor, the papilla landmarks were defined in section 4.3 as 'the highest (away from the incisal edge) points on the side of the tooth surface before being obscured by the interdental papilla'. The gingival margin landmark was defined as 'on the gingival margin at the end of the long axis of the clinical crown (LACC)'. If we are to use images of teeth where the crown is fully visible to simulate possible landmark positions, we require extensions to these landmark definitions; namely that the positions along the long axis of the crown and sides of the tooth surface should not be higher than the CEJ (i.e. on the root).

8.2.1.2 Teeth where entire crown is not visible

If we used images of unextracted teeth to simulate different positions of the gum, it would allow use of the data collected for the reliability study in chapter 4, whereas images of completely visible crowns would have to be collected specially. Incisally, gum landmarks could be simulated along the long axis of the clinical crown or down the visible sides of the tooth from those observed. However, in the other (cervical) direction, we would have no idea of how far beneath the gingival margin the landmark at the end of the long axis could be positioned, or of the shape of the sides of the tooth beneath the papillae. Fig. 8.2 indicates some of the different possibilities for a typical unextracted tooth.

An alternative approach considered was to create a single realistic possibility of a full crown outline from each image of a patient's unextracted tooth and then proceed as in fig. 8.1, simulating different possible gum positions as if we had the outline from the

fully visible crown. This would however, introduce an additional source of possible inaccuracy to the procedure.

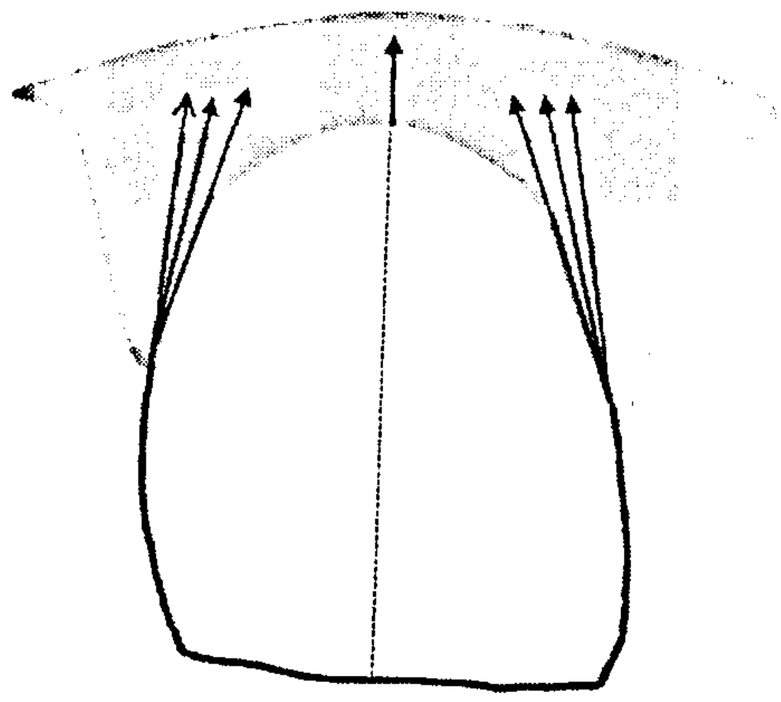


Fig. 8.2: Other possible landmark positions beneath the gingiva (gum tissue) will be constrained by the tangent directions of the tooth sides at the observed papilla landmarks (outer arrows) but could also lie at an unknown distance along any of the other directions shown.

In light of the guesswork involved in using unextracted tooth images, it was decided that images where the entire crown surface is visible would be the preferable option. Having the full actual outline of the crown also enabled us to determine whether or not each final configuration (produced by the semi-landmark method being considered) was still a feasible representation of the tooth in question.

8.2.1.3 Sources of images

Images of unextracted teeth, where the gingival margin had receded due to gingival recession and the entire surface of the crown was visible, were provided by the Department of Restorative Dentistry, School of Clinical Dentistry, University of Sheffield. Note that gingival recession has no relation to tooth shape (Rawlinson, A, personal communication). An additional source of images was from books. Periodontal books by Lindhe (1990), Wilson & Kornman (1996), Waite & Strahan (1989) and Heasman (1997) contain images of unextracted teeth as described above. Dental anatomy texts by Wheeler (1962, 1974) contain images of the crown surfaces of extracted teeth from different patients, often with some medical history. For each source of data, the same inclusion criteria used for the reliability study in chapter 4 were adhered to, with care taken to ensure that images were of sound healthy teeth with no evidence of attrition and were orientated in a consistent manner.

8.2.2 Acquisition of crown outlines, non-gingival landmarks and cemento-enamel junction 'markers'

Once images were obtained, the outline of each crown was extracted using the edge detection trace facility in Image Pro Plus v3.01 (Media Cybernetics, USA), see fig. 8.3 below. This information was then converted to full strings of outline coordinates, as described below, from which we can then simulate different possible positions of the interdental papilla landmarks. Note that in some instances, the edge detection method required assistance from the user along the CEJ where the boundary between root and crown was difficult to identify by differences in greyscale.

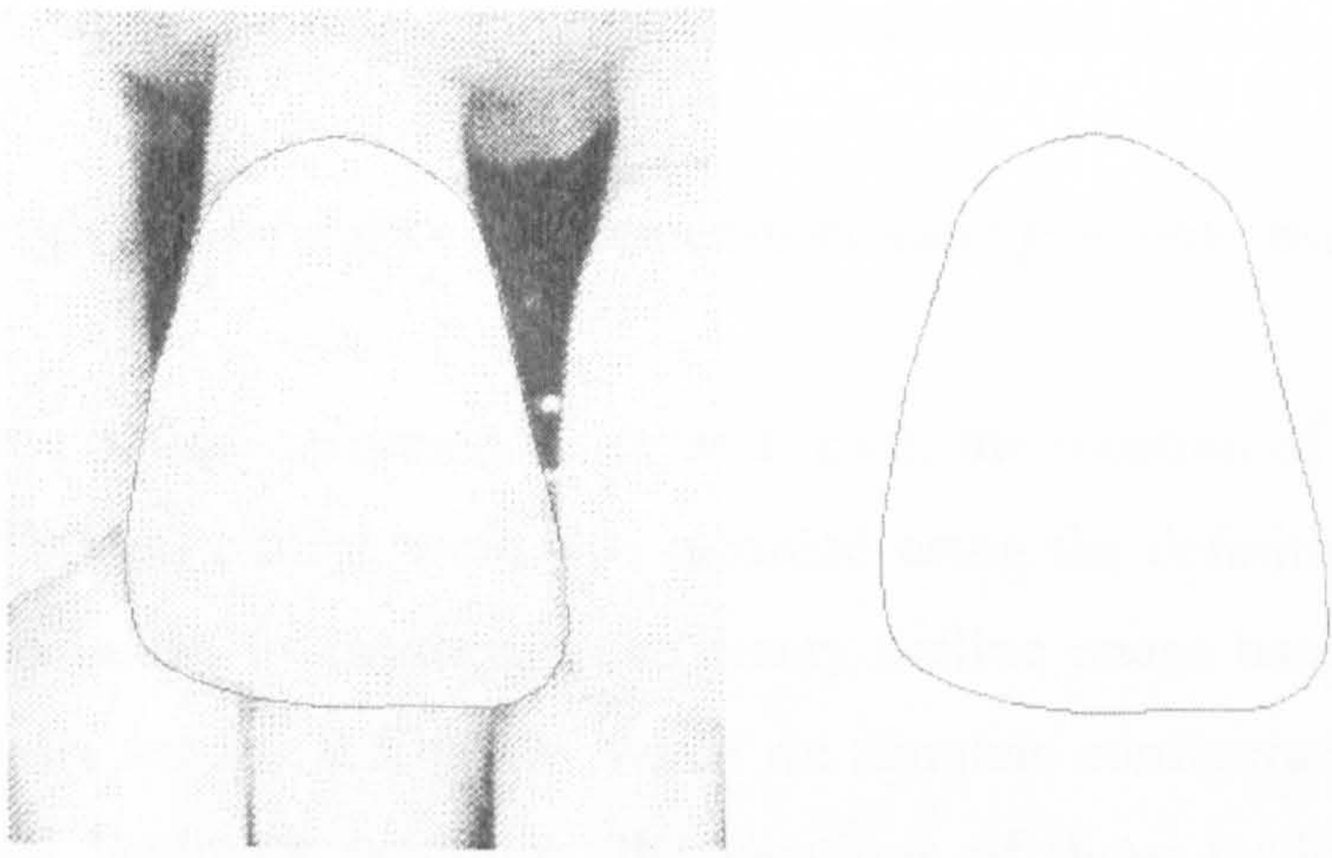


Fig. 8.3: (Left) Outline trace in Image Pro Plus. (Right) Extracted binary outline. (x,y) pixel axes are in the directions of the horizontal and vertical edges of the image.

8.2.2.1 Conversion of outline files

Image Pro Plus exports outlines as a text file of coordinate positions, where each point represents the start or finish of a run of coordinates with either no change in the x-position, y-position or where the absolute change in both x and y positions is identical. For example, the string of coordinates in fig. 8.4 would be recorded as (0,0), (0,2), (3,5), (7,5), (8,6). After importing these coordinates into S-plus a routine was written to fill in the 'missing' gaps in the coordinates of the outlines, up to the resolution of the image. So, for the string of coordinates above we would then have (0,0), (0,1), (0,2), (1,3), (2,4), (3,5), (4,5), (5,6), (6,5), (7,5), (8,6).

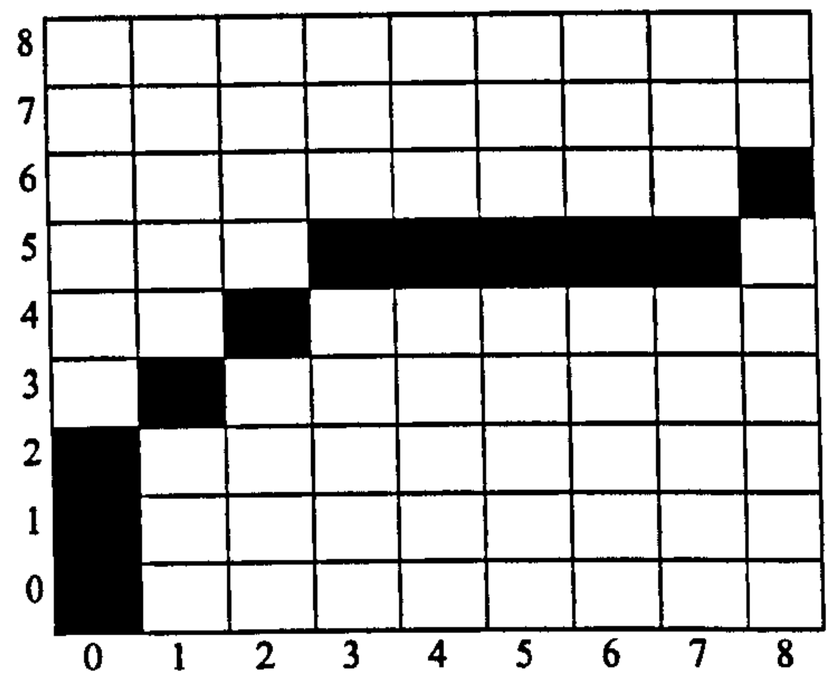


Fig. 8.4: Example outline string of pixel coordinates

The total number of pixels generated will depended on the different resolution of the images. Reasons why the full list of outline coordinates are required, rather than the representation produced by Image ProPlus, are described in 8.2.4.

8.2.2.2 Non-gingival landmarks and cemento-enamel junction 'markers'

In addition to the outline information for each case, the location of the five non-gum landmarks on the (x,y) outline were also recorded using the definitions in section 4.3. Identification was aided by overlaying the binary outline image back onto the original image of the tooth, see fig. 8.5 below. When we simulate configurations with different possible gingival landmark positions, the locations of these landmarks will remain unchanged.

Also recorded were the positions of three markers on the CEJ part of the outline; one at either side of the root and one roughly in the centre at the cervical end of the long axis of the crown. These were used later in the simulation of landmarks to provide a cervical (upper) bound for the papilla landmarks and the position of the LACC landmark.

In order to ensure that each set of non-gingival landmarks and the CEJ positions lie exactly on the outline string, a routine was written to identify the 'nearest point' on the outline to the recorded position of these points; on each occasion where the point location was a few pixels away from the outline, the nearest point position was used instead. Note that calibration of the coordinates was not performed since scales were often absent from the images used and differences in size are to be removed anyway.

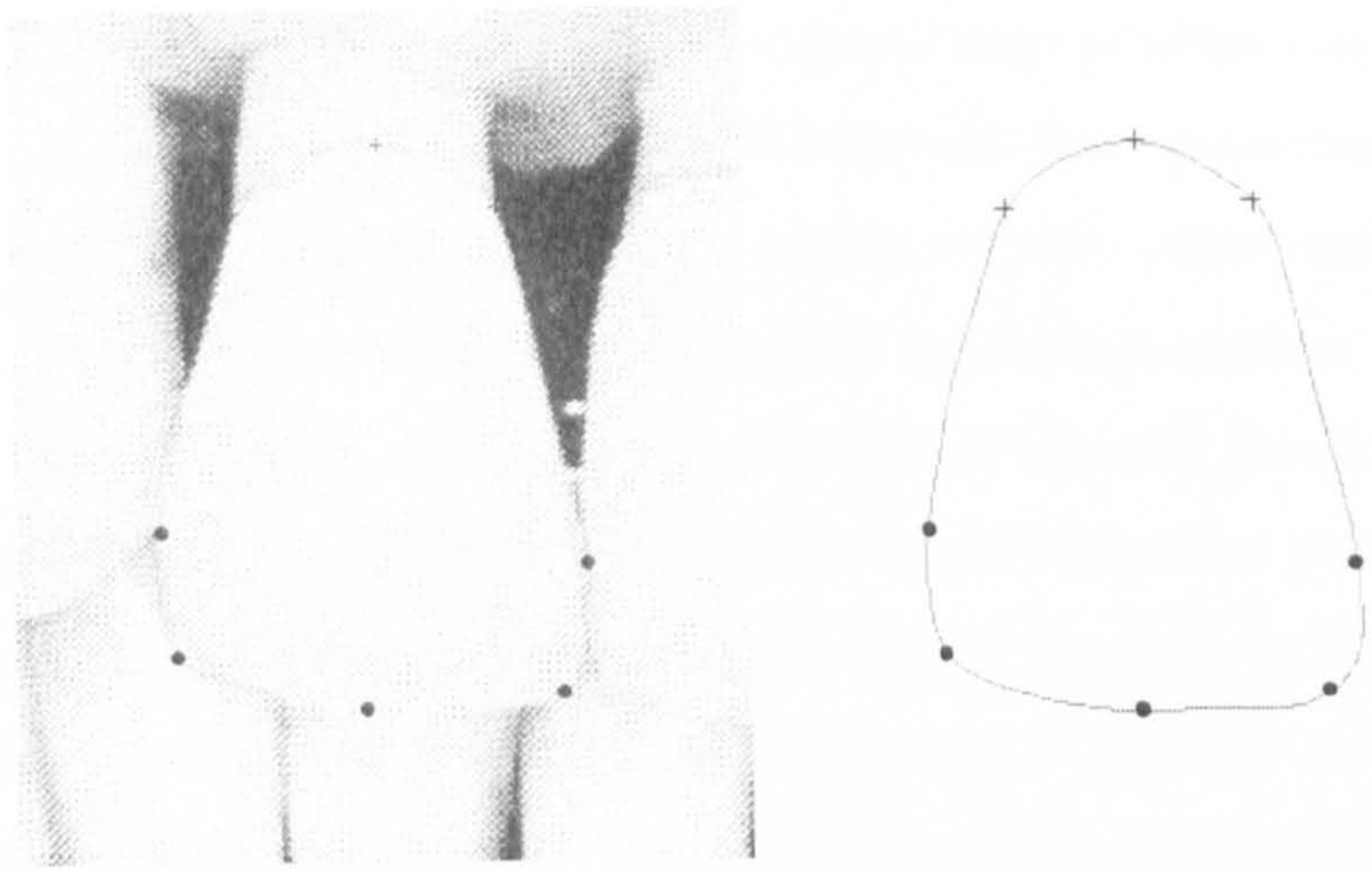


Fig. 8.5: Non-gingival landmarks (filled circles) and cemento-enamel junction markers (crosses) transferred from ImagePro Plus locations onto outline string.

8.2.3 Ranges of variation

So that our simulations represent feasible possibilities of what may be observed when the gum is present, conceivable ranges of variation for both the gingival margin landmark (along the long axis) and papillae landmarks (along the sides of the tooth) were required. The level and positions of the gingival margin and inter-dental papillae, the soft tissue profile, vary throughout an individual's life once teeth are fully erupted and at any one time can depend on many factors. These include gum health, underlying bone morphology, tooth brushing technique, age and the contact relationships with neighbouring teeth. For example, with increasing age or periodontal disease, gingival recession becomes more common (Wilson & Kornman, 1996).

For the gingival margin landmark at the end of the LACC, Wilson & Kornman (1996) state that the normal range of variation after complete tooth eruption is approximately 0.5-2.0 mm from the centre of the cemento-enamel junction. Linde (1990) provided the same range, whereas Heasman (1997) suggested 0.5-3.0mm, roughly in agreement with Wheeler (1974), who describes a 'normal' gingival margin as covering the cervical third of the tooth crown. On the basis of these values, and assuming an LACC of around 10mm, simulated positions between roughly 0% and 30% along the long axis of the full crown from the cemento enamel junction should therefore represent a reasonably realistic range of possibilities for the gingival margin landmark.

A formal range of variation for each of the interdental papillae was somewhat harder to find. However, information on the factors affecting these positions, expert opinion and the use of existing images of unextracted teeth were able to provide us with an adequate indication. Along the most prominent part of the buccal surface, the gingival margin normally follows a similar curvature to the cemento-enamel junction of crown and root (Wheeler, 1974). Consequently, on any given tooth outline we would expect the papilla landmarks to be at least the same distance down the sides of the teeth, towards the incisal edge, from the ends of the cemento-enamel junction as the gingival margin landmark is along the long axis. Where the gingival margin becomes the interdental papillae however, away from the most prominent part of the crown, their exact shape and positions are also determined by the contact relationships with the neighbouring teeth (Linde, 1990). Wheeler (1974) similarly states that the positions of the interdental papillae will depend on the relative positions and nature of the interproximal spaces between the adjacent teeth, which are filled by the papilla tissue, the 'apex' of the papillae (where we define the landmark) then typically being in the area of contact. The same positions were also suggested by Wilson & Kornman (1996), who stated that the 'coronal height' of the papilla generally resides immediately apical to the contact area of the adjacent teeth.

Another indication of the ranges in papillae positions, in relation to the gingival margin landmark and LACC, was obtained using the unextracted buccal surface configurations from the reliability study of chapter 4. We can translate, scale and rotate the configurations so that the endpoints of the LACC are aligned in the same positions in a manner similar to Bookstein's coordinate system (Bookstein, 1991), using equation (2.6). Fig. 8.6 shows the resulting scatter of papilla landmarks from one operator's representations of 19 upper central incisors after the ends of the LACC are transformed to positions (0,0) and (0,1). As can be seen, the papilla landmarks are typically between 0.4 and 0.9 of the way along the LACC from the incisal edge, with a suggestion that the distal landmarks (on the right of fig. 8.6) are generally more occlusally located than those on the mesial side. This would be expected since the contact area on the distal side is usually more towards the incisal edge. Note also that we would expect there to be a notable correlation in the position of the papilla landmarks on any particular tooth, which must be considered when simulating.

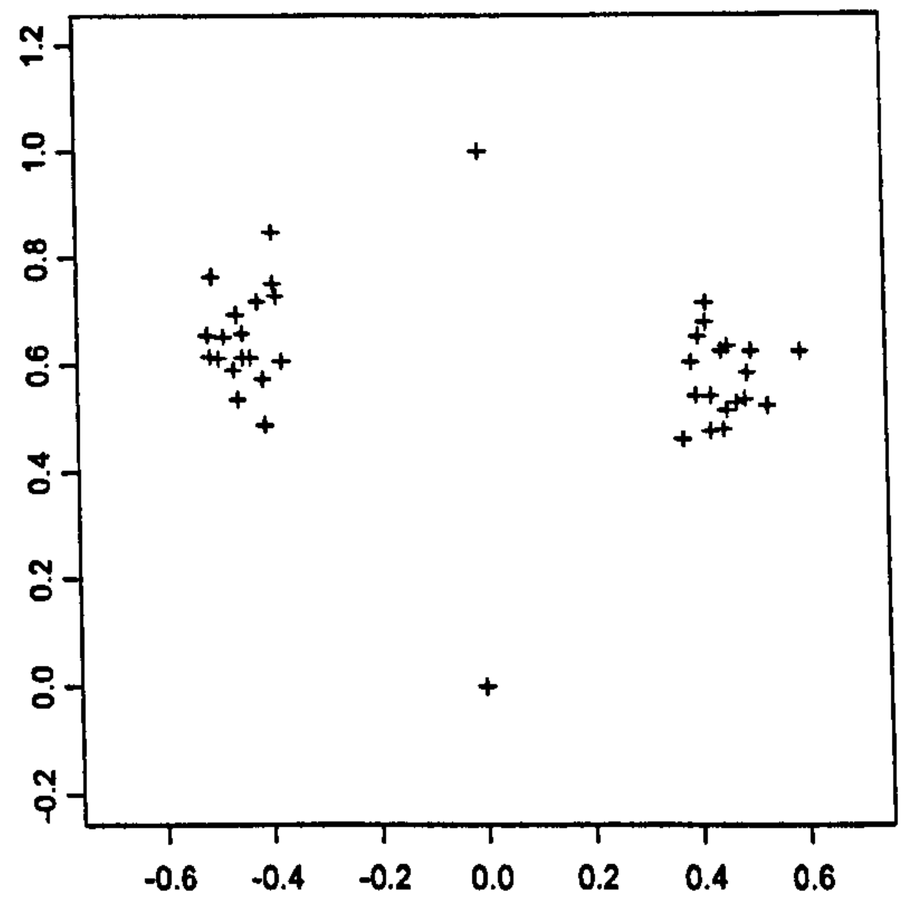


Fig. 8.6: Scatter of papilla landmarks on upper right central incisor following Bookstein-type superimposition using LACC endpoints.

Based on the data in fig. 8.6, the relationship between the y-coordinates of the aligned papilla landmarks could be described by the following simple linear regression equation (with standard errors of estimates in brackets):

$$\textit{Mesial papilla y-coord.} = 0.223\ (0.168) + 0.602\ (0.285) \times \textit{Distal papilla y-coord.} \quad (8.1)$$

A test of the significance of the slope term, or equivalently the regression sum of squares, gave p=0.049, suggesting evidence of a linear relationship between the aligned (standardised so that LACC is of length 1) mesial and distal papilla y-coordinates. The correlation coefficient between mesial and distal papilla y-coordinates, based on this data, was estimated as 0.456. A paired t-test between the y-coordinates in fig. 8.6 however failed to provide evidence of a difference in mean position between the mesial and distal sides of the upper central incisor. However, such a difference may still need to be considered on other teeth and so our method of simulation will make allowances for this, should it be necessary. Note that very similar results for the regression equation and t-test were produced when we used each of the three other operators' data.

Another way of acquiring information on the likely ranges of variation for all three gingival landmarks was to present several images of extracted teeth or crown outlines to dental researchers who then placed indicators on the images denoting possible ranges of variation. For the papilla landmarks, the results of an investigation using 10 different crown outlines, shown to three different dentists from within the School of Clinical Dentistry, University of Sheffield, were in good agreement with the ranges derived from

fig. 8.6, with an average range of approximately 50 to 80% along the LACC from the incisal edge being recorded for both landmarks on each upper central incisor.

To allow for any realistic possibility of the papillae positions when simulating, they were to be located at least the same distance from the ends of the cemento-enamel junction as the simulated gingival landmark was from its centre and at positions corresponding to no more than 40% of the way along the LACC from the incisal edge. In addition, the lower 40% bound should not be incisal to (below) the position of the corresponding endpoint of the mesiodistal width (by definition of the MD endpoints).

8.2.4 Simulation procedure

From each case outline and set of cemento-enamel junction markers we simulate sets of three interdependent gingival landmark positions. Each one of these sets is then combined with the same five non-gingival landmarks (identified in 8.2.2.2) to form a set of configurations, which have identical landmarks at the MD endpoints, incisal corners and incisal end of the LACC, but different positions of the three gingival landmarks, representing variation in the position of the gingival tissue.

For each set of gum positions we simulate a gingival margin landmark along the long axis of the tooth and then two papilla landmarks along the outlines of the sides of the crown, the latter pair being within ranges dependent on the position of first landmark. To do this, the ranges for the different landmark positions determined above need to be converted into appropriate ranges of values in the space of the outline and landmark coordinates. Each set of outline coordinates (and set of landmarks along them) was rotated, scaled and translated by a Bookstein type transformation so that the CEJ and incisal edge landmarks at the end of the LAAC, are sent to points (0,1) and (0,0) respectively, (see fig. 8.7). For each landmark, the simulation is then one-dimensional since it is only the distance in the y-direction along the long axis of the crown or along either side of the outline string that now needs to be generated. Each of the landmark coordinate positions can be uniquely identified by its y-coordinate alone, since, given whether the landmark is on the mesio or distal side or along the long axis of the tooth, the associated x-value can then be determined.

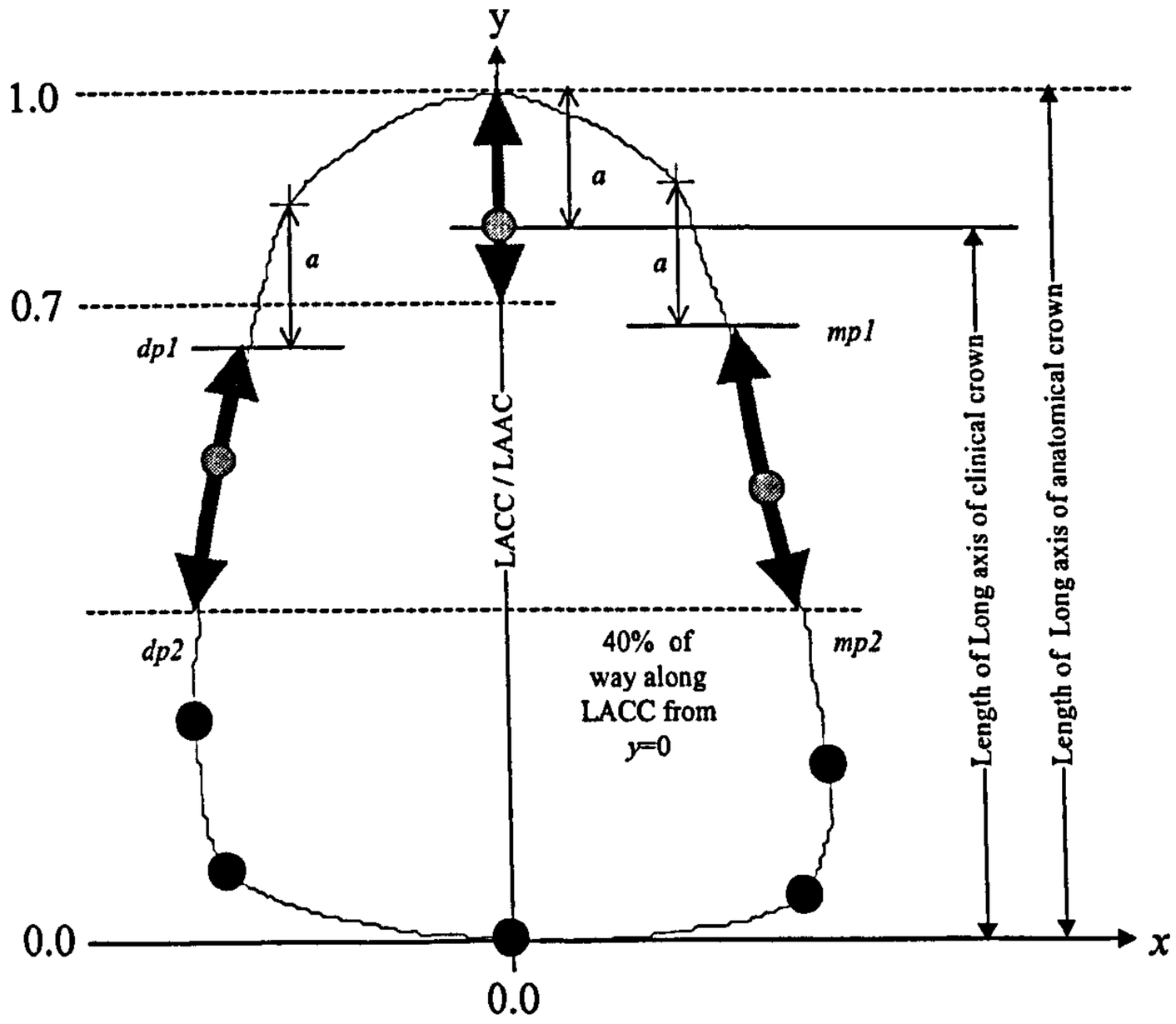


Fig. 8.7: Ranges of simulation for each of the three gum landmarks. Plusses denote cemento-enamel junction 'markers'. Filled circles are the non-gingival landmarks (endpoint of MD, incisal corners and incisal endpoint of LACC).

8.2.4.1 Central gingival margin landmark

For the gingival margin landmark, at the top of the LACC, the two *y*-values defining a range of positions at 0% and 30% of the way along the long axis of the crown, between which we will uniformly simulate *y*-coordinates, are then at positions with *y*-coordinates 1 and 0.7 respectively. Because the long axis of the tooth has been aligned between positions (0,0) and (0,1), the corresponding *x*-coordinate will always be 0.

8.2.4.2 Papillae landmarks

Having simulated a uniform *y*-value between 0.7 and 1, the range of values for the two corresponding papilla landmarks along the mesio and distal sides then depend on this position. For the cervical (upper) bounds, the *y*-distance between the simulated gingival landmark and the cemento-enamel junction at the top of the long axis was calculated. This is distance '*a*' in fig. 8.7. This is then added to the *y*-position of the markers at each end of the CEJ, to provide values '*mp1*' and '*dp1*' in fig. 8.7. For the two occlusal (lower) bounds, the larger of either the *y*-coordinate corresponding to a point at 40% of the way along the LACC from the incisal edge, or the *y*-coordinate of the mesio or distal

endpoint of the MD width was used. (The LACC is between the central incisal edge landmark and the simulated value for the gingival margin landmark). These are positions '*mp2*' and '*dp2*' in fig. 8.7. If one papilla landmark is expected to be located more occlusal or cervically than the other then the value of the occlusal bound may be based on different percentage points.

Positions for the mesio and distal papillae landmarks were then simulated by uniformly sampling *y*-values between the ranges (*mp1*,*mp2*) and (*dp1*,*dp2*). However, an alternative to simply using two independent uniform distributions was required, as this would ignore the correlation in the two papilla positions stated in 8.2.3 and, would not allow one position to be more cervical or occlusal located than the other (if desired). One possibility would be to simulate the first papilla position (uniformly) and then the second from a narrower (smaller variance) uniform distribution, conditional on the first position, with centre given by the regression relationship in equation (8.1). Alternatively, we could consider simulating directly from a joint distribution which has marginal uniforms, that may be non-symmetric if one *y*-value is expected to be higher than the other. However, to take account of both these features of the data we can make use of a readily available option in S-plus by simulating two correlated values from a bivariate normal distribution. The values produced would then have univariate normal marginal distributions which, by the reverse of the 'Probability Integral Transformation' (PIT), can be transformed to have correlated Uniform densities.

The bivariate normal distribution for zero mean, unit variance variates z_1 and z_2 has p.d.f.:

$$f(z_1,z_2) = \frac{1}{2\pi\sqrt{(1-\rho^2)}}\exp\left\{-\frac{1}{2(1-\rho^2)}(z_1^2+z_2^2-2\rho z_1 z_2)\right\} \qquad (8.2)$$

where ρ represents the correlation between the two variables. The marginal distributions each have standard normal p.d.f.s $f(z_i)$ and corresponding c.d.f.s $F(z_i)$, i=1,2, $-\infty < z_i < \infty$, given by:

$$f(z_i)=\phi(z_i)=\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}(z_i^2)\right\},\ F(z_i)=\Phi(z_i)=\int_{-\infty}^{z_i}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}t^2\right\}dt.$$

The PIT states that whatever probability density function $f(z)$ a random variable Z has, its cumulative density function $F(Z)$ has a Uniform (0,1) distribution. Consequently, simulated binormal values z_1 and z_2 can each be transformed to Uniform (0,1) distributions by simply calculating $v_1 = F(z_1)$ and $v_2 = F(z_1)$. These values are then re-scaled so that the ranges of the two distributions are transformed to the two pairs of *y*-coordinate limits, (*mp1*, *mp2*) and (*dp1*, *dp2*) stated above. (One of the values could also be shifted occlusally or cervically using the regression equation (8.1)). This then gives us a pair of simulated *y*-coordinates within the required ranges, whose distributions are correlated. Note that the sample correlation between the *y*-coordinates will not be the same as that specified for ρ in the bivariate normal distribution from which they are simulated. However, by trial and error, a value of $\rho = 0.5$ in equation (8.2) was found to produce the required correlation in the region of 0.46 between the simulated pairs.

To determine the *x*-coordinates corresponding to simulated *y*-coordinates of the papilla landmarks, the nearest *y*-coordinate within the relevant part of the string of outline data was identified and used instead along with its associated unique *x*-value. Alternatively, we could have identified the two (x,y) coordinates in the outline string whose *y*-coordinates are either side of the simulated value and used interpolation to obtain an *x*-coordinate, but this seemed unnecessary, given the resolution of the outlines involved.

8.2.4.3 Repeating the simulation

To produce another simulation from the same crown outline, the process is repeated. A new gingival margin location is obtained and then two papilla landmarks produced, from within a range dependent on the gingival margin landmark. After simulating different sets of 3 interdependent gingival landmarks by this process, these are then combined with the five non-gingival landmarks (identified in 8.2.2.2) to form a set of configurations which represent the same case, but with the gingival margin and papilla tissue in different possible positions. The process is then repeated for each other crown outline obtained in 8.2.2, to produce multiple possible realisations for each of the buccal surface images of upper central incisors (cases), collected in 8.2.1.3.

Fig. 8.8(left) below shows five simulated sets of possible gingival landmark positions, based on the outline, non-gingival landmarks and CEJ markers of one of the cases. Fig. 8.8(right) shows the GPA registration of the five configurations, representing the same case, but with the gingival landmarks in different possible positions.

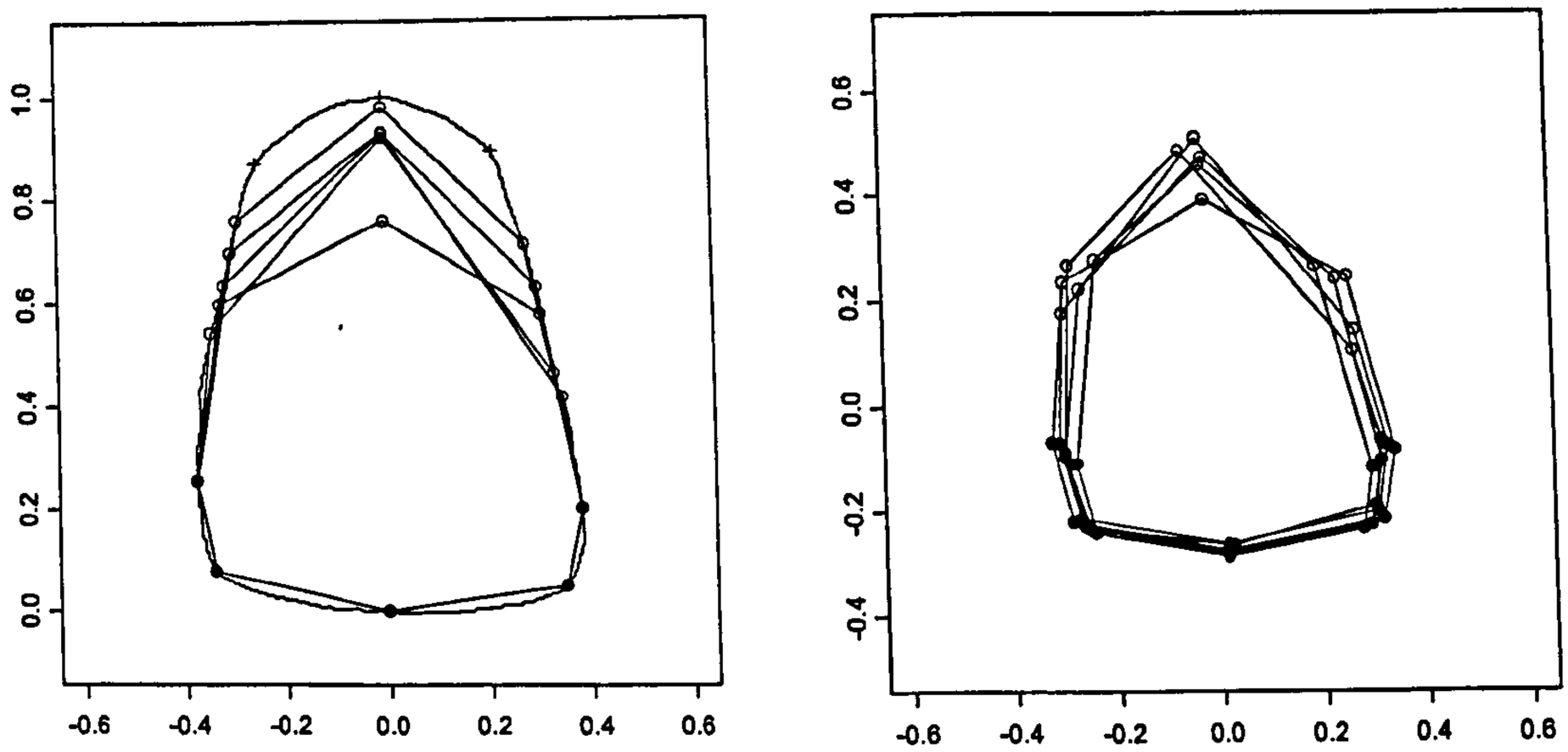


Fig. 8.8: Left: (Hollow circles) Five simulated sets of gingival landmarks positions from same case. (Solid circles) Fixed non-gingival landmarks at ends of MD width, incisal corners and incisal endpoint of LACC. (Plusses) Cemento enamel junction 'markers'. Right: Procrustes superimposition of the five simulated configurations, representing the same case.

8.2.5 Other data considerations and choice of summary measures

Once we are able to simulate multiple realisations of different teeth, the next important consideration was how the configurations should be used to investigate the different semi-landmark methods and what summary measures should be calculated and compared in order to evaluate their success.

The key aim is the reduction of variation in the Procrustes fits both 'between' and 'within' cases due to differences in the positions of the gingival landmarks along the direction of the LAAC or along the sides of the tooth. We have no interest in any differences in shape between cases that can be attributed to variation in the positions of the gingival landmarks in these directions. In addition we require that configurations representing the same tooth, but with differences in the positions of the gingival landmarks along the simulated directions still be regarded as the same shape.

8.2.5.1 Data choice

Each semi-landmark method aims to remove variation about a reference configuration, in practice the mean of the sample being considered. Having simulated multiple representations of different teeth with gingival landmarks in different possible positions, we have several choices of how the data should be used. In particular, we could run each method:

1. On the entire set of simulations from different cases;
2. Separately on the simulations from each case at a time; or
3. On sets of single representations from each case.

Obviously, each method must be run on the same set of data, so that any summary statistics obtained from the results, are directly comparable.

In order to examine whether we have reduced the variation 'between cases' options 1 and 3 would be preferable. Bookstein (1996a,d,e) looked at the usefulness of his semi-landmark technique by demonstrating a reduction in variation along unwanted directions between single representations of different brain cases (option 3). However, he did not study what happens 'within' cases which we are also interested in here. To examine whether we have reduced the variation within cases as well as between, option 3 is clearly less appealing, despite being closest to the situation we would have in practice. It would be difficult to calculate summary measures to show a reduction 'within cases' if there is only one representation per case. Consequently, for this problem, options 1 and 2 are the more obvious choices.

If we consider only one set of configurations from one case at a time (option 2) the final positions of the semi-landmarks will only be determined relative to each case mean. In practice, the methods would be run on representations from different cases and, by definition the semi-landmark positions determined with respect to their overall mean shape. Consequently, we need to ensure that different cases are used in the investigations here and allow the landmarks to move relative to the mean configuration obtained from different patients.

One reason for this is that when determining the new positions of the semi-landmarks, using 'global' criteria such as the minimum bending energy (BE) or Full Procrustes (FP) method, the result is dependent on the arrangement of all landmarks in a configuration. In addition, when aligning configurations, prior to computing the nearest point (NP) on each semi-landmark's chord, to the corresponding landmark in the reference shape, there will also be a dependency on all landmarks in each configuration, including the fixed ones (here the non-gingival landmarks). As was seen in 5.6.2 and 7.3.2, the arrangement of the fixed landmarks, relative to the reference shape, plays a major role in the determination of the new semi-landmark positions. Use of a variety of different cases should therefore ensure that there is adequate and realistic variation in the relative positions of the non-gingival landmarks represented in the data, in order to allow for different possible scenarios which may be encountered with certain particular tooth shapes. Otherwise, if we were to consider only one set of simulated configurations from each case at a time (option 3), the shape of the non-gingival landmarks would be identical in each dataset. It is perhaps then more likely that the semi-landmarks on the different simulations of the same case would all adopt the same positions (as desired), but these positions would have no meaningful relevance when compared 'between cases'.

In light of these considerations we run each semi-landmark method on the entire set of multiple simulations from the different cases (option 1). In this way, measures of variation both between and within cases, relative to the same overall mean can then be obtained. Appropriate numbers of cases and simulations per case in light of the issues discussed here are considered in section 8.2.6.

8.2.5.2 Choice of GPA method and summary measures

Summary measures are based on the final Procrustes fits of the s simulations of each of the n cases, registered to their final overall estimate of mean shape $\hat{\mu}_{SL}$ and denoted as $X_{im}^{P_{SL}}$, $i = 1,\ldots,n$, $m=1,\ldots,s$. (Recall that in the 'semi.it.new' routine, described in 7.4, each iteration ends with a GPA step, on which convergence is assessed, so that the final configurations returned are the Procrustes fits to their final estimated mean shape, in preparation for any subsequent tangent space analysis). These summary measures are

then to be compared between the different semi-landmark methods and with the same summary measures based on the initial Procrustes registration of the simulated data, with all landmarks fixed. In addition, the variation between and within cases is displayed in terms of scatter diagrams of the final Procrustes fits from each method and also compared to result of the initial (ordinary) GPA of the data.

In order for the results to be comparable, we require that the same GPA method be used throughout the investigations. In 7.3 we considered the effects of using different Procrustes registration methods on the determination of semi-landmark positions, using the original and new semi-landmark methods. The new positions of semi-landmarks determined by using the BE and FP criteria, have no dependency on the alignment of the individual and reference configuration and so for these methods it does not matter which GPA approach we use. For the NP criterion, section 7.3.2 showed how slight differences can occur in the results obtained using different GPA methods and argued how if one were to choose, then it would perhaps make more sense to use a Procrustes mean which is the arithmetic mean of the Procrustes fits as the reference shape, rather than a version which was larger than each of the Procrustes fits.

Use of the iterative GPA method (ii) with estimated mean $\hat{\mu} = \sum_{i=1}^{n}\sum_{m=1}^{s} X_{im}^{P_{SL}} / ns$ also means that a partition of $RSS = \sum_{i=1}^{n}\sum_{m=1}^{s}\left\| X_{im}^{P_{SL}} - \hat{\mu}_{SL}\right\|^2$ into sums of squared Euclidean norms between cases and within cases, using the 1-way ANOVA identity in section 4.4.2.3 is a partition of $RSS(\hat{\mu}_{SL}) = \sum_{i=1}^{n}\sum_{m=1}^{s}\left\| X_{im}^{P_{SL}} - \hat{\mu}_{SL}\right\|^2 = \sum_{i=1}^{n}\sum_{m=1}^{s} d_F^2(X_{im}, \hat{\mu}_{SL})$, the same quantity on which we assess convergence of each method (see section 7.2). That is, from equation (4.13) we can partition the sum of squared full Procrustes distances or sums of squared Euclidean norms of the final configurations $X_{im}^{P_{SL}}$, $i = 1,\ldots,n$, $m=1,\ldots,s$, about their estimated mean shape $\hat{\mu}_{SL}$ using the identity:

$$EucSS(total) = EucSS(cases) + EucSS(within)$$

where

$$EucSS(total) = RSS = \sum_{i=1}^{32}\sum_{m=1}^{50}\left\| X_{im}^{P_{SL}} - \overline{X}^{P_{SL}} \right\|^2 = \sum_{i=1}^{32}\sum_{m=1}^{50} d_F^2(X_{im}^{P_{SL}}, \hat{\mu}_{SL}),$$

$$EucSS(cases) = \sum_{i=1}^{32}\sum_{m=1}^{50}\left\| \overline{X}_i^{P_{SL}} - \overline{X}^{P_{SL}} \right\|^2 \approx \sum_{i=1}^{32}\sum_{m=1}^{50} d_F^2(\overline{X}_i^{P_{SL}}, \hat{\mu}_{SL}),$$

$$EucSS(within) = \sum_{i=1}^{32}\sum_{m=1}^{50}\left\| X_{im}^{P_{SL}} - \overline{X}_i^{P_{SL}} \right\|^2 \approx \sum_{i=1}^{32}\sum_{m=1}^{50} d_F^2(X_{im}^{P_{SL}}, \overline{X}_i^{P_{SL}}).$$

Having run each method of semi-landmark registration, it is clear that the resulting variation 'within cases', i.e. *EucSS*(*within*) should be as small as possible. Representations of the same tooth, differing only in the possible positions of the gingival landmarks along (almost) the same directions, need to be considered as the same shape. Since the non-gum landmarks do not change 'within-cases' any change in this variance measure will be due to changes in the gum positions.

However, simply minimising *EucSS*(*within*) alone may not always produce desirable results. A smaller *EucSS*(*within*) may result if all the simulations from the same case end up converging to the same unrealistic shape. An alternative option considered was to treat the 'within case' means of the original simulated positions as each cases 'true shape' and to measure the variation of the final configurations around these 'true shapes' after running each of the semi-landmark methods. However, minimisation of this criterion would stop the simulations from being able to adopt positions in accordance with our principal aim; that any variation in shape between representations from different cases, about the overall mean shape, attributable to differences in gum position be removed. Consequently, we still need to allow the within-case means to be able to change from their original positions and so we regard semi-landmark positions anywhere within the bounds of their simulated positions as still being representative of their 'true' shape.

In order to reject unrealistic shapes, at the end of running each method, the final three gum landmark positions of each configuration were required to be still within the ranges or bounds they were simulated between. This would then ensure that each final configuration is still a 'plausible' representation of the tooth from which it was simulated (in reality recorded from). For example, this would ensure that the gingival landmarks do not move above the cemento-enamel junction (CEJ).

In section 5.6.2, the resulting configurations using the original semi-landmark method were (subjectively) deemed as 'unrealistic', although we had no way of measuring how 'unrealistic' they were. Here we have the entire 'true' shape of the tooth surface and information on the feasible ranges of the gingival semi-landmarks, given the position of the CEJ, which was previously obscured, and so we are able to check the realism of the final positions of the semi-landmarks. This involved re-registering each of the final configurations by a Bookstein-type transformation so that two of the non-gingival landmarks (e.g. the MD endpoints) are sent to the positions they were in when each set of simulations was produced, for the particular case in question. A check is then made that the *y*-coordinate of the gingival margin landmark, at the top of the LACC, is still within 0.7 and 1.0, as in fig. 8.7, and that the *y*-coordinates of the papilla landmarks are still between the bounds (*mp1*,*mp2*) and (*dp1*,*dp2*), the feasible positions between which these landmarks were simulated.

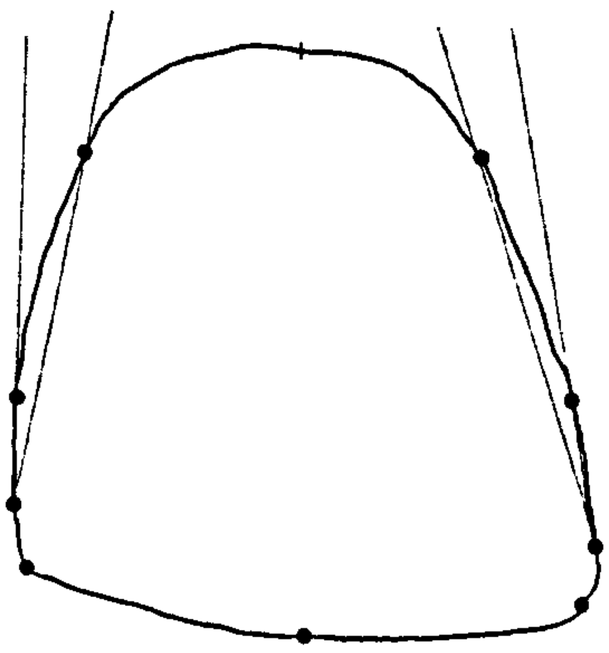
To check that a method of registration also reduces the variation 'between cases' due to differences in gum position, we naturally seek a reduction in the value of *EucSS*(*cases*). However, we also consider the variation in individual configurations about the grand mean as well, i.e. *EucSS*(*total*) since in practice we will have only single representations from each case, rather than the mean of each configuration. Note however, that because of the aim of reducing variation between cases, it is not sensible here to use reliability measures (as in the semi-landmark example of 5.6.1).

Finally, it may also be worthwhile considering the distances of the final positions of the papilla landmarks from the actual outline of the tooth they were simulated on. On some tooth surfaces the use of chord directions may result in new semi-landmark positions far away from the original outline. This would occur with those cases where the sides of the teeth have high curvature as illustrated in fig. 8.9 below.

In general, the chord directions of the semi-landmarks may change during the iteration, if the landmarks which define them are also able to move. In this application however, the directions are defined through the mesio/distal MD endpoints, which are always fixed, and the mesio/distal papilla semi-landmarks and so do not change from the initial directions determined from the semi-landmarks' original starting positions.

The effect of differences in the starting positions of the papilla landmarks and their associated chord directions is illustrated in fig. 8.9 below for case '30', which was found to have the highest degree of curvature along the sides of the crown outline. Papilla landmarks are shown with starting positions at the upper and lower ends of their range of simulated variation. The extent of horizontal variation remaining at the final semi-landmark positions will clearly be greater the more variation there is in the starting positions of the papilla landmarks along the outline (and so the greater the angle between the different directions) and when the final positions of the semi-landmarks are towards the upper (cervical) end of their ranges. Recall also that by definition the chord directions will always slope inwards from the MD endpoints to the papilla landmarks, rather than outwards since the MD endpoints define the maximal width of the tooth approximately perpendicular to the LACC or LAAC.

Fig. 8.9: Differences in chord directions and potential semi-landmark positions, depending on starting locations of the papilla landmarks



However, for the majority of buccal surfaces of the upper central incisor, the sides of the teeth are relatively straight and so the use of chord directions is not expected to lead to this being too great a problem for most cases.

8.2.6 Sample sizes

From the various sources described above it was possible to obtain 32 images of different cases satisfying the inclusion criteria used in the reliability study of chapter 4. From each extracted outline, 50 configurations were generated, each comprising the same non-gingival landmarks at the MD endpoints, incisal corners and incisal endpoint of the LACC, and a set of simulated gingival landmark positions.

The number of cases exceeds the 19 used in the reliability study and so should adequately encompass a variety of different possible patterns in the non-gingival landmarks, which were found to play an important role in the determination of new semi-landmark positions with each criterion used. For each case it was possible to produce any number of simulated sets of gingival landmarks representing different possible positions for the LACC point on the gingival margin and papilla landmarks. However, after some preliminary investigation it was decided that 50 simulations per case would adequately produce a wide variety of simulations across the full ranges of possible positions for each landmark specified in 8.2.4, without slowing down the performance of the computer used too greatly. (With 32×50=1600 configurations, the slowest method, using the Full Procrustes criterion, took just over 24 hours to perform 2 full iterations using a machine with a 500MHz Celeron processor and 128Mb RAM).

Fig. 8.10 (top) shows 50 simulated sets of possible gingival landmark positions, based on the outlines, non-gingival landmarks and cemento-enamel junction markers of four of the 32 cases. Fig. 8.10 (bottom) shows the Procrustes fits of the configurations, when registered to the overall mean of all 32×50 simulated configurations.

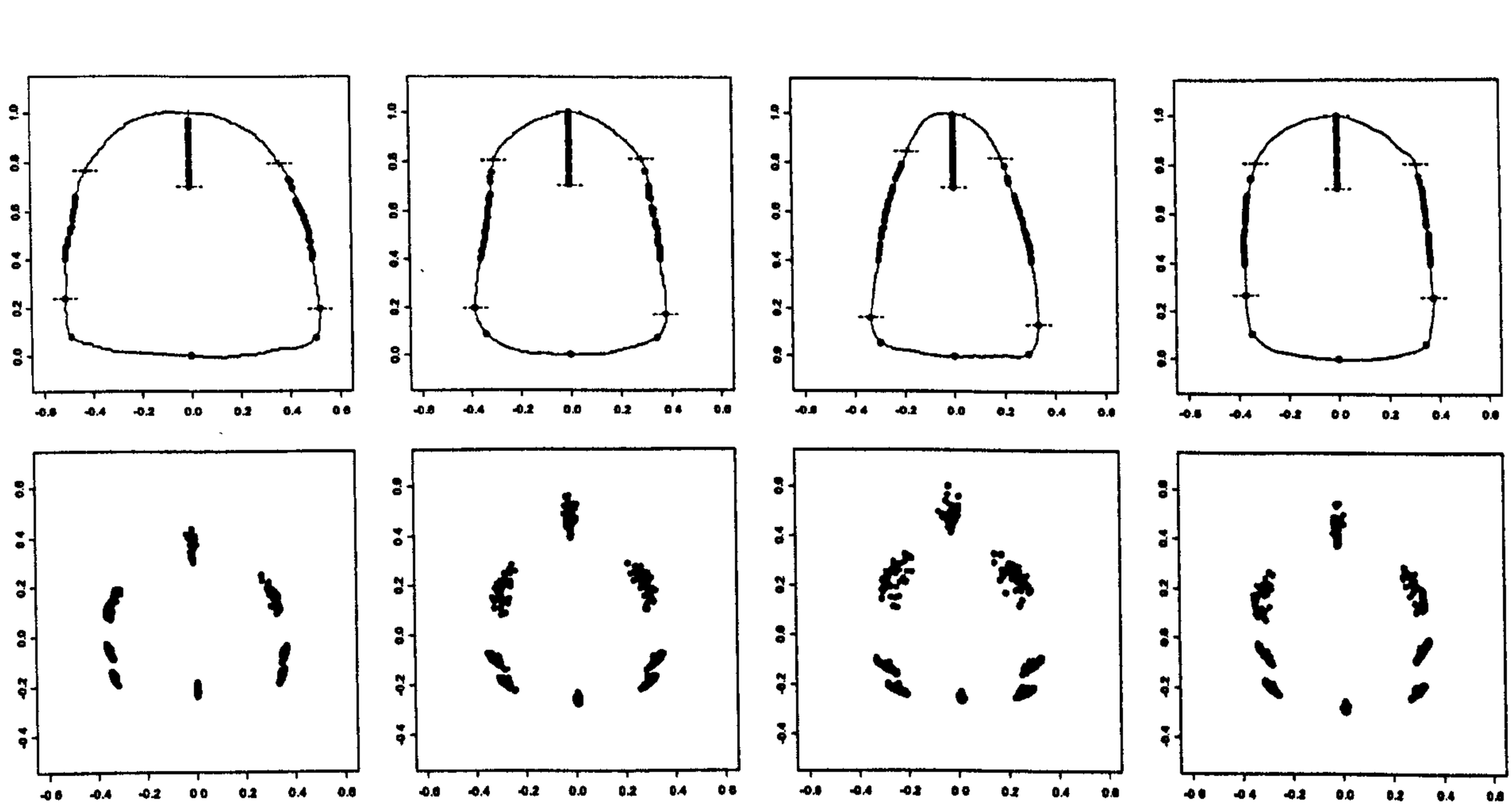


Fig. 8.10: (Top) 50 simulated sets of gingival landmarks positions on each of 4 cases. (Dashed lines) Upper and lower limits of simulation ranges. (Plusses) CEJ 'markers'. Landmarks at ends of MD width, incisal corners and incisal endpoint of LACC same for each simulated configuration. (Bottom): Original Procrustes fits of the simulated configurations, to overall mean of all simulations.

8.3 Results

8.3.1 Overall summary measures and convergence

Table 8.1 below details the iteration history from running each semi-landmark method until the change in RSS, denoted as ΔRSS, slows to a stable level, upon entering it's 'slow manifold' or 'uninformative regime', as described in sections 5.5.5 and 7.4. (Here $RSS = RSS(\hat{\mu})$ since GPA method (ii) is used, where $\hat{\mu} = \overline{X}^P$.) The iteration on which this first occurs is indicated by a shaded box, which may be identified (say) by considering the values of ΔRSS as a scree plot in the same way as the number of PC's is chosen in a principal components analysis. The level of convergence reached on this iteration can then be checked to see if this is acceptable, based on the recommendations made in section 5.5.6. If so, then we can then re-run each semi-landmark process at a pre-specified level of convergence or number of iterations to achieve these values, without the danger of the configurations becoming unrealistic in shape from running the routine for too long.

Upper central incisor (Gingival landmarks as semi-landmarks)						
Iteration	Bending energy		Full Procrustes distance		Nearest point	
	RSS	ΔRSS	RSS	ΔRSS	RSS	ΔRSS
Original	25.8535		25.8535		25.8535	
1	23.0209	2.8326	4.3370	21.5165	8.7424	17.1111
2	**23.0399**	**-0.0190**	**4.3321**	**0.0048**	5.7529	2.9895
3	22.9478	0.0921	4.3273	0.0048	4.7940	0.9590
4	22.8548	0.0931	4.3225	0.0048	4.4821	0.3118
5	22.7612	0.0936	4.3176	0.0049	4.3826	0.0996
6	22.6672	0.0940	:	:	4.3512	0.0314
7	22.5729	0.0943	:	:	**4.3413**	**0.0099**
8	22.4783	0.0946	:	:	4.3381	0.0032
9	22.3834	0.0949	:	:	4.3369	0.0011
10	22.2883	0.0951	:	:	4.3365	0.0005
11	22.1931	0.0952	:	:	4.3362	0.0003
12	:	:	:	:	4.3360	0.0002
13	:	:	:	:	4.3358	0.0002
14	:	:	:	:	4.3356	0.0002

Table 8.1: Iteration history of RSS and ΔRSS values for the different semi-landmark methods.

Based on the discussion of section 5.5.6, a convergence criterion for ΔRSS when running each semi-landmark method on 1600 configurations, corresponding to the suggested value of at most 2×10^{-5} for the change in mean RSS would be in the region of 0.03 (i.e. $\Delta RSS \leq 1600\times0.00002$). As can been seen from table 8.1, this means that each method converges to an acceptable level before entering its 'slow manifold'.

(ΔRSS<0.03 first occurs at the values in bold, coinciding with the boxes shaded grey). In general, if each method had been considered in isolation, then the obvious choice would be to use the convergence level shaded in grey (providing the level of convergence is acceptable). However, in this study, so that the results are comparable, we would want to use the same convergence level for each of the semi-landmark methods, this being the strictest possible common value below the suggested 0.03 level for which each procedure still avoids entering its 'uninformative regime'.

Table 8.1 shows that the Full Procrustes (FP) criterion and bending energy (BE) method both took fewer iterations to reach ΔRSS<0.03 than the nearest point (NP) method. Notice however that ΔRSS for the BE method is negative at convergence and would begin to increase if the process were run for any longer. Recall from 5.5.5 that Rohlf (personal communication) reported this happening with certain examples he had tried, when developing the 'TPSrelw' program. When the data here is used in 'TPSrelw', the program simply stops upon finding a negative ΔRSS and the process is declared as converged. It does however seem reasonable to assume that since the increase in *RSS* occurs below the chosen convergence level, this has a negligible effect on the configurations and so is deemed small enough to be of no concern.

In terms of processing time, the FP criterion took considerably longer than the other two methods, despite the NP method requiring more iterations. This is because the BE and NP method have explicit solutions for the determination of the semi-landmark positions during each iteration, whereas the FP criterion requires iterative optimisation using the 'nlmin' function. However, even if 'nlmin' is used for the BE and NP methods, the FP method was still found to take considerably longer.

Fig. 8.11 plots the Procrustes fits about their estimated grand mean shape for the original simulated data (with all landmarks fixed) and after using each of the semi-landmark methods. Table 8.2 presents the chosen summary measures from 8.2.5 for each of the methods, i.e. the values of $EucSS(cases), EucSS(within)$ and $EucSS(total)$ based on the final Procrustes fits $X_{im}^{P_{SL}}$, i=1,...,32, m=1,...,50, and their estimated mean shape $\hat{\mu}_{SL}=\overline{X}_{im}^{P_{SL}}$. For comparison, the first column reports these figures for the original (initial) Procrustes registration of the data.

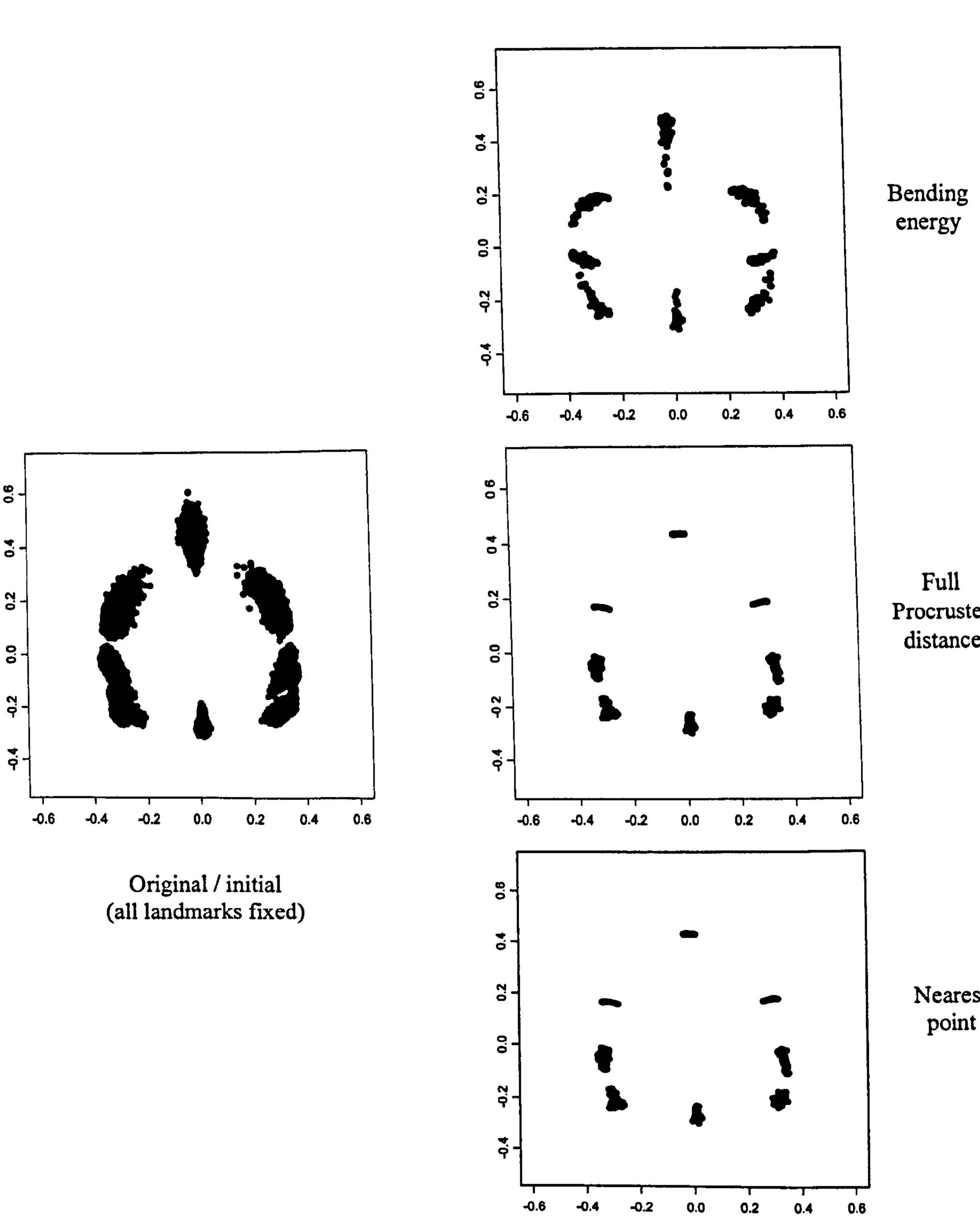


Fig 8.11: (Left) Initial Procrustes registration or the 32×50 configurations (all landmarks fixed). (Right) Procrustes registration after running each of the new semi-landmark methods.

	Semi-landmark criterion			
	None (all fixed)	Bending energy	Full Proc. distance	Nearest point
EucSS (cases)	10.06	22.943	4.241	4.251
EucSS (within)	15.79	0.097	0.091	0.090
EucSS (total)=*RSS*	25.85	23.040	4.332	4.341
Overall reliability	0.38	0.996	0.979	0.979

Table 8.2: Summary measures based on final Procrustes fits produced by each of the different semi-landmark methods.

Recall that since the $X_{im}^{P_{SL}}$ are obtained by GPA method (ii), then:

$$EucSS(total) = EucSS(cases) + EucSS(within)$$

where each term is as defined in section 8.2.5.2. The final row of table 8.2 reports the overall reliability figures for each set of final Procrustes fits, based on these Euclidean sums of squares. These are included to illustrate why, in this application, an improvement to the reliability figures is not a good indication of how successful a method has been. From equations (4.16) and (4.17) in 4.4.2.3, for the one-way case, these are calculated as:

$$\frac{\frac{1}{50}\left(\frac{1}{32-1}EucSS(cases)-\frac{1}{50\times32-32}EucSS(within)\right)}{\frac{1}{50}\left(\frac{1}{32-1}EucSS(cases)-\frac{1}{50\times32-32}EucSS(within)\right)+\frac{1}{50\times32-32}EucSS(within)}.$$

8.3.2 Bending energy method

For a successful outcome we would hope that any variation between the 32×50 configurations in the positions of the gingival landmarks would be reduced or eliminated. After using the BE method however, fig. 8.11 indicates that there is still variation in the Procrustes fits at the three gingival landmark positions. In fact, the occlusal-gingival (vertical) variation in the landmark on the gingival margin at the top of the LACC actually appears to have increased slightly.

The values in table 8.2, suggest that despite only a small reduction in the total variation in shape, the variation within cases has been reduced to almost zero, satisfying one of the goals we set out to achieve. Since the fixed, non-gingival landmarks were simulated at the same positions on each case, we can conclude that any variation in the position of the gum within each set of simulations has been almost eliminated, with representations of the same tooth ending up as the same shape, as required (we explain why in section 8.3.2.1). However, as can be seen from the first row of table 8.2, with the BE method this appears to have been at the expense of a two-fold increase in the amount of variation between cases.

An examination of the Procrustes fits of each set of 50 simulations from each case, confirms that the 'within case' variation in shape has been clearly been reduced. Fig. 8.12 shows the Procrustes fits of the 50 simulations of each of six selected cases, registered to the overall mean of the 32×50 configurations, both before (first column) and after (second column) running the semi-landmark method. In the second column, each set of representations of each of the teeth have all ended up as almost the same shape, as we would hope for. The only apparent variation remaining is in the final horizontal positions of the papilla landmarks, which results from differences in the directions of the escribed chords, determined by their original starting positions (see section 8.2.5.2). The largest extent to which this was observed was on those cases where the sides of the teeth have high curvature and where the final positions of the semi-landmarks are towards the upper (cervical) end of the ranges they were simulated between. These cases are those in rows 1, 3 and 4 of fig. 8.12.

Notice however that that the variation 'between cases' is now more notable. The reduction in variation 'within cases' has been achieved in many cases by each set of simulations becoming the same infeasible shape, as the gingival landmarks adopt positions regarded as 'unrealistic', when referred back to the original tooth outline they are supposed to represent.

In the third column of fig. 8.12, each set of Procrustes fits has been superimposed back into the space of the outline of the cases from which they were simulated. This uses a Bookstein-type transformation of the form of equation (2.6), to transform the MD endpoints of each configuration (and hence each of the other non-gingival landmarks) back to the original positions used for the simulation. The final positions of each configuration's gingival landmarks are now shown in relation to the CEJ and sides of the teeth they were simulated along. The dashed lines on each outline denote the cervical (upper) and incisal (lower) limits of variation that each set of landmark positions were simulated between, representing the feasible positions for gingival landmarks. Recall that the cervical bounds occur at the positions of the CEJ (here indicated as plusses, as in fig. 8.7). The incisal bounds are (for the gingival margin landmark at the top of the LACC), at a position 30% of the way along the LACC from the CEJ and (for the papilla landmarks) at the positions of the MD landmarks, since by definition, these cannot be located below them.

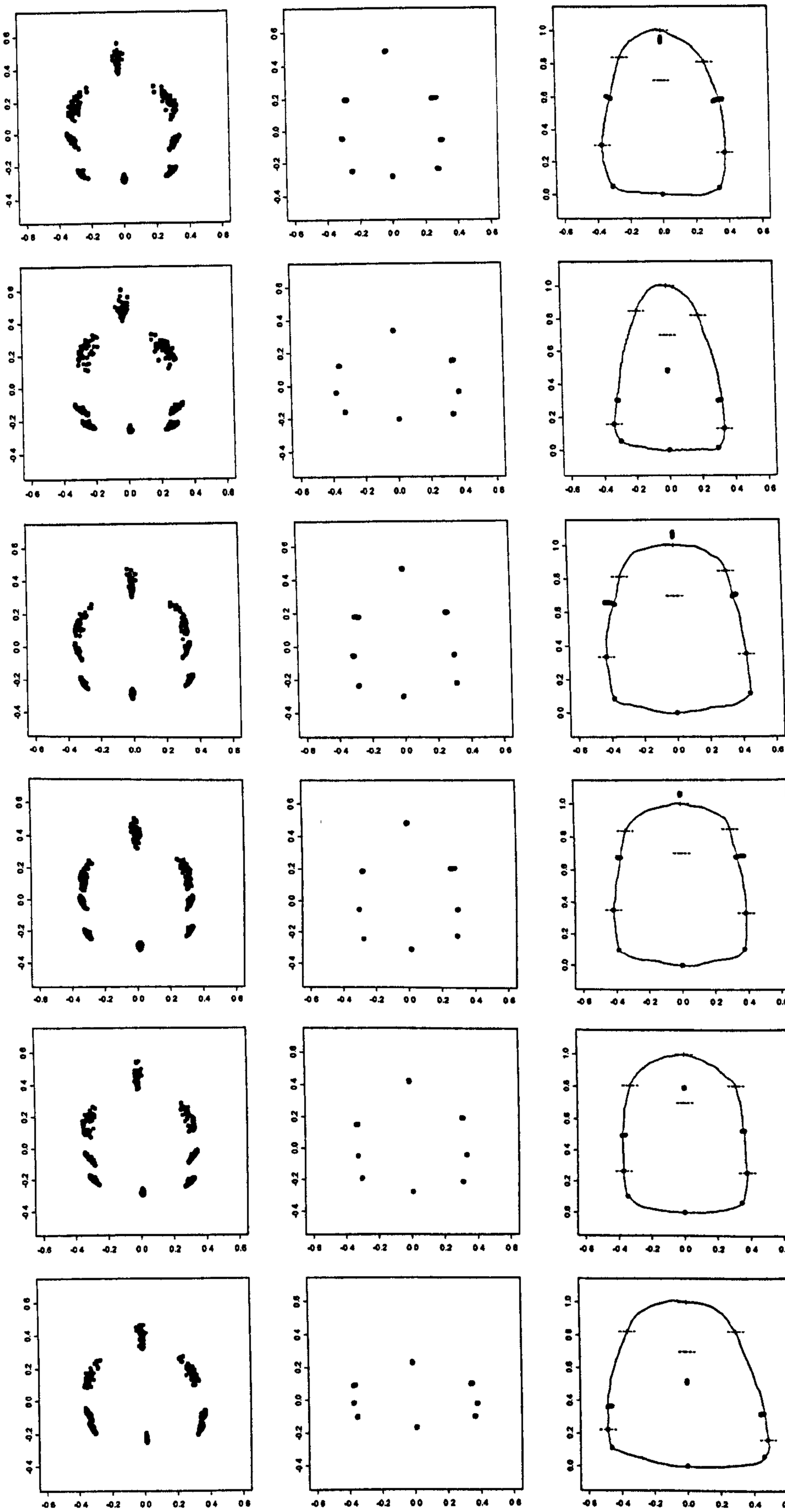


Fig 8.12: Procrustes fits for the 50 simulations of each of 6 cases (registered to overall mean of the 32×50 configurations). (First column) After initial GPA. (Second column) Final fits after using Semi-landmark method with BE criterion. (Last column) Edge superimposition of final fits (to match MD endpoints) back onto original tooth outline. Dashed lines denote the cervical (upper) and incisal (lower) ranges that each set of gingival landmarks were simulated between

For the six example cases in fig. 8.12, the first (row 1) and fifth (row 5) sets of configurations have shapes with gingival landmark positions well within the ranges of acceptability indicated by the dashed lines. For the cases in the second, third, fourth and final rows however, each of the final configurations representing these cases have gingival-margin landmarks (at the top of the LACC) outside the ranges of what would be regarded as feasible positions on these teeth. In particular, the final configurations of the third and fourth rows have LACC endpoints that would not actually lie on the crown of the teeth, whereas the cases in the second and final rows of fig. 8.12 have gingival margin/LACC endpoints that would be highly unlikely in practice, given the proximity of the position of the CEJ.

For the remaining sets of simulations from the rest of the 32 cases, the outcome was very similar, with all representations of each case converging to almost exactly the same shape. In exactly half the cases (16 out of 32), these configurations had final gingival landmark positions within the acceptable ranges of positions when referred back to the original crown outlines they were simulated from. In 8 of the cases the final gingival margin/LACC landmarks were found to be in a position that would be cervical to (above) the CEJ of the crown outline being represented and in the remaining 8 cases, this landmark was always in a position that would be incisal to (below) the lower bound of feasible positions, given the position of the CEJ. In section 8.3.2.2 we investigate when and why all representations of a particular case either 'collapse' or 'stretch' to an unrealistic shape outside the original ranges of simulation or converge to a 'feasible' shape within them.

8.3.2.1 Why representations of the same case end up the same shape

Figs 8.13 and 8.14 show the first iteration of the semi-landmark method for two representations of case 8 (previously in fig. 8.12, second row) and of case 23 (fig. 8.12, fourth row), when considered as part of the entire 32×50 configuration sample. Only the first iteration is shown since when using the BE criterion, the main movement of the semi-landmarks occurred on the first iteration, with convergence at the end of the second, after only minor further changes to their positions. For illustration, the two representations, *A* and *B*, of each case are such that the three gingival landmarks are at their cervical (upper) bounds of simulation or their incisal (lower) bound, i.e. for

representation A, the gingival landmarks are at the positions of the markers along the CEJ and for representation B, the papilla landmarks on the outline of the crown and the gingival margin landmark at the top of the LACC are at positions 40% of the way along the LACC and 70% of the way along the LAAC from the incisal edge, respectively.

Following the usual notation, the first and fourth panels in the top row of both figs 8.13 and 8.14 display the initial registration of A and B to the initial estimate of mean shape $\hat{\mu}_{(0)}$ obtained from all 32×50 configurations. This gives the first set of 'nominal' positions of each set of landmarks, denoted for each configuration as $Y^0 = X^P_{(0)}$. Panels 2 and 5 then show the movement of the semi-landmarks to their new positions (with respect to $T = \hat{\mu}_{(0)}$) to produce configurations $Y^{new} = X^{new}_{(1)}$. The third and sixth panels then show the Procrustes registration of each $X^{new}_{(1)}$ to the new updated mean shape $\hat{\mu}_{(1)}$, again obtained from all 32×50 configurations, at the end of the iteration, to give a new Procrustes fit $X^P_{(1)}$. In the second row, PTPS transformations between each pair of configurations in the panel above are displayed as deformation grids, mapping the space from $\hat{\mu}_{(0)}$ to $X^P_{(0)}$ (panels 1 & 4), $\hat{\mu}_{(0)}$ to $X^{new}_{(1)}$ (panels 2 and 5) and $\hat{\mu}_{(1)}$ to $X^P_{(1)}$ (panels 3 & 6). On each grid, the 'target' configuration $X^P_{(0)}$, $X^{new}_{(0)}$ or $X^P_{(1)}$ is displayed as filled circles. The third and fourth rows display the affine and non-affine components of the PTPS transformations as deformation grids, as described in section 5.3.3. i.e. for Q, S, Γ^{21} and Γ^{11} as defined in (5.31) and (5.32), these represent the mapping of the space from $\hat{\mu}_{(0)}$ to $Q\Gamma^{21}X^P_{(0)}$ or $Q\Gamma^{21}X^{new}_{(1)}$ (3rd row, panels 1 and 4 or 2 and 5) or $\hat{\mu}_{(1)}$ to $Q\Gamma^{21}X^P_{(1)}$ (3rd row, panels 3 and 6), $Q\Gamma^{21}X^P_{(0)}$ to $X^P_{(0)}$ or $Q\Gamma^{21}X^{new}_{(1)}$ to $X^{new}_{(1)}$ (4th row, panels 1 and 4 or 2 and 5), and $Q\Gamma^{21}X^P_{(1)}$ to $X^P_{(1)}$ (4th row, panels 3 and 6). In the third row, the affine fits of $\hat{\mu}_{(0)}$ to $X^P_{(0)}$, $\hat{\mu}_{(0)}$ to $X^{new}_{(1)}$ and $\hat{\mu}_{(1)}$ to $X^P_{(1)}$, are displayed as diagonal crosses, with the 'target' configuration $X^P_{(0)}$, $X^{new}_{(1)}$ and $X^P_{(1)}$ displayed on the grid as filled circles, so as to show the size of the residuals (which form the remaining non-affine (bending) parts of each PTPS transformation, represented in panel below).

Fig 8.14: First iteration of semi-landmark method for representations *A* and *B* of case 8 (filled circles, solid lines), with gingival landmarks at upper & lower simulation limits. BE criteria used to determine new semi-landmark positions. All 32×50 configurations used to compute mean (plusses, dashed lines). (Empty circles) Nominal semi-landmark positions. (Dotted lines) directions of escribed chords. (Crosses) Fitted values given by affine component of each PTPS transformation.

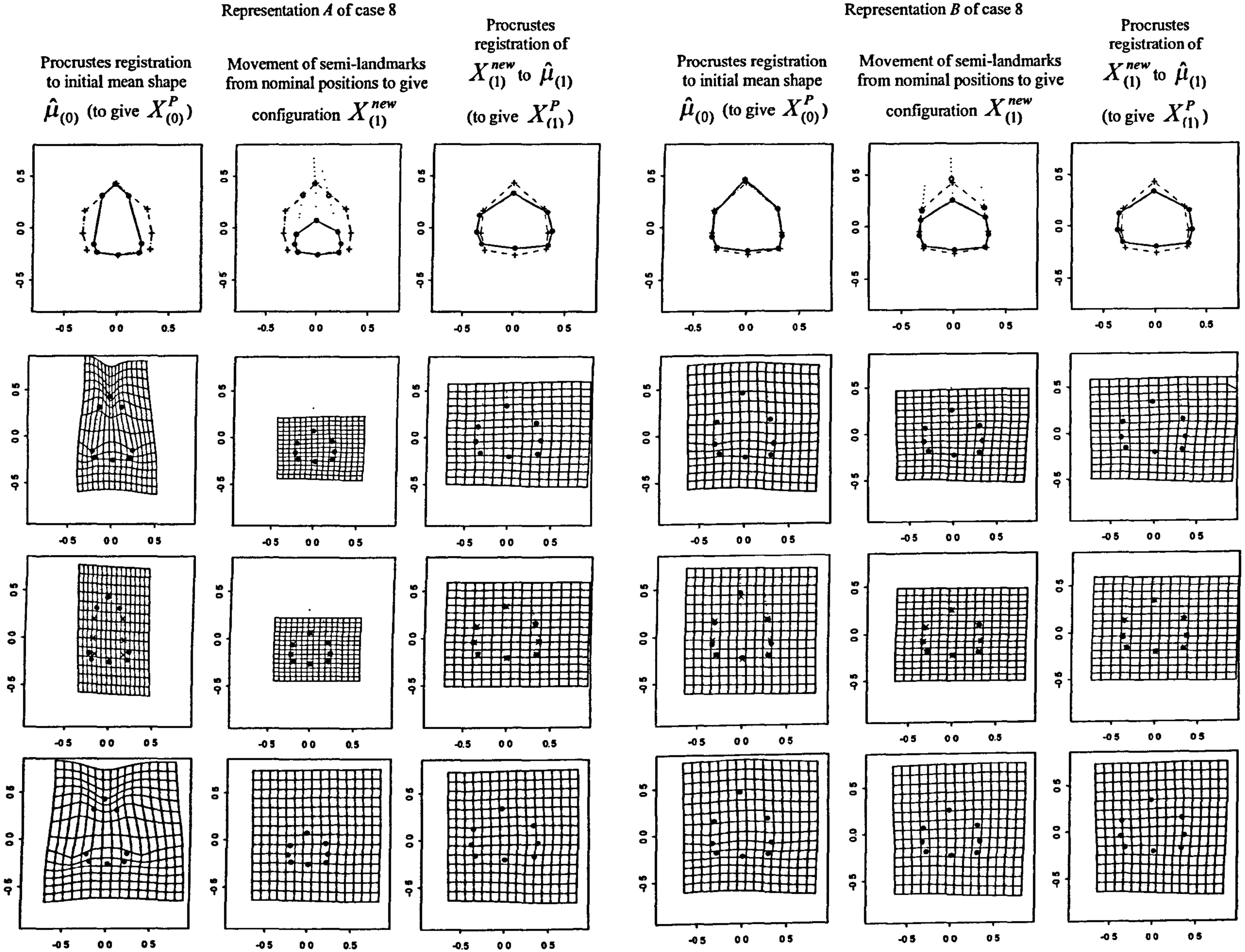
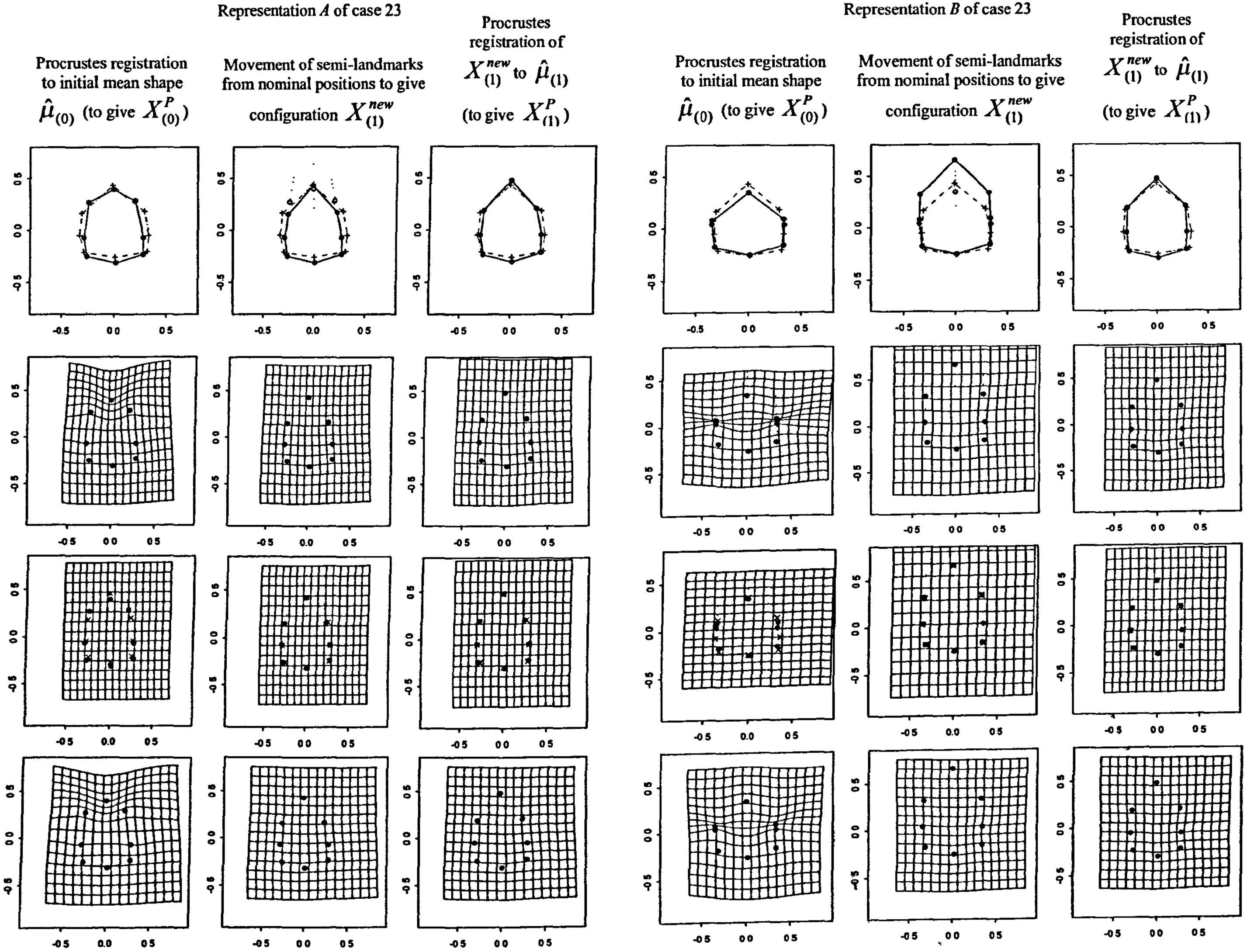


Fig 8.13: First iteration of semi-landmark method for representations *A* and *B* of case 23 (filled circles, solid lines), with gingival landmarks at upper & lower simulation limits. BE criteria used to determine new semi-landmark positions. All 32×50 configurations used to compute mean (plusses, dashed lines). (Empty circles) Nominal semi-landmark positions. (Dotted lines) directions of escribed chords. (Crosses) Fitted values given by affine component of each PTPS transformation.

Deformation grid for PTPS transformation from $\hat{\mu}_{(0)}$ to (L-R): $X^{P}_{(0)}$, $X^{new}_{(1)}$ & $\hat{\mu}_{(1)}$ to $X^{P}_{(1)}$

Deformation grid for affine component of PTPS transformation from $\hat{\mu}_{(0)}$ to affine fit to (L to R): $X^{P}_{(0)}$, $X^{new}_{(1)}$ & $\hat{\mu}_{(1)}$ to affine fit to $X^{P}_{(1)}$

Deformation grid for non-affine component of PTPS transformation. i.e. from affine fit to (L to R): $X^{P}_{(0)}$, $X^{new}_{(1)}$ & $X^{P}_{(1)}$



Notice that the size of the deformation grids from $\hat{\mu}_{(0)}$ to $X_{(1)}^{new}$ in columns 2 and 5 is affected by the size of $X_{(1)}^{new}$, particularly when the new positions of the semi-landmarks result in a much smaller/larger configuration than $X_{(0)}^{P}$. In order to standardise the grids for easier visual comparison, we could re-superimpose each $X_{(1)}^{new}$ back to $\hat{\mu}_{(0)}$ and display the transformation grids of the map from $\hat{\mu}_{(0)}$, as these would have the same appearance as those from $\hat{\mu}_{(0)}$ to $X_{(1)}^{new}$. However, since the mean shape hardly changes between iterations (from $\hat{\mu}_{(0)}$ to $\hat{\mu}_{(1)}$), the displays in columns 3 and 6 essentially serve this purpose and to talk about re-registering new configurations $X_{(1)}^{new}$ back to the old mean $\hat{\mu}_{(0)}$ seems somewhat misleading.

As described in section 5.6.3, the BE is minimised over λ by the semi-landmarks adopting positions which minimise the generalised residual sum of squared distances between the 'affine' mapping of $\hat{\mu}_{(0)}$, given by $Q\Gamma^{21}X_{(1)}^{new} = Q\Gamma^{21}vec^{-1}(vec(X_{(0)}^{P}) - U\lambda)$ and $X_{(1)}^{new} = vec^{-1}(vec(X_{(0)}^{P}) - U\lambda)$, given by equation (5.62) as:

$$tr(X_{(1)}^{new} - Q\Gamma^{21}X_{(1)}^{new})^{T} S^{-1}(X_{(1)}^{new} - Q\Gamma^{21}X_{(1)}^{new})^{T} = \text{BE}, \qquad (8.3)$$

where S is as defined in (5.31), here for configuration $\hat{\mu}_{(0)}$. The minimised non-affine displacements of the PTPS mapping, i.e. residuals of the GLS affine fit, then ensure that the landmarks in $\hat{\mu}_{(0)}$ are mapped exactly to the fixed landmarks and positions along the semi-landmarks chords in $X_{(1)}^{new}$.

In both figs 8.13 and 8.14, the two representations of each case with semi-landmarks at different starting positions end up almost the same shape after the first iteration. Comparing the third and sixth panels in the first row of both figures, it appears that the only difference between the new versions of representations A and B of each case, is in the new horizontal positions of the papilla semi-landmarks.

Regardless of the PTPS transformation from $T = \hat{\mu}_{(0)}$ to $Y^{0} = X_{(0)}^{P}$, shown in the first and fourth panels of the second row, the PTPS transformations with least BE from $T = \hat{\mu}_{(0)}$ to

$Y^{new} = vec^{-1}(vec(X^{P}_{(0)}) - U\lambda) = X^{new}_{(1)}$ in the second and fifth columns, are nearly identical for *A* and *B* (ignoring the effects of location, orientation and size, as discussed above), including their affine and non-affine components. Since the fixed landmarks are in the same arrangement in both representations of the same case and the chord directions along which the semi-landmarks may move are in approximately in the same directions and locations, then regardless of the nominal starting positions of the semi-landmarks, new configurations result with almost identical shape. The slight difference in the resulting $X^{new}_{(1)}$ for *A* and *B* is from the differences in the locations and directions of the chords used on each target configuration, determined by the starting positions of the semi-landmarks, as depicted in fig. 8.9. (Here they were chosen at the opposite ends of their possible simulation ranges so as to maximise this effect). Close examination of the deformation grids in rows three and four of figs 8.13 and 8.14, shows how this affects the optimal PTPS mappings and their components (panels 2 and 5) for configurations *A* and *B*. For example, in fig. 8.13, the deformation grid representing the non-affine component of the mapping from $\hat{\mu}_{(0)}$ to the new version of representation *A* indicates that the horizontal component of the global affine needs to be slightly different at the new semi-landmark positions from the rest of the configuration, because the papilla chords of *A* slope inwards more as the semi-landmarks move cervically (upwards) in accordance with the vertical affine mapping determined by the fixed landmarks. However, less bending (variation in the shape of the non-affine grid) is required for configuration *B* where the affine component is better able to simultaneously fit the more vertical chord positions and fixed landmarks.

8.3.2.2 When and why unrealistic shapes are produced: An investigation using discriminant analysis

Having considered why representations of the same case end up as the same shape, we now examine why these configurations all either 'collapse' or 'stretch' to the same unrealistic shape or result in a configuration which is still regarded as 'feasible'. The outcome for each set of simulations from each case can be categorised into one of three such groups, with 8 'collapsing', 16 'feasible' and 8 'stretching'. We can then investigate the dependency of these outcomes on the shape of the actual crown from which each set of 50 simulations were produced, using the three CEJ markers as landmarks on each of the 32 teeth, rather than possible gum positions.

Let $X^P_{i_{(l_g)}}$ denote the Procrustes fit of the ith, i=1,...,32, actual crown, each of which falls into one of three sublists l_g, g=1,2,3, where l_1 comprises the 'collapsing' cases (with $i_{(l_1)}$=1,2,5,8,...), l_2 comprises the cases which produced 'feasible' tooth shapes ($i_{(l_2)}$=3,6,7,...) and l_3 those which 'stretched' ($i_{(l_3)}$=4,13,16,...). The number of cases in each sub-list will be denoted by n_g, with n_1=8, n_2=16, n_3=8. The mean shapes and Procrustes fits of the three groups obtained from an ordinary GPA registration of the 32 cases to their overall mean, are shown in fig. 8.15 below. Each mean shape $\overline{X}^P_{l_1}$, $\overline{X}^P_{l_2}$ and $\overline{X}^P_{l_3}$ is calculated as the arithmetic mean of the Procrustes fits belonging to each group, as is usual in tangent space methods for comparing groups (see section 3.4.2). As can be seen in fig. 8.15, this exploratory analysis suggests that on average, cases where each set of simulations 'collapse' when used with the BE method, have MD endpoints located relatively closer to the incisal edge, compared to those which 'stretch'. However, a more formal investigation was performed to investigate the differences between groups in more detail.

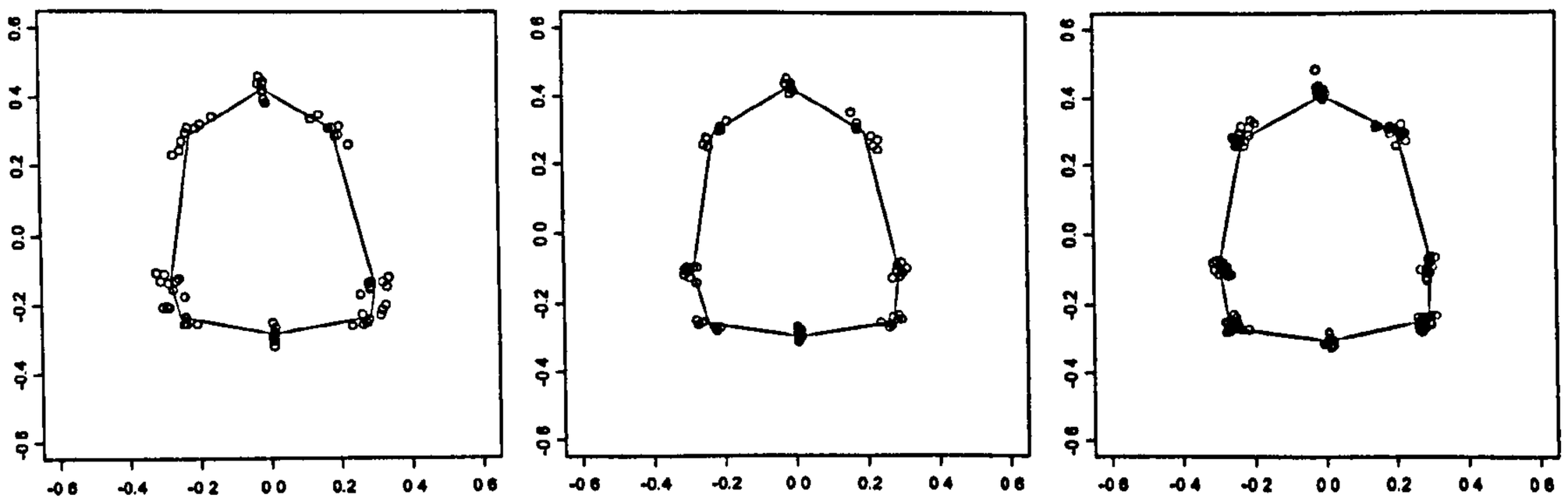


Fig. 8.15: (Left to right) Arithmetic mean shapes and Procrustes fits for the 'collapsing', 'feasible' and 'stretching' crown outlines, using the CEJ markers as landmarks at the cervical (top) end of each crown. Configurations are registered to the overall mean shape of the 32 cases and not their group means.

As was seen in chapters 3 and 4, the variation in shape in a set of Procrustes fits, can be parameterised by calculation of principal components. Tests of differences between the mean component scores along each PC can then be made using one-way ANOVA, using contrasts to investigate the possibility of a trend in mean scores from the 'collapsing' to the 'stretching' group. Along all but one of the 13 non-zero PC's for these configurations, no evidence of a difference or trend was found between the mean

component scores (p values were always above 0.7). However, for the component scores along the second PC, a significant difference and evidence of a trend across the groups was found (p<0.00001). The PC in question is plotted in fig. 8.16 by displaying configurations (icons) corresponding to standardised scores either side of the mean shape as described in section 3.4.2.2. This can be seen to contrast the occlusal-cervical (vertical) positions of the MD endpoints with the landmarks along the incisal edge. 'Collapsing cases' were found to have lower mean PC scores and therefore by fig. 8.16, have MD endpoints and incisal corners/edge landmarks located closer together along these directions. Conversely, the 'stretching' cases, had higher PC scores and therefore MD endpoints and incisal corners further apart along these directions.

UNIVERSITY
OF SHEFFIEL
LIBRARY

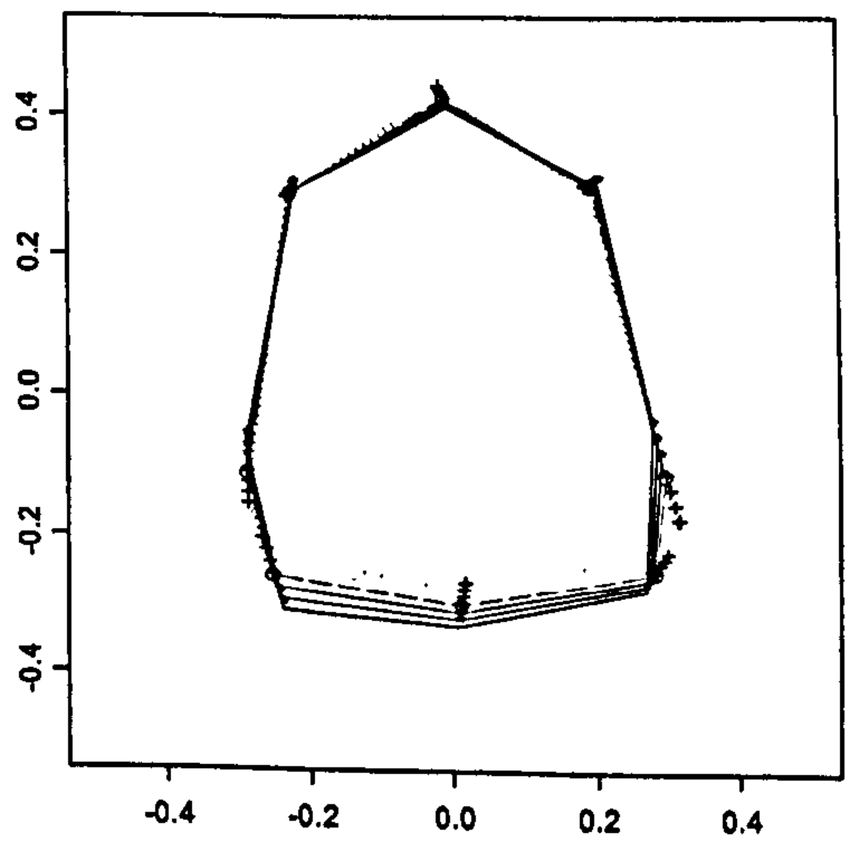


Fig. 8.16: (Dotted lines, crosses): Shapes at -3,-2,-1 standard deviations along the 2nd PC of variation of the 32 actual crown configurations. (Solid lines, stars): Shapes at +1,+2,+3 s.d's along the 2nd PC. (Dashed line, empty circles): Procrustes mean shape.

Note however, that the difference between the three groups has only been related to one particular projection of the Procrustes fits, obtained by considering transformations to orthogonal dimensions which describe successively smaller proportions of variation and ignoring any information on group membership. Instead, calculation of Fisher's linear discriminant function allows us to find the projection of the data that best highlights the differences between the groups. Let $v_{i_{(l_g)}}$ denote a suitable choice of tangent coordinates. Here we have used $v_{i_{(l_g)}} = vec(X^P_{i_{(l_g)}})$, the vectorised Procrustes fits. If W and B are the 'within group' and 'between group' covariance matrices of the $v_{i_{(l_g)}}$, with

$$W=\frac{1}{n-1}\sum_{g=1}^{3}(v_{i_{(l_g)}}-\bar{v}_{l_g})(v_{i_{(l_g)}}-\bar{v}_{l_g})^T \text{ and } B=\frac{1}{n-3}\sum_{g=1}^{3}n_g(\bar{v}_{l_g}-\bar{v})(\bar{v}_{l_g}-\bar{v})^T,$$

we project the data onto vector a (by $a^T v_{i_{(l_g)}}$) where a maximises:

$$F=\frac{a^T Ba}{a^T Wa}, \tag{8.4}$$

the ratio of the between to within groups sums of squares (i.e. the F statistic of a one-way ANOVA after projection of the data by a). Since F is invariant to scalar multiplication, a is constrained so that $a^T Wa=1$.

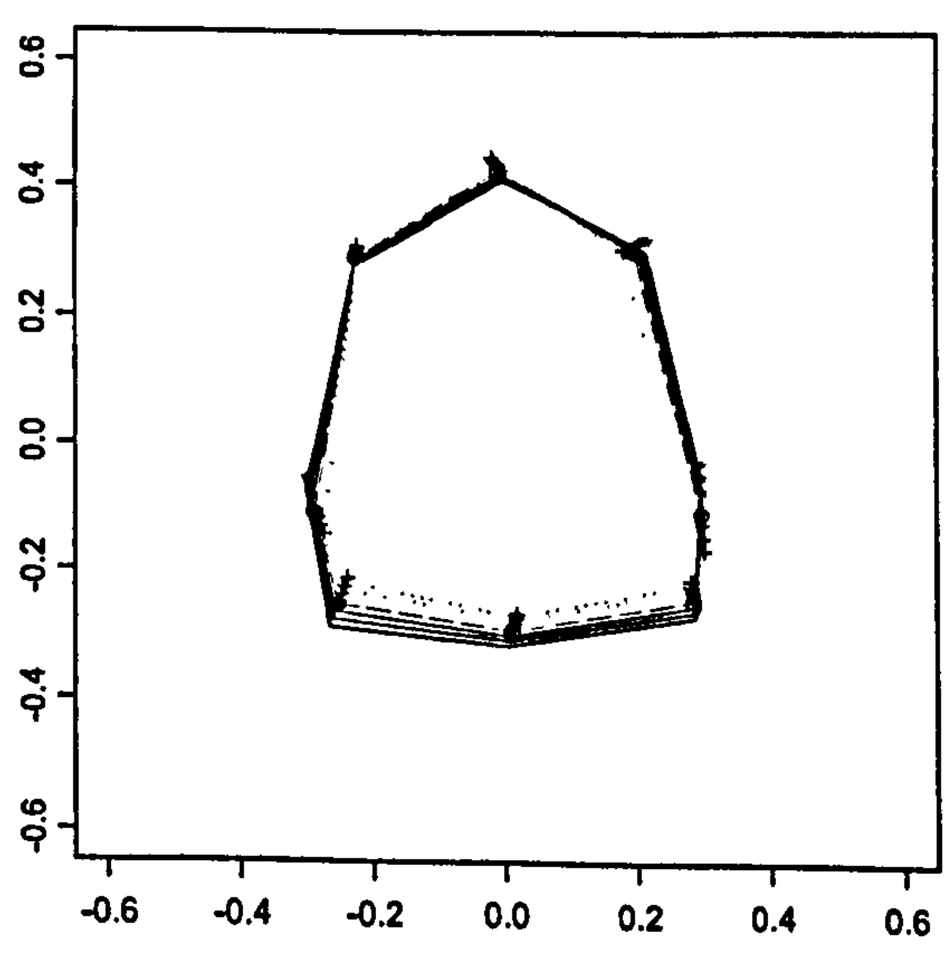
Following, for example, Mardia *et al.* (1979), a is given by the eigenvector corresponding to the largest eigenvalue λ_1, of $W^{-1}B$. However since we have used the Procrustes fits as our choice of tangent coordinates, W is singular and so the Moore-Penrose generalised inverse must to be used to obtain W^{-1}. (An alternative would be to perform the discriminant analysis on the $2k$-3 PC scores and project the resulting discriminant function a back from PC space to configuration space). The values $a^T v_{i_{(l_g)}}$ are then the discriminant scores for each configuration and have maximal variation between groups (compared to within) over all such projections. For the three groups here, it was found that in all but 3 out of 32 cases $a^T v_{i\in i_{(l_1)}} < a^T v_{i\in i_{(l_2)}} < a^T v_{i\in i_{(l_3)}}$ indicating that the differences between the groups can be almost completely explained by the variation in scores along a.

Note that when there are 3 groups, the matrix $W^{-1}B$ has one other non-zero eigenvalue, λ_2, with corresponding eigenvector describing the second discriminant function. Although the non-orthogonality of the eigenvectors means that the amount of discrimination cannot be partitioned exactly to each function, an examination of the relative sizes of λ_1 and λ_2 indicates the relative importance of each projection of the data. In practice more than two eigenvalues result when using the Moore-Penrose inversion to obtain $W^{-1}B$, but if we perform the analysis using the 13 sets of PC scores for each case, as suggested above, then only two result. Here λ_1 was found to account for over 90% of $\lambda_1+\lambda_2$, suggesting that the projection of the data given by the first eigenvector a (of $W^{-1}B$) describes almost all of the variation between the groups. (The fact that only one set of PC scores showed any sign of significance (with $p<0.0001$) also

suggests that the difference in shape between the three groups is best described by a single projection of the data).

If standardised discriminant scores are calculated as $a^T vec(v_{i_{(l_g)}} - \bar{v}) / \sqrt{\lambda_1}$, then we can visualise the discriminant function in terms of variation in the Procrustes fits by constructing a series of plots analogous to the principal component displays in 3.4.2.2 and fig. 8.16 above. We can plot configurations (icons) corresponding to standardised scores $c\lambda_1^{1/2}a$ with c = –3,-2,-1,0,+1,+2, +3, here by calculating $vec^{-1}(c\lambda_1^{1/2}a + \bar{v})$ since the $v_{i_{(l_g)}}$ are the vectorised Procrustes fits. Fig. 8.17 displays these 'icons' for the discriminate function a determined above. This again shows that the incidence of stretching or collapsing produced by using the bending energy criteria is related to the proximity of the MD endpoints to the incisal edge of the tooth. Standardised scores for the 'collapsing' group were found to be lower, corresponding to cases where the MD endpoints and incisal corners are located closer together. Conversely, the 'stretching' cases had higher scores, corresponding to cases where the landmarks are (vertically) further apart.

Fig. 8.17: (Dotted lines, crosses): Shapes at -3,-2,-1 standard deviations along a standardised version of Fisher's linear discriminant function.
(Solid lines, stars): Shapes at +1,+2,+3 s.d.'s along the 2nd PC.
(Dashed line, empty circles): Procrustes mean shape.



The principal pattern of variation in the Procrustes fits, which maximises (8.3) and is displayed in fig. 8.17, corresponds very nearly to the variation represented by the principal component displayed in fig. 8.16. An examination of the elements of a, the largest eigenvector of $W^{-1}B$, revealed a very similar structure to the loadings of the second eigenvector (principal component) of $B+W$. The result also ties in with the observation made in 8.3.3.1 and 5.6.3, that the occurrence of collapsing or stretching,

when trying to represent gingival landmark variation by semi-landmarks and chords, is determined by the fact that the semi-landmarks adopt positions in accordance with the PTPS transformation between the fixed landmarks, in the direction of the vertical chords (and here, the fixed landmarks are always in the same arrangement on each representation of any particular case, relative to the mean shape).

8.3.3 Full Procrustes and nearest point methods

For the Full Procrustes (FP) and nearest point (NP) methods, fig. 8.11 (right) shows that in comparison to the original Procrustes registration of the configurations, both methods have been successful in reducing the variation in gingival landmark positions along the specified directions. The plots of the $X_{im}^{P_{SL}}$ for both methods appear almost identical. All remaining variation in the gingival positions is in directions perpendicular to the final chord directions, between the mesio/distal papilla landmarks and the mesio/distal endpoints of the MD width and along the direction of the LACC. In addition, the variation in the fixed landmarks, at the ends of the MD width, incisal corners and incisal endpoint of the LACC, is also reduced; we shall consider why in section 8.3.3.3 below.

Table 8.2 indicates that for both methods the total variation in shape is reduced from 25.85 down to 4.33 or 4.34 (a reduction of just over 80%). The variation between cases has been reduced from 10.06 to 4.24 or 4.25 (a reduction of 58%) and that within cases from 15.79 down to around 0.091 (a reduction of over 99%).

For each method, fig. 8.18 shows the original and final Procrustes fits of the 50 simulations from five of the 32 cases. As before the fits are registered to the overall mean of the 32×50 configurations, both before (first column) and after (second and fourth columns) running each of the semi-landmark methods. The third and fifth columns then show each set of fits superimposed back onto the outline of the case they were simulated from, with dashed lines indicating the cervical (upper) and incisal (lower) ranges of variation that were used. The simulations of cases in rows 1-4 are four of the cases previously used to illustrate the results of the BE method and are those where each set of gingival landmarks adopted 'unrealistic' positions when referred back to their original tooth outline (fig. 8.12, rows 2, 3, 4 & 6).

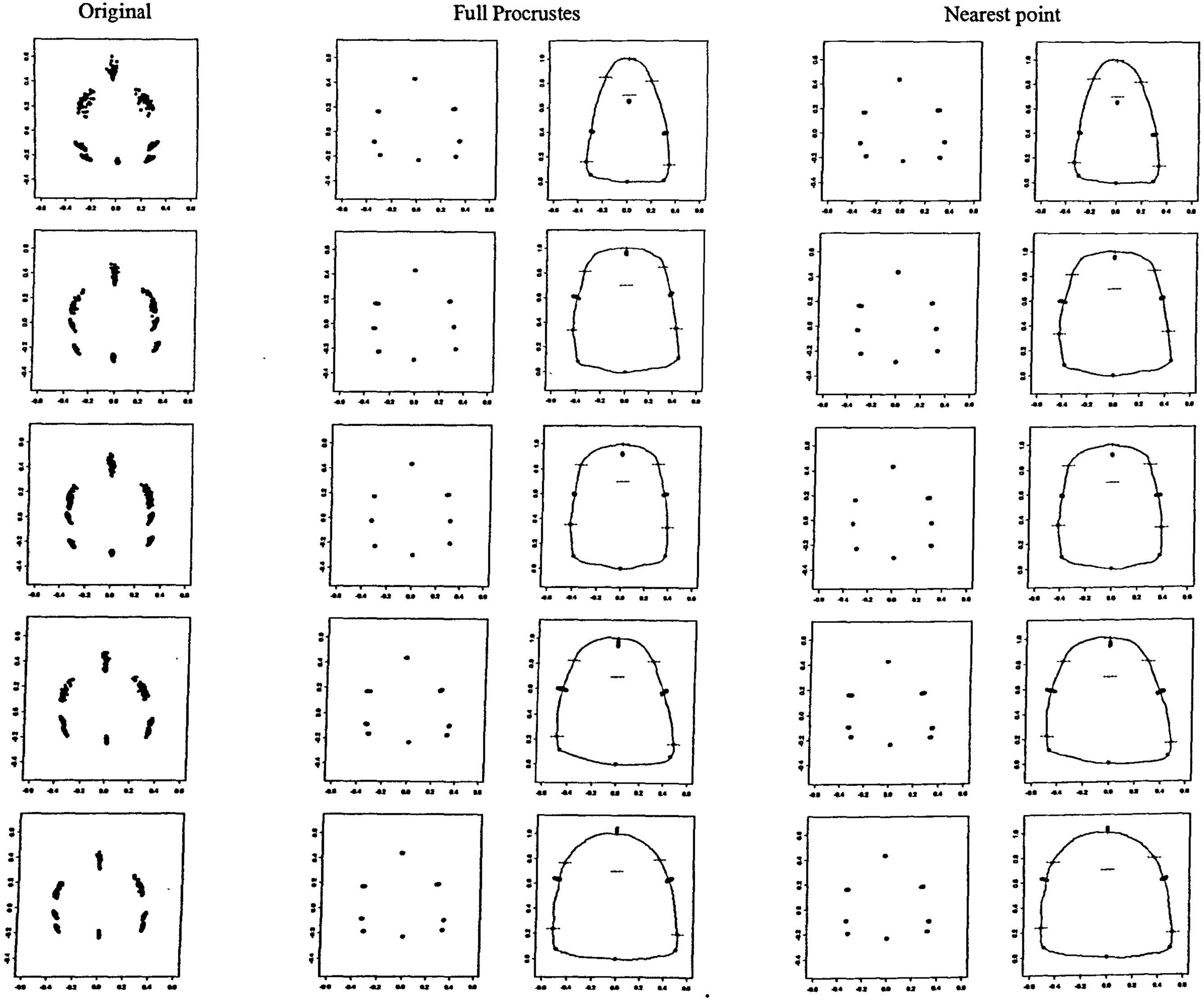


Fig 8.18: Procrustes fits for the 50 simulations of each of 4 cases (registered to the overall mean of the 32×52 configurations. (First column) After initial GPA. (Second and fourth columns) Final fits after using the semi-landmark method with full Procrustes and nearest point criteria. (Third and fifth columns) Edge superimposition of final fits (to match MD endpoints) back onto original tooth outlines. Dashed lines indicate ranges that each set of gingival landmarks were simulated between.

As suggested by the reduction in *EucSS(within)* in table 8.2, the second and fourth columns of fig. 8.18 show that, for both methods, all configurations simulated from the same case end up as being almost identical in shape, satisfying one of the key aims of the investigation. We shall consider why this occurs and why the FP and NP criteria result in almost exactly the same sets of final configurations in section 8.3.3.1. Notice that unlike with the BE method, this is not achieved at the expense of an increase in *EucSS(cases)*, as can be seen in fig. 8.11, where the variation between cases at the gingival landmarks has clearly been reduced.

When superimposed back onto the space of the outline of the case being represented, the final positions of the gingival landmarks of each set of simulations from 30 out of the 32 cases were found to be within the ranges of feasible variation they were originally simulated between, see fig. 8.18 rows 2-4 for examples with both methods. However, for the sets of representations of two of the cases, here in the top and bottom rows of fig. 8.18, the landmark on the gingival margin at the end of the LACC always ends up in a position incisal (top row) or cervical (bottom row) to the range of realistic positions. Again this is observed for both methods, since the resulting final sets of configurations are almost identical. We shall consider the results for these two particular cases in more detail in section 8.3.3.2.

8.3.3.1 Why representations of the same case end up the same shape and both new methods give similar results

For the FP criterion, the final shapes of each configuration are determined on more or less the first iteration, as seen in table 8.1 and illustrated in fig. 8.19 below. This shows the various stages of the first iteration for two representations *A* and *B* of case '8', when included in the entire sample of 32×50 configurations. As in 8.3.2.1, the gingival landmarks of *A* and *B* are at the cervical and incisal bounds of possible positions, i.e. for case *A*, at the CEJ markers and for case *B*, at positions 40% of the way along the LACC (papilla landmarks) and 70% of the way along the LAAC from the incisal edge (gingival margin landmark). As with any two simulated configurations from the same case, the non-gingival landmarks are in identical positions.

In the usual notation, let $X_{(0)}^{P}$ denote the original 'nominal' positions of a configuration after Procrustes registration to the initial estimate of mean shape $\hat{\mu}_{(0)}$ and $vec^{-1}(vec(X_{(0)}^{P}) - U\lambda) = X_{(1)}^{new}$ denote the new version of a configuration after moving the semi-landmarks. As shown in the second and third columns of fig. 8.19, the semi-landmarks move so as to minimise $\left\| SG(vec^{-1}(vec(X_{(0)}^{P}) - U\lambda)) - \hat{\mu}_{(0)} \right\|^2$ as the semi-landmarks vary along their chords by distance λ and $vec^{-1}(vec(X_{(0)}^{P}) - U\lambda)$ varies under the action of the 'similarity group' of transformations.

Original GPA registration of $X_{(0)}^{P}$ to initial mean shape $\hat{\mu}_{(0)}$

Centre left: Movement of semi-landmarks from nominal positions to give configuration $X_{(1)}^{new}$ whose shape minimises $d_F^2(X_{(1)}^{new}, \hat{\mu})$ or $\left\| SG(X_{(1)}^{new}) - \hat{\mu}_{(0)} \right\|^2$ as shown in centre right.

GPA registration of $X_{(1)}^{new}$ to updated mean shape $\hat{\mu}_{(1)}$, obtained from all new configs.

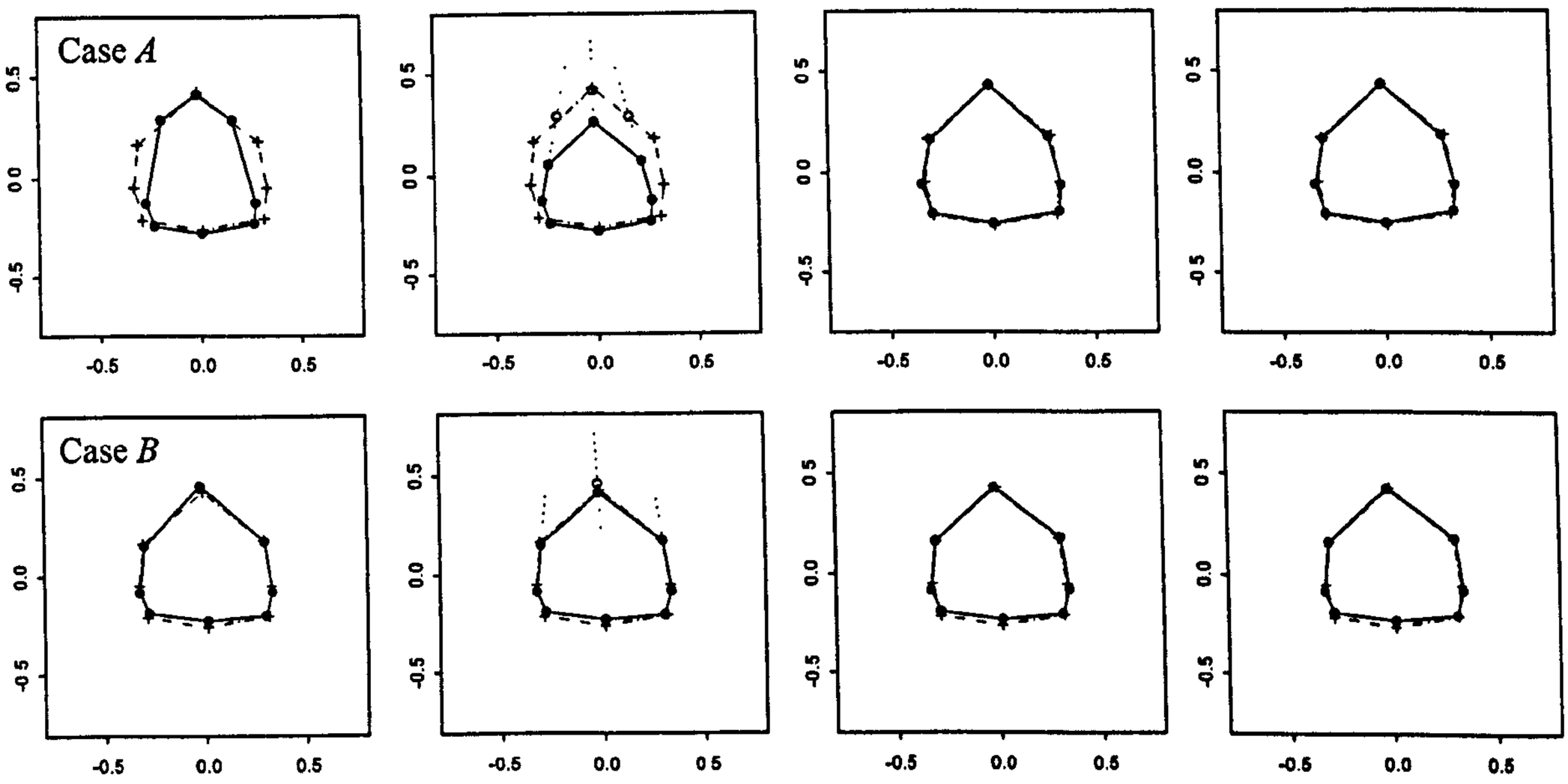


Fig. 8.19: First iteration of semi-landmark procedure for a two representations *A* and *B* of case '8' (filled circles, solid lines), with gingival landmark positions at opposite ends of the simulation ranges. Full Procrustes criterion used to determine new semi-landmark positions and all 32×50 configurations used to compute $\hat{\mu}_{(0)}$ and $\hat{\mu}_{(1)}$ (plusses, dashed lines). (Empty circles) Nominal semi-landmark positions. (Dotted lines) directions of semi-landmarks chords.

For each case *A* and *B*, $vec^{-1}(vec(X_{(0)}^{P}) - U\lambda)$ and $\hat{\mu}_{(0)}$ have identical configurations of fixed landmarks. If these were the only landmarks to be superimposed then the two configurations could be matched exactly. However, stopping this are the directions and

positions (relative to the other landmarks) of the escribed chords, along which locations must be found to simultaneously minimise the total sum of squared Euclidean distances between the both the fixed and semi-landmarks in $\hat{\mu}_{(0)}$ and $vec^{-1}(vec(X^P_{(0)}) - U\lambda)$. As there are five squared distances contributing to $\left\| SG(vec^{-1}(vec(X^P_{(0)}) - U\lambda)) - \hat{\mu}_{(0)} \right\|^2$ (between the fixed landmarks), compared to just three between the semi-landmarks and the gingival landmarks in $\hat{\mu}_{(0)}$, the superimposition is still to some extent, dominated by the fixed landmarks. However, since the semi-landmarks can only move in a roughly vertical direction, the match to $\hat{\mu}_{(0)}$ is constrained by the horizontal positions which the semi-landmarks must adopt along their chords (the contribution to the sum of squares at these positions being perpendicular to the chords). Since the configuration of fixed landmarks is identical in representations *A* and *B* of case '8' and the directions and relative locations of the semi-landmark's chords roughly in the same, the actual starting position of the semi-landmarks along the chords do not matter. Any two representations of the same case will end up almost exactly the same in shape. (See fig. 8.19, 3rd col.). The only difference in the new (and final) positions of the semi-landmarks is due to the slight differences in the chord directions and locations, as explained in 8.2.5.2. (As before, the gingival landmarks in *A* and *B* were set at their upper and lower simulation limits so as to maximise this effect). Note also that in the last two columns of fig. 8.19, the updated mean shape, obtained from all 32×50 configurations at the end of the iteration, is almost unchanged and that the Procrustes superimposition of $X^{new}_{(1)}$ to $\hat{\mu}_{(0)}$ is almost identical to the GPA registration of $X^{new}_{(1)}$ to $\hat{\mu}_{(1)}$. Therefore the process quickly converges.

For the NP method, the final configurations of each representation *A* and *B* of cases 8 and 23 end up almost identical to those obtained using the FP method, with very similar values of $d_F^2(X^{P_{SL}}, \hat{\mu}_{SL})$ or $\left\| X^{P_{SL}} - \hat{\mu}_{SL} \right\|^2$. However, since on any iteration, the new positions of the semi-landmarks along their chords are determined by minimising $\left\| vec^{-1}(vec(X^P) - U\lambda) - \hat{\mu} \right\|^2$ without allowing $vec^{-1}(vec(X^P) - U\lambda)$ to also be rotated, translated or scaled to $\hat{\mu}$ (this is done at the next GPA step with respect to a practically unchanged mean shape), it usually takes several more iterations before positions which also minimise $\left\| SG(vec^{-1}(vec(X^P) - U\lambda)) - \hat{\mu} \right\|^2$ are attained. See fig. 8.20.

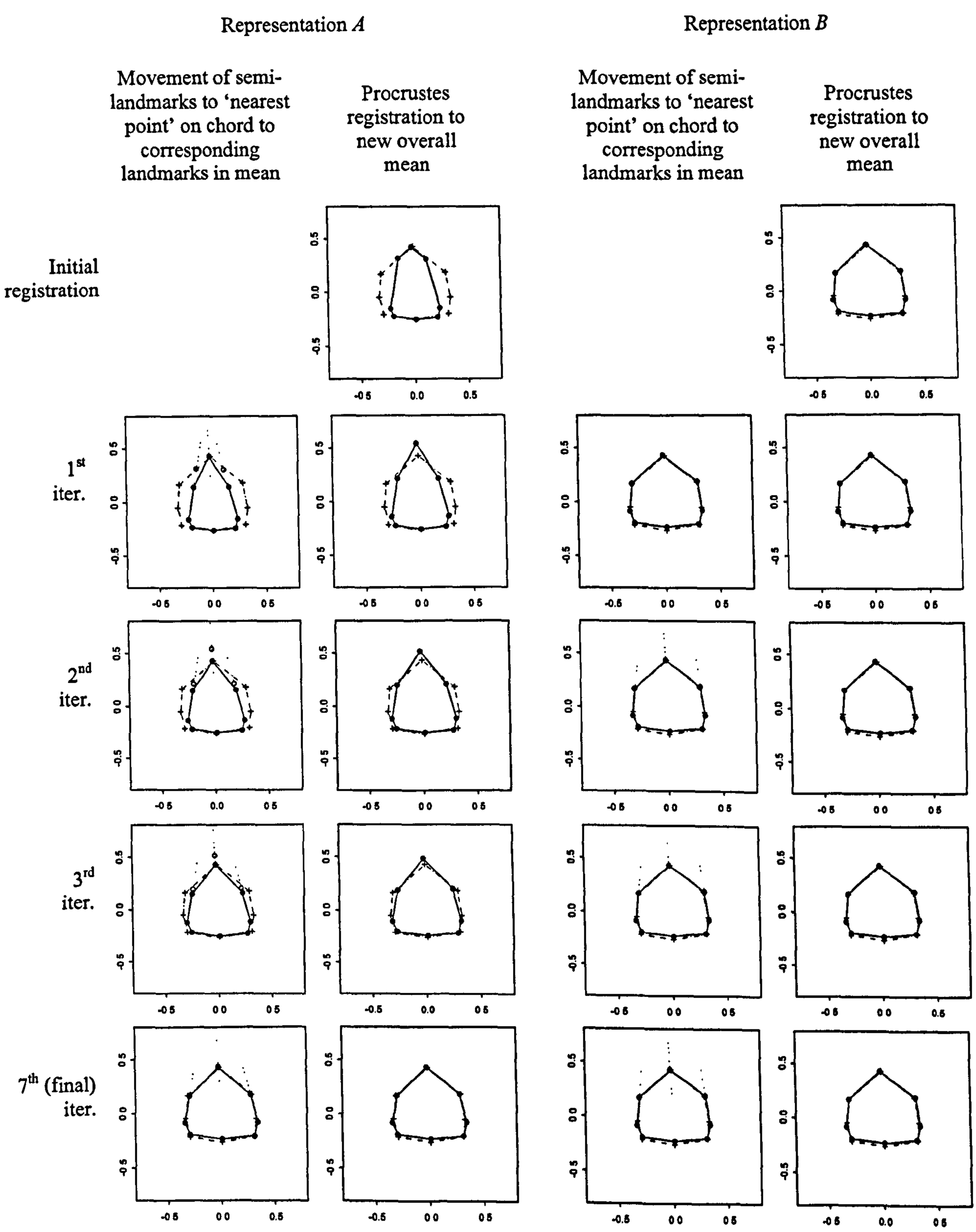


Fig 8.20: Iterations of 'nearest point' method for representations *A* and *B* of case 8 (filled circles, solid lines), with gingival landmarks at upper & lower simulation limits. All 32×50 configurations used to compute Procrustes mean shapes (plusses, dashed lines). (Empty circles) Nominal semi-landmark positions at start of each iteration. (Dotted lines) directions of escribed chords.

8.3.3.2 Gingival margin landmarks outside ranges of realistic variation

In fig. 8.18 it was seen that all final representations of cases 2 and 8 (top row and bottom row), had gingival margin landmark positions outside of the ranges considered as 'realistic' when referred back to the outlines of the crowns they were originally simulated from. For case 2, all configurations had gingival margin landmarks which would be cervical (above) the CEJ and for case 8, these positions were all incisal (below) the bound at 70% of the way along the LAAC from the incisal edge. However, for the final fits of the other 30 sets of simulations, this problem does not occur.

As can be seen in the first row, first panel of fig. 8.19 or fig. 8.20, the Procrustes registration of representation A of case 8 ($X^{P}_{(0)}$) to the initial mean shape of all 32×50 configurations ($\hat{\mu}_{(0)}$) does not 'fit' particularly well, but by allowing the gingival landmarks to move, so as to minimise the FP or NP criterion, at the end of the iteration (and each subsequent iteration), an improved fit to the practically unchanged mean shape is achieved. However, for any representation of this case, regardless of the initial simulated positions of the gingival landmarks, this always ultimately results in the gingival margin semi-landmark moving out of its feasible range on the actual outline of the crown it was simulated from (here to below the limit at 30% of the way along the LAAC from the CEJ). This is because the gingival margin landmark in the mean shape, to which the configuration has been matched, is itself in a position outside the range of feasible positions for this case, as can be seen when the original outline and gingival simulation ranges are reconstructed at the final Procrustes fits, as show in fig. 8.21(top). Similarly, while all representations of case 2 result in final semi-landmark positions which minimise squared distances from their chords to corresponding positions in the mean shape, this results in a final gingival margin landmark position which would be above the CEJ of this case. Again this is because the corresponding landmark in the same overall mean shape is in a position outside the range of feasible positions for this case, as show in fig. 8.21(bottom). (Fig. 8.21 was constructed by calculating the Bookstein-type transformation required to send the MD endpoints on the original outlines of cases 2 and 8 to one pair of these landmarks in the final Procrustes fits of each set of simulations, then applying this to all other outline coordinates and the simulation ranges).

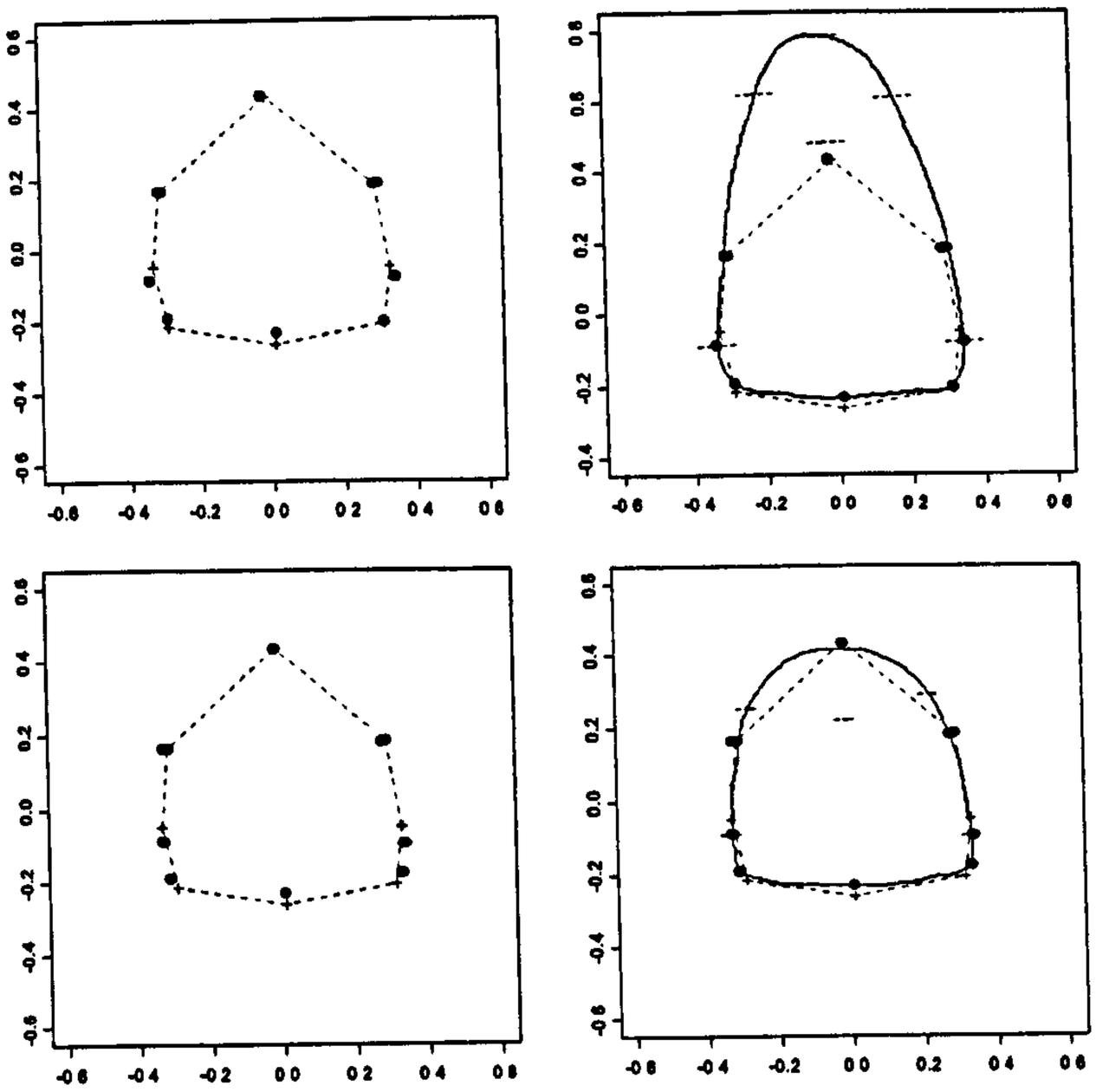


Fig. 8.21: (left) Final Procrustes fits about overall estimated mean shape. (Right) reconstruction of crown outline using positions of fixed landmarks of Procrustes fits. (Top) for representations of case 8 (Bottom) for representations of case 2. (Dashed lines, plusses) mean shape. (Dashed horizontal lines) cervical and incisal range limits that each set of gingival landmarks were simulated between.

If we consider the original crown shapes of cases 2 and 8 alongside the other 30 crown configurations using the CEJ markers as landmarks, rather than possible gum positions (as in 8.3.3.2), then GPA registration of the 32 configurations shows that cases 2 and 8 are actually 'outliers'. i.e. their shapes are a long way from the mean of the sample. An examination of the residual sums of squares or the squared full Procrustes distances of the fits X_i^P of the i=1,...,32 crown configurations about their mean shape $\hat{\mu}$, showed that cases 2 and 8 had the largest values (see fig. 8.22 below). These differences in shape are roughly in the direction of the semi-landmark chords and so the simulated gingival landmark positions based on the CEJ locations of each case are able to move to positions which provide an improved fit to the mean shape. However, this results in new positions outside of their feasible ranges.

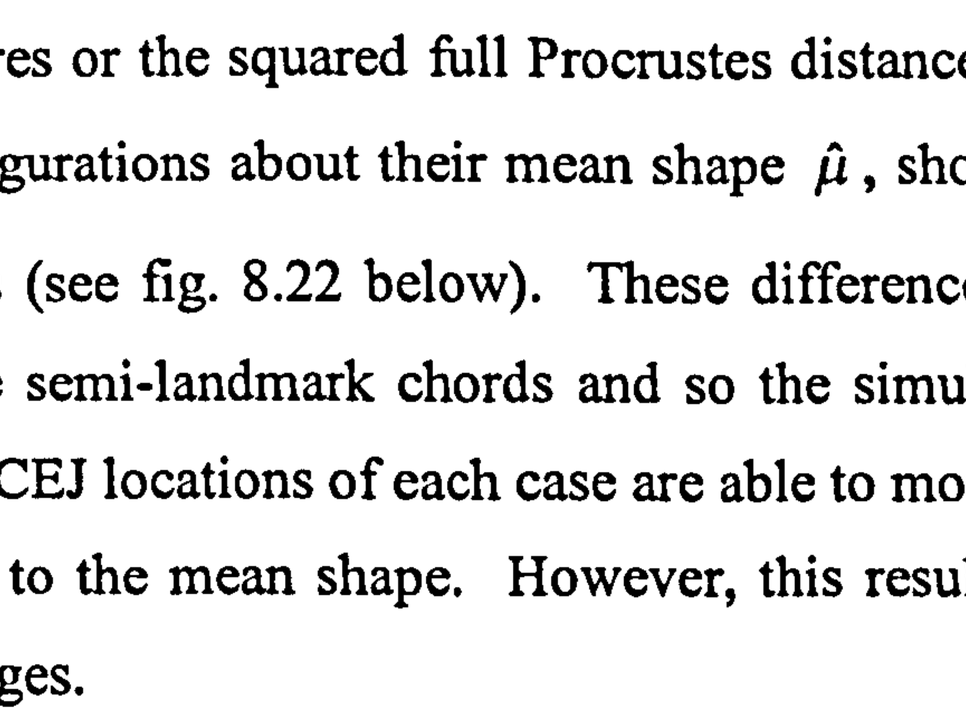


Fig. 8.22: Squared Full Procrustes distances of each crown configuration from estimated mean shape.

To stop this, one easy solution would be to constrain the landmarks not to move outside these bounds. In practice however, on unextracted teeth we will not know the positions of the CEJ markers on which these ranges are based. Landmarks also need to be homologous between cases and so it would not be beneficial to have final positions of these landmarks corresponding to 'standardised gingival positions' on some cases, and the CEJ markers on others. Since there were only 2 from 32 cases for which this situation occurred and the final positions of the gingival landmarks were only just outside their feasible ranges, then this should not be too much of a problem with any similar sample (and it appears from fig. 8.22 that cases with these actual crown shapes are rare). Another way to consider this issue is that in choosing the chord directions for the semi-landmarks to move along, we are effectively declaring that any variation in shape along these directions is of no interest and can be removed. This therefore not only includes variation in these landmarks due to the positions of the patient's gum, but also that resulting from any other biological features such as the positions of the CEJ markers. Since the landmarks only move marginally outside their feasible ranges and reduce the variation at these positions zero, one could interpret the final positions as simply lacking any information along these directions, regarding them as 'standardised' positions relative to the mean shape.

8.3.3.3 Reduction in variation at fixed landmarks

By reducing the unwanted variation at the gingival landmark positions, fig. 8.11 indicated that after running both the FP and NP methods, the variation in the fixed landmarks at the ends of the MD width, incisal corners and incisal endpoint of the LACC, was also reduced. To investigate this further, we considered a smaller sample of 32 configurations, each one randomly selected from the set of the 50 simulations, and re-ran both procedures. A smaller, but representative sample of the entire simulation dataset was used so that the various frequency comparisons of sums of squared differences used below, could be more easily produced. The results presented were obtained using the iterative method (ii) for GPA (with $RSS(\hat{\mu})=RSS$). Where a result would have been different had GPA method (i) been used, details are provided.

The sum of squared Euclidean distances between the $j=1,\ldots,k$ landmarks of a configuration Y and their corresponding positions in an estimated mean shape $\hat{\mu}$, may

be separated into separate contributions from the semi-landmarks, which form a sublist $j_{(l)}$, l=1,…,3 and the fixed landmarks, $j \notin j_{(l)}$, with:

$$\|Y-\hat{\mu}\|^2=\sum_{\substack{semi\\ j\in j_{(l)}}}\|y_j-\hat{\mu}\|^2+\sum_{\substack{fixed\\ j\notin j_{(l)}}}\|y_j-\hat{\mu}\|^2 .$$

8.3.3.3.1 Nearest point method

At each movement stage of the NP method (at the start of iteration r), the sum of squared Euclidean distances between the semi-landmarks and their corresponding positions in $\hat{\mu}_{(r-1)}$ is reduced as they move to positions along their chords at least Euclidean distance to the corresponding landmarks in $\hat{\mu}_{(r-1)}$, i.e. we obtain $X_{(r)}^{new}=vec^{-1}(vec(X_{(r-1)}^{P})-U_{(r)}\lambda_{(r)})$ satisfying:

$$\left\|X_{(r)}^{new}-\hat{\mu}_{(r-1)}\right\|^2=\min_{\lambda_{(r)}}\left\|vec^{-1}(vec(X_{(r-1)}^{P})-U_{(r)}\lambda_{(r)})-\hat{\mu}_{(r-1)}\right\|^2 . \qquad (8.5)$$

This always leads to:

$$\sum_{\substack{semi\\ j\in j_{(l)}}}\left\|x_{j(r)}^{new}-\hat{\mu}_{j(r-1)}\right\|^2\le\sum_{\substack{semi\\ j\in j_{(l)}}}\left\|x_{j(r-1)}^{P}-\hat{\mu}_{j(r-1)}\right\|^2 \qquad (8.6)$$

and since the positions of the fixed landmarks are unchanged,

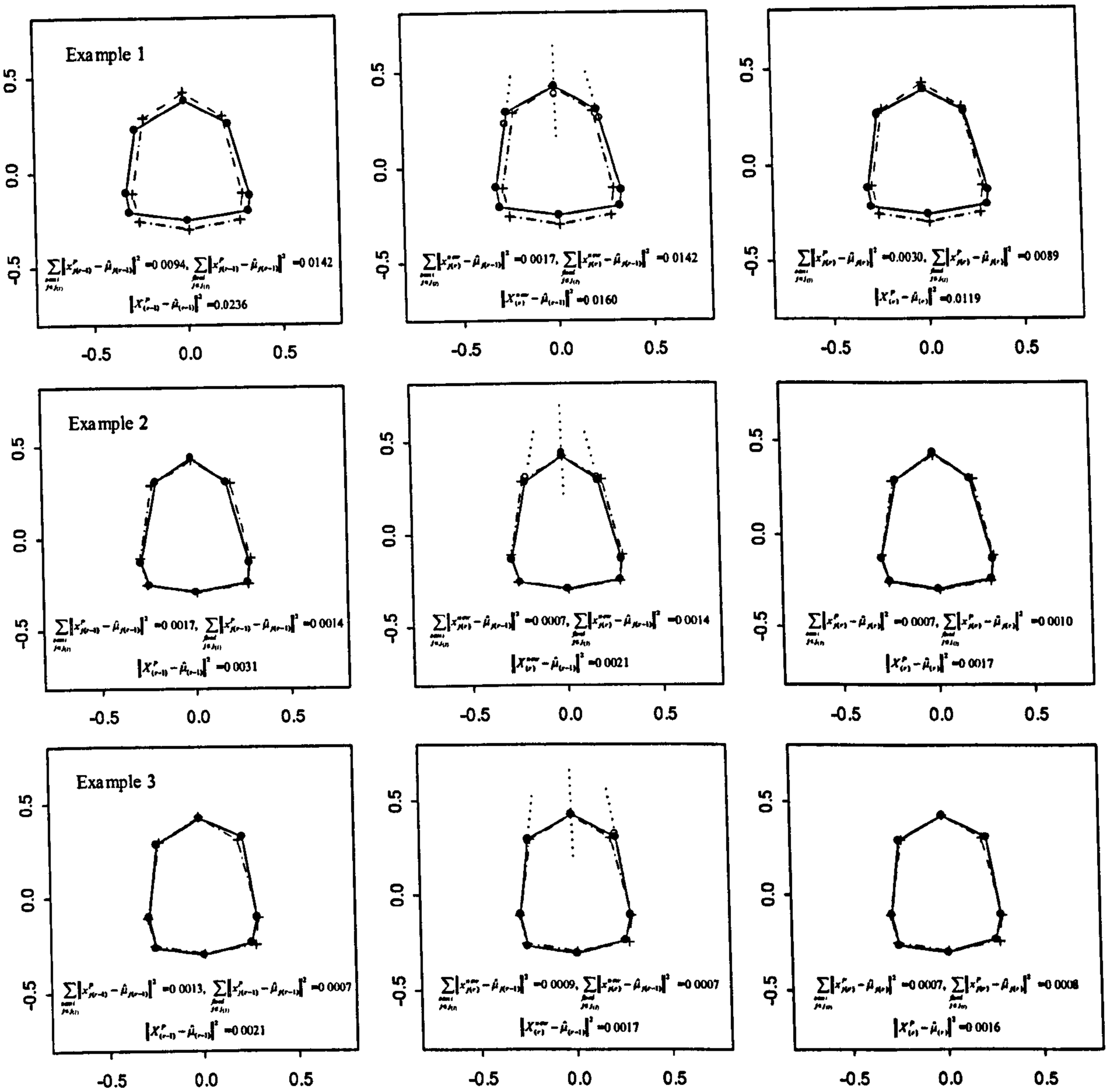
$$\left\|X_{(r)}^{new}-\hat{\mu}_{(r-1)}\right\|^2\le\left\|X_{(r-1)}^{P}-\hat{\mu}_{(r-1)}\right\|^2 . \qquad (8.7)$$

At the end of iteration r, $X_{(r)}^{new}$ is re-registered by GPA to a new estimate of mean shape $\hat{\mu}_{(r)}$ (with $\hat{\mu}_{(r)}\approx\hat{\mu}_{(r-1)}$) to obtain $X_{(r)}^{P}$. Because the ordinary least squares (OLS) matching gives equal weight to all pairs of corresponding landmarks as $X_{(r)}^{P}$ is registered to $\hat{\mu}_{(r)}$, the reduction in sum of squared distances between the semi-landmarks in $X_{(r)}^{new}$ and corresponding landmarks in $\hat{\mu}_{(r-1)}$ ($\approx\hat{\mu}_{(r)}$) in (8.6) is distributed around the rest of the configuration, allowing an improved overall fit to the essentially unchanged mean. Since $\hat{\mu}_{(r)}$ hardly changes we expect that for each configuration:

$$\left\|X^{P}_{(r)} - \hat{\mu}_{(r)}\right\|^2 \leq \left\|X^{P}_{(r-1)} - \hat{\mu}_{(r-1)}\right\|^2 \tag{8.8}$$

since $\left\|X^{P}_{(r)} - \hat{\mu}_{(r)}\right\|^2 \approx \min_{SG}\left\|SG(X^{new}_{(r)}) - \hat{\mu}_{(r)}\right\|^2 \leq \left\|X^{new}_{(r)} - \hat{\mu}_{(r-1)}\right\|^2 \leq \left\|X^{P}_{(r-1)} - \hat{\mu}_{(r-1)}\right\|^2$,

as depicted in fig. 7.3 (a) (top left). See fig. 8.23, first two columns. In other words, the contribution to *RSS* should always be reduced between iterations.



Original GPA registration of individual configuration (filled circles, solid lines) to mean shape (plusses, dashed line)

Semi-landmark movement from nominal positions (empty dots) to new positions (filled dots, solid line) at least Euclidean distance to corresponding landmarks in mean shape (plusses, dashed line)

GPA registration of new configuration (filled circles, solid line) to updated mean shape (plusses, dashed line)

Fig. 8.23: Redistribution of minimised sums of squared Euclidean distances during semi-landmark movement and after superimposition to updated mean shape.

However, as we described in 7.4.2, this is only true up until the process enters the 'slow manifold' for values of ΔRSS, when the size of change in $\hat{\mu}$, although small, becomes important relative to the change in X^P.

Tables 8.3 and 8.4 shows that when $X^{new}_{(r)}$ is registered to $\hat{\mu}_{(r)}$ to obtain $X^P_{(r)}$, the reduction in sums of squares, given by (8.10), results in one of four scenarios, (a) to (d), for the values of $\sum_{\substack{semi \\ j\in j_{(l)}}} \left\| x^P_{j(r)} - \hat{\mu}_{j(r)} \right\|^2$ and $\sum_{\substack{fixed \\ j\notin j_{(l)}}} \left\| x^P_{j(r)} - \hat{\mu}_{j(r)} \right\|^2$, when compared to the previous iteration, and one of four scenarios, (i) to (iv), when compared to the values of $\sum_{\substack{semi \\ j\in j_{(l)}}} \left\| x^{new}_{j(r)} - \hat{\mu}_{j(r-1)} \right\|^2$ and $\sum_{\substack{fixed \\ j\notin j_{(l)}}} \left\| x^{new}_{j(r)} - \hat{\mu}_{j(r-1)} \right\|^2$, where the former is the quantity optimised by the NP criterion and $\sum_{\substack{fixed \\ j\notin j_{(l)}}} \left\| x^P_{j(r-1)} - \hat{\mu}_{j(r-1)} \right\|^2 = \sum_{\substack{fixed \\ j\notin j_{(l)}}} \left\| x^{new}_{j(r)} - \hat{\mu}_{j(r-1)} \right\|^2$.

	$\sum_{\substack{semi \\ j\in j_{(l)}}} \left\| x^P_{j(r)} - \hat{\mu}_{j(r)} \right\|^2$ vs. $\sum_{\substack{semi \\ j\in j_{(l)}}} \left\| x^{new}_{j(r)} - \hat{\mu}_{(r-1)} \right\|^2$	$\sum_{\substack{fixed \\ j\notin j_{(l)}}} \left\| x^P_{j(r)} - \hat{\mu}_{j(r)} \right\|^2$ vs. $\sum_{\substack{fixed \\ j\notin j_{(l)}}} \left\| x^{new}_{j(r)} - \hat{\mu}_{(r-1)} \right\|^2$	Iteration											
			1	2	3	4	5	6	7	8	9	10	11	12
(a)	≤	≤	12	9	11	8	8	6	5	4	2	2	2	2
(b)	≥	≤	17	16	14	14	14	14	14	14	13	13	12	11
(c)	≤	≥	3	7	7	10	10	12	12	13	14	14	14	13
(d)	≥	≥	0	0	0	0	0	0	1	1	3	3	4	6

Table 8.3: Frequencies of changes in fixed and semi-landmark sums of squares following movement and GPA steps of nearest point method. (Shaded columns) converged at ΔRSS<0.0001.

	$\sum_{\substack{semi \\ j\in j_{(l)}}} \left\| x^P_{j(r)} - \hat{\mu}_{j(r)} \right\|^2$ vs. $\sum_{\substack{semi \\ j\in j_{(l)}}} \left\| x^P_{j(r-1)} - \hat{\mu}_{(r-1)} \right\|^2$	$\sum_{\substack{fixed \\ j\notin j_{(l)}}} \left\| x^P_{j(r)} - \hat{\mu}_{j(r)} \right\|^2$ vs. $\sum_{\substack{fixed \\ j\notin j_{(l)}}} \left\| x^P_{j(r-1)} - \hat{\mu}_{(r-1)} \right\|^2$	Iteration											
			1	2	3	4	5	6	7	8	9	10	11	12
(i)	≤	≤	29	23	22	16	15	11	8	5	3	3	3	2
(ii)	≥	≤	0	2	3	6	7	9	11	13	12	12	11	11
(iii)	≤	≥	3	7	7	10	10	12	13	14	16	16	17	18
(iv)	≥	≥	0	0	0	0	0	0	0	0	1	1	1	1

Table 8.4: Frequencies of changes in fixed and semi-landmark sums of squares between GPA steps of successive iterations of nearest point method. (Shaded columns) converged at ΔRSS<0.0001.

Table 8.3 shows how the reduction in sums of squares between the semi-landmarks and those in $\hat{\mu}_{(r-1)}$, achieved by moving the landmarks to their nearest point along the chords to those in $\hat{\mu}_{j(r-1)}$, is distributed around all landmarks on the following GPA step, to an approximately unchanged mean, $\hat{\mu}_{(r)}$, resulting in either an increase or decrease in $\sum_{\substack{semi \\ j \in j_{(l)}}} \left\| x^{P}_{j(r)} - \hat{\mu}_{j(r)} \right\|^2$, compared to the minimised value of $\sum_{\substack{semi \\ j \in j_{(l)}}} \left\| x^{new}_{j(r)} - \hat{\mu}_{(r-1)} \right\|^2$.

The tables show that that using the current sample, at the end of each iteration for the majority of configurations, GPA results in scenario (a) or (b) in table 8.3, with corresponding frequency total to that of (i) and (ii) in table 8.4. In particular, during the first few iterations, when most of the semi-landmark movement occurs,

$$\sum_{\substack{fixed \\ j \notin j_{(l)}}} \left\| x^{P}_{j(r)} - \hat{\mu}_{j(r)} \right\|^2 \leq \sum_{\substack{fixed \\ j \notin j_{(l)}}} \left\| x^{new}_{j(r)} - \hat{\mu}_{j(r-1)} \right\|^2 = \sum_{\substack{fixed \\ j \notin j_{(l)}}} \left\| x^{P}_{j(r-1)} - \hat{\mu}_{j(r-1)} \right\|^2 . \qquad (8.9)$$

As the variation at the semi-landmarks is reduced, so too is the effect it had on the variation at the other landmarks where GPA distributed unwanted variation around the rest of the configuration. Here this results in the final Procrustes registered configurations having smaller variance at the fixed landmarks, compared to their original GPA registration, as well as the semi-landmarks, as seen in fig. 8.11. (Since there are 5 non-gingival landmarks and only 3 gingival landmarks, the OLS registration typically concentrates on matching the 5 landmarks around the lower outline, rather than the 3 gingival ones in $X^{new}_{(r)}$ to $\hat{\mu}_{(r)}$). See, for example, fig. 8.23 (top two rows).

For the semi-landmarks, scenarios (i) and (iii) in table 8.4 most frequently occur (until convergence) with:

$$\sum_{\substack{semi \\ j \in j_{(l)}}} \left\| x^{P}_{j(r)} - \hat{\mu}_{j(r)} \right\|^2 \leq \sum_{\substack{semi \\ j \in j_{(l)}}} \left\| x^{P}_{j(r-1)} - \hat{\mu}_{j(r-1)} \right\|^2 . \qquad (8.10)$$

In particular, during the first few iterations, when most of the semi-landmark movement occurs, scenario (i) nearly always results, where the reduction in sums of squares between $X^{P}_{(r)}$ and $\hat{\mu}_{(r)}$ is achieved at both the semi and fixed landmarks, (see for example, second row of fig. 8.23). However, as the process continues towards

convergence and the movement of the semi-landmarks becomes smaller, there are fewer cases for which (i) occurs and an increasing number with scenario (ii) or (iii) occurring, where only one of $\sum_{\substack{semi \\ j \in j_{(i)}}} \left\| x^P_{j(r)} - \hat{\mu}_{j(r)} \right\|^2$ or $\sum_{\substack{fixed \\ j \in j_{(i)}}} \left\| x^P_{j(r)} - \hat{\mu}_{j(r)} \right\|^2$ decreases between iterations. For example, with scenario (ii), the variation in the fixed landmarks is still reducing, but the sums of squares of the semi-landmarks in the Procrustes fits about the mean have increased. With scenario (iii) the converse is true, compared to the previous GPA step (see example in row 3 of fig. 8.23). However, until convergence, the overall sums of squares of all landmarks is still reducing since we still have $\left\| X^P_{(r)} - \hat{\mu}_{(r)} \right\|^2 \leq \left\| X^P_{(r-1)} - \hat{\mu}_{(r-1)} \right\|^2$ for each configuration, as shown in table 8.5(i).

$\left\| X^P_{(r)} - \hat{\mu}_{(r)} \right\|^2$ vs. $\left\| X^P_{(r-1)} - \hat{\mu}_{(r-1)} \right\|^2$	Iteration											
	1	2	3	4	5	6	7	8	9	10	11	12
≤	32	32	32	32	32	32	32	30	25	21	19	12
≥	0	0	0	0	0	0	0	2	7	11	13	20

Table 8.5(i): Frequencies of changes in total sums of squares between GPA steps on successive iterations of nearest point method. (Shaded column) converged at ΔRSS<0.0001

$\left\| X^P_{(r)} - \hat{\mu}_{(r)} \right\|^2$ vs. $\left\| X^{new}_{(r)} - \hat{\mu}_{(r-1)} \right\|^2$	Iteration											
	1	2	3	4	5	6	7	8	9	10	11	12
≤	32	30	31	31	29	23	22	19	17	15	10	9
≥	0	2	1	1	3	9	10	13	15	17	22	23

Table 8.5(ii): Frequencies of changes in total sums of squares of nearest point method, following GPA of resulting configurations from movement step. (Shaded column) converged at ΔRSS<0.0001

Table 8.5(ii) shows that as we approach convergence (and beyond) an increasing proportion of configurations are produced with:

$$\left\| X^P_{(r)} - \hat{\mu}_{(r)} \right\|^2 \geq \left\| X^{new}_{(r)} - \hat{\mu}_{(r-1)} \right\|^2, \tag{8.11}$$

i.e. where GPA registration of the $X^{new}_{(r)}$ results in a larger sum of squares to new mean $\hat{\mu}_{(r)}$ compared to that minimised at the movement step between $X^{new}_{(r)}$ and $\hat{\mu}_{(r-1)}$, as described in 7.4.2. When

$$\left\| X^P_{(r)} - \hat{\mu}_{(r)} \right\|^2 \approx \min_{SG} \left\| SG(X^{new}_{(r)}) - \hat{\mu}_{(r)} \right\|^2 \geq \left\| X^{new}_{(r)} - \hat{\mu}_{(r-1)} \right\|^2 \geq \min_{SG} \left\| SG(X^{new}_{(r)}) - \hat{\mu}_{(r-1)} \right\|^2$$

(with equality on the LHS if GPA method (i) is used), this means that $\hat{\mu}_{(r)}$ has changed to be more different in shape from $X^{new}_{(r)}$ than $\hat{\mu}_{(r-1)}$ was from $X^{new}_{(r)}$, as depicted in fig. 7.3(a) (bottom left), although up until convergence, for each configuration, we still have $\left\|X^{P}_{(r)} - \hat{\mu}_{(r)}\right\|^2 \leq \left\|X^{P}_{(r-1)} - \hat{\mu}_{(r-1)}\right\|^2$, and so ΔRSS is still decreasing. Beyond convergence however, as we enter the flat part of the plot of ΔRSS versus r, new configurations eventually begin to be produced for which not only is (8.11) true but the change in X^P and $\hat{\mu}$ results in:

$$\left\|X^{P}_{(r)} - \hat{\mu}_{(r)}\right\|^2 \geq \left\|X^{P}_{(r-1)} - \hat{\mu}_{(r-1)}\right\|^2 \tag{8.12}$$

(as in fig 7.3 (a) (right)). That is, the NP method results in an increase in the contribution to RSS for some configurations, between iterations. As the proportion of cases for which (8.11) is true increases further, so does the proportion for which (8.12) is. The systematic change in $\hat{\mu}$, produced by systematic changes in the X_i^{new}, despite being very small, becomes larger than the change in shape of the X_i (from $\hat{\mu}$), as was depicted in fig 7.3 (a) (right, top or bottom). The effect of occurrences of (8.12) is to offset any decreases in RSS which are produced, slowing the convergence. As described in 7.4.2, the systematic changes in $\hat{\mu}$ and the X_i stop the value of ΔRSS dropping below a certain level, although the movement of the semi-landmarks is now very small, with ΔRSS <0.00001, so in practice this does not matter.

8.3.3.3.2 Full Procrustes method

When using the FP criterion instead, the final configurations are almost identical, although fewer iterations are required to obtain them. At each movement stage of an iteration we obtain $X^{new}_{(r)} = vec^{-1}(vec(X^{P}_{(r-1)}) - U_{(r)}\lambda_{(r)})$ for each configuration, satisfying:

$$\left\|SG_{(r)}(X^{new}_{(r)}) - \hat{\mu}_{(r-1)}\right\|^2 = \min_{SG_{(r)},\lambda_{(r)}} \left\|SG_{(r)}(vec^{-1}(vec(X^{P}_{(r-1)}) - U_{(r)}\lambda_{(r)})) - \hat{\mu}_{(r-1)}\right\|^2 \tag{8.13}$$

with

$$\left\|SG(X^{new}_{(r)}) - \hat{\mu}_{(r-1)}\right\|^2 \leq \left\|SG(X^{P}_{(r-1)}) - \hat{\mu}_{(r-1)}\right\|^2 \approx \left\|X^{P}_{(r-1)} - \hat{\mu}_{(r-1)}\right\|^2 \tag{8.14}$$

(with equality on the RHS, if GPA method (i) used). The optimised full Procrustes superimposition corresponding to (8.13) above, again results in one of four different scenarios, (a) to (d), here for $\sum_{\substack{semi \\ j\in j_{(l)}}}\left\|SG(x^{new}_{j(r)})-\hat{\mu}_{j(r-1)}\right\|^2$ and $\sum_{\substack{fixed \\ j\notin j_{(l)}}}\left\|SG(x^{new}_{j(r)})-\hat{\mu}_{j(r-1)}\right\|^2$, compared to $\sum_{\substack{semi \\ j\in j_{(l)}}}\left\|SG(x^{P}_{j(r-1)})-\hat{\mu}_{j(r-1)}\right\|^2$ and one of scenarios (i) to (iv) for $\sum_{\substack{semi \\ j\in j_{(l)}}}\left\|x^{P}_{j(r)}-\hat{\mu}_{j(r)}\right\|^2$ and $\sum_{\substack{fixed \\ j\notin j_{(l)}}}\left\|x^{P}_{j(r)}-\hat{\mu}_{j(r)}\right\|^2$, compared to $\sum_{\substack{semi \\ j\in j_{(l)}}}\left\|x^{P}_{j(r-1)}-\hat{\mu}_{j(r-1)}\right\|^2$ and $\sum_{\substack{fixed \\ j\notin j_{(l)}}}\left\|x^{P}_{j(r-1)}-\hat{\mu}_{j(r-1)}\right\|^2$. The frequencies (out of 32) with which each of these scenarios occurred with our sample of 32 cases are detailed in tables 8.6 and 8.7 below.

	$\sum_{\substack{semi \\ j\in j_{(l)}}}\left\|SG(x^{new}_{j(r)})-\hat{\mu}_{j(r-1)}\right\|^2$ vs. $\sum_{\substack{semi \\ j\in j_{(l)}}}\left\|SG(x^{P}_{j(r-1)})-\hat{\mu}_{j(r-1)}\right\|^2$	$\sum_{\substack{fixed \\ j\notin j_{(l)}}}\left\|SG(x^{new}_{j(r)})-\hat{\mu}_{j(r-1)}\right\|^2$ vs. $\sum_{\substack{fixed \\ j\notin j_{(l)}}}\left\|SG(x^{P}_{j(r-1)})-\hat{\mu}_{j(r-1)}\right\|^2$	Iteration						
			1	2	3	4	5	6	7
(a)	≤	≤	25	0	0	0	0	0	0
(b)	≥	≤	0	18	18	18	18	18	18
(c)	≤	≥	7	14	14	14	14	14	14
(d)	≥	≥	0	0	0	0	0	0	0

Table 8.6: Frequencies of changes in fixed and semi-landmark sums of squares following optimisation of FP superimposition to current mean. (Shaded column) converged at $\Delta RSS<0.0001$.

	$\sum_{\substack{semi \\ j\in j_{(l)}}}\left\|x^{P}_{j(r)}-\hat{\mu}_{j(r)}\right\|^2$ vs. $\sum_{\substack{semi \\ j\in j_{(l)}}}\left\|x^{P}_{j(r-1)}-\hat{\mu}_{j(r-1)}\right\|^2$	$\sum_{\substack{fixed \\ j\notin j_{(l)}}}\left\|x^{P}_{j(r)}-\hat{\mu}_{j(r)}\right\|^2$ vs. $\sum_{\substack{fixed \\ j\notin j_{(l)}}}\left\|x^{P}_{j(r-1)}-\hat{\mu}_{j(r-1)}\right\|^2$	Iteration						
			1	2	3	4	5	6	7
(i)	≤	≤	25	20	19	19	19	19	19
(ii)	≥	≤	0	10	12	12	12	12	12
(iii)	≤	≥	7	0	0	0	0	0	0
(iv)	≥	≥	0	2	1	1	1	1	1

Table 8.7: Frequencies of changes in fixed and semi-landmark sums of squares between GPA steps of successive iterations of Full Procrustes method. (Shaded column) converged at $\Delta RSS<0.0001$.

As we would expect, on the first iteration, when practically all of the semi-landmark movement occurs, for each configuration we have:

$$\sum_{\substack{semi\\ j\in j_{(l)}}}\left\|SG(x^{new}_{j(r)})-\hat{\mu}_{j(r-1)}\right\|^2 \leq \sum_{\substack{semi\\ j\in j_{(l)}}}\left\|SG(x^{P}_{j(r-1)})-\hat{\mu}_{j(r-1)}\right\|^2 .$$

For the majority of configurations the sums of squares at the fixed landmarks is also able to be reduced (scenario (a) in table 8.6), as the OLS criterion of (8.13) optimises the fit to all landmarks in $\hat{\mu}_{(r-1)}$ as the semi-landmarks vary along their chords. Since $\hat{\mu}_{j(r)} \approx \hat{\mu}_{j(r-1)}$, then following GPA at the end of first iteration, this results in scenario (i) or (iii) in table 8.7, just as with the NP method. The sums of squares at the semi-landmarks are always reduced and since this variation is no longer distributed about the rest of the landmarks by GPA, for the majority of configurations, a reduction in variance at the fixed landmarks is also able to be achieved, as observed in fig. 8.11.

As noted in 7.4.1, for many of the configurations, on each iteration,

$$\left\|X^{P}_{(r)}-\hat{\mu}_{(r)}\right\|^2 \approx \min_{SG}\left\|SG(X^{P}_{(r)})-\hat{\mu}_{(r)}\right\|^2 \geq \min_{SG}\left\|SG(X^{new}_{(r)})-\hat{\mu}_{(r-1)}\right\|^2 \qquad (8.15)$$

(with equality on the LHS if GPA method (i) used). i.e. GPA registration, at the end of the rth iteration, results in an increased sum of squares between the Procrustes fit of $X^{new}_{(r)}$ to $\hat{\mu}_{(r)}$, compared to the value optimised when $X^{new}_{(r)}$ was superimposed to $\hat{\mu}_{(r-1)}$, i.e. $\hat{\mu}_{(r)}$ changes to be more different in shape from $X^{new}_{(r)}$ than $\hat{\mu}_{(r-1)}$ was from $X^{new}_{(r)}$. (See fig 7.3(a) (bottom left). As noted in 7.4.1, this is not unexpected since each $X^{new}_{(r)}$ is determined to minimise the OLS between its landmarks and those of $\hat{\mu}_{(r-1)}$ and not $\hat{\mu}_{(r)}$. As table 8.8 shows however, for all configurations, on the first iteration,

$$\left\|X^{P}_{(r)}-\hat{\mu}_{(r)}\right\|^2 \leq \left\|X^{P}_{(r-1)}-\hat{\mu}_{(r-1)}\right\|^2 . \qquad (8.16)$$

i.e. the contribution to RSS by each configuration is reduced and ΔRSS decreases, since if $\hat{\mu}_{(r)} \approx \hat{\mu}_{(r-1)}$, then by (8.12),

$$\left\|X^{P}_{(r)}-\hat{\mu}_{(r)}\right\|^2 \approx \left\|SG(X^{P}_{(r)})-\hat{\mu}_{(r)}\right\|^2 \leq \left\|SG(X^{P}_{(r-1)})-\hat{\mu}_{(r-1)}\right\|^2 \approx \left\|X^{P}_{(r-1)}-\hat{\mu}_{(r-1)}\right\|^2$$

(with '=' instead of '≈' if GPA method (i) is used).

$\left\|X^P_{(r)} - \hat{\mu}_{(r)}\right\|^2$ vs. $\left\|X^P_{(r-1)} - \hat{\mu}_{(r-1)}\right\|^2$	iteration						
	1	2	3	4	5	6	7
≤	32	30	30	30	30	30	30
≥	0	2	2	2	2	2	2

Table 8.8: Frequencies of changes in total sums of squares between GPA steps on successive iterations of Full Procrustes method. (Shaded column) converged at ΔRSS<0.0001.

However, the size of the change in $\hat{\mu}$ (relative to the changes in the X_i), despite being very small, again becomes important when the process fails to converge. On the second iteration and beyond, the movements of the semi-landmarks are very small and the sums of squares is only able to be reduced at either the semi or the fixed landmarks when superimposing $X^{new}_{(r)} = vec^{-1}(vec(X^P_{(r-1)}) - U_{(r)}\lambda_{(r)}$ to $\hat{\mu}_{(r-1)}$, while allowing the semi-landmarks to move along their chords, but never both (see table 8.6). As table 8.8 shows, from the second iteration onwards, in this sample there are two configurations where the change in $\hat{\mu}$ and X^P always results in

$$\left\|X^P_{(r)} - \hat{\mu}_{(r)}\right\|^2 \geq \left\|X^P_{(r-1)} - \hat{\mu}_{(r-1)}\right\|^2 \geq \left\|SG_{(r)}(X^{new}_{(r)}) - \hat{\mu}_{(r-1)}\right\|^2 \qquad (8.17)$$

and an increase in the contribution to ΔRSS (or $\Delta RSS(\hat{\mu})$) for these configurations.

Again, despite being very small, the systematic change in $\hat{\mu}$, produced by systematic changes in the X^{new}_i, is larger than the change in shape of the X^P_i, as depicted in fig. 7.3 (a, right and b). Convergence is again very slow with the continual changes in $\hat{\mu}$ and X_i stopping the value of ΔRSS (or $\Delta RSS(\hat{\mu})$) falling below a certain level. Since there are always configurations for which (8.17) is true, this can also result in ΔRSS (or $\Delta RSS(\hat{\mu})$) increasing between iterations, as seen in fig 7.2 (left). Again however, the changes in X^P_i and hence $\hat{\mu}$, are very small, so in practice this issue is unlikely to cause problems.

8.4 Summary and discussion

In this chapter we have investigated how well the original bending energy (BE), and newly proposed nearest point (NP) and full Procrustes (FP) semi-landmark methods perform in addressing the problem of unwanted variation in the gingival margin and interdental papillae landmarks, when analysing the shape of configurations representing the buccal surface of upper central incisors. Based on images where the entire crown outline is visible, multiple configurations representing each case were simulated, each with the same non-gingival landmarks, but different possible sets of the gingival landmarks, generated according to prior knowledge from the dental literature and the experience of periodontologists. The key aim was that variations in shape due to differences in the positions of the gingival landmarks be removed (filtered out), by allowing the gingival landmarks to move iteratively along chords representing the LAAC (or LACC) or sides of the buccal surface. Although we have concentrated on the upper central incisor, the procedures and investigations described would be suitable for similar future investigations of the buccal surfaces of other tooth types.

For the BE method it was found that, although the variation within cases is eliminated as required, this is at the expense of an increase in variation between cases. Representations of the same case all converge to the same unrealistic shapes, when referred back to the outlines of the crowns from which they were simulated, with new semi-landmark locations chiefly determined by the configuration of the non-gingival landmarks in the directions of the semi-landmarks chords.

Using the NP and FP criteria, both methods were successful in substantially reducing variation in the gingival landmark positions in the directions of the specified chords, both within and between cases and also at the fixed non-gingival landmarks. In fact, both method give very similar results, in terms of the final sets of Procrustes registered configurations and the various summary measures based on them.

With both methods, movement of the semi-landmarks results in the residual sums of squares about the estimated mean shape being reduced at not only the semi-landmarks, but also at the fixed landmarks for most configurations. This may be achieved in just

one or two iterations using the FP method or several iterations of the NP method. Note however, that despite taking more iterations, the NP method was found to be notably faster, in terms of processing time.

For 2 of the 32 cases considered, the shape change required by configurations simulated from these cases, resulted in configurations with gingival margin landmarks just outside their ranges of 'feasible' variation, when referred back to their original crown outlines. However it was argued that this should not present too much of a problem in future applications of these methods on upper central incisors, when wishing to allow for differences between cases in the position of a patient's gingival tissue.

Note that throughout this investigation, we have only considered the unwanted variation in the gingival landmarks, using the same locations for the non-gingival landmarks on each representation (simulation) of the same case. In chapter 4, it was seen that the landmarks at the ends of the MD width and corners of the incisal edge were affected by inter-operator inconsistency and in chapter 5, that the original (minimum bending energy) method was successful in reducing this unwanted variation. We will consider the use of the new semi-landmark methods in addressing problems of unreliability in the following chapter but here we have solely concentrated on the main problem of concern caused by the effects of gingival tissue variation on our analysis of shape.

In chapters 9 and 10, we consider how the different semi-landmark methods perform in other applications, both in the study of tooth shape (with different tooth types and problems) and when applied to configurations generated from simple geometric forms. We return to the data and results of the reliability study in chapter 4 and investigate use of the different choices of objective function in addressing issues of operator inconsistency in the positioning of landmarks and in removing variation in shape resulting from orientation inconsistencies at the imaging stage. We then consider use of the different semi-landmark criteria in filtering out unwanted patterns of variation simulated on configurations generated from basic known shapes, to illustrate how use of the new methods may be extended to other applications.

Chapter 9

Application to reliability problems in the study of tooth shape

9.1 Introduction

In this and the following chapter we consider further applications and uses of the various semi-landmark methods, both in the study of tooth shape and when applied to distorted configurations generated from simple geometric forms.

In this chapter we return to the data and results of the reliability study of chapter 4 and investigate use of the different choices of semi-landmark criteria in addressing some of the main issues identified in section 4.5.3. For the tooth surfaces considered, PCA of the within-case variation in Procrustes fits (and an examination of operators' images) suggested two main issues resulting from operator inconsistency in their representation as landmark configurations. In the buccal view, discrepancies in the positions of certain landmarks around the edges of the tooth were found to be a common problem. In the occlusal view, unwanted variation in shape appeared to result from orientation differences in the bucco-lingual direction at the imaging stage.

In section 9.2, we reconsider two of each of the buccal and occlusal surface data sets from chapter 4 and investigate how semi-landmarks and chord directions may be used to be represent the inconsistencies and lack of homology in certain landmarks, resulting from the reasons described above. For orientation problems, we introduce the idea of constrained semi-landmark movement, so that the configuration of a subset of semi-landmarks (e.g. those along the incisal edge) is always retained as the landmarks move along their chords. As in the previous chapter, we consider how the available data should be used in order to investigate the ability of the different semi-landmark methods to filter out this variation during iterative registration, and what summary measures

should be calculated and compared (with the results of the original GPA registration of each dataset in chapter 4), in order to evaluate the success of each method.

Section 9.3 presents the results of each of the applications, in the form of plots and summary tables. From these, and further investigations, we identify similarities and differences in how the new and original semi-landmark methods perform in different situations. Section 9.4 then considers what happens when we also include semi-landmarks and chord directions representing variation in the locations of landmarks, due to differences in the position of patients' gingival tissue. This includes a demonstration of how the means of two populations of upper central incisor configurations may be compared, having used the new semi-landmark methods to remove unwanted variation, resulting from both operator inconsistencies and differences in the position of patients' gingival tissue.

9.2 Materials and methods

9.2.1 Choice of semi-landmarks and chord directions

9.2.1.1 Buccal surfaces

In section 5.6.1 we saw how use of the original semi-landmark routine, with minimum bending energy (BE) used to determine the new positions of the semi-landmarks at each step, was able to improve the overall reliability of different operators' representations of the buccal surface of upper central incisors. A pictorial examination of the within-case variation in Procrustes fits (and operator feedback) suggested inconsistencies in the positions of the mesio-distal endpoints and landmarks at the corners of the incisal edge, as the cause of most unwanted variation in shape. This variation appeared to be in directions around the edges of the tooth and so at every step of the iteration, each of the four landmarks was allowed to move along its 'escribed chord' approximating the outline of the surface in its immediate vicinity.

In fig 4.6 (4^{th} row), the same patterns of variation were also evident for the lower canine, with the first two PC's accounting for approximately 57% of within case

variation in these directions. For the upper central incisor this figure was 68%. For the lower central incisor (fig 4.6, 2nd row), inconsistencies in the positions of the mesio-distal endpoints were also evident on the first PC (50% of within case variation), but here the landmarks at the corners of the incisal edge appeared more reliable. This would be expected since these teeth have squarer corners than the other two buccal surfaces, allowing the landmark to be located more consistently.

Since application to the lower canine would involve the same combination of fixed and semi-landmarks and chord directions as the upper central incisor, we investigate the new semi-landmark methods on the upper and lower central incisor data sets from chapter 4, using the semi-landmarks and associated chord directions defined in fig 9.1. In contrast to the upper central, the lower incisor has just two semi-landmarks, rather than four, with the corners of the incisal edge fixed. As was seen in 5.6.1, the semi-landmarks and chords are chosen to represent the variation in the locations of landmarks along the outlines of these teeth, evident in the plots of the within-case PC's in fig. 4.6.

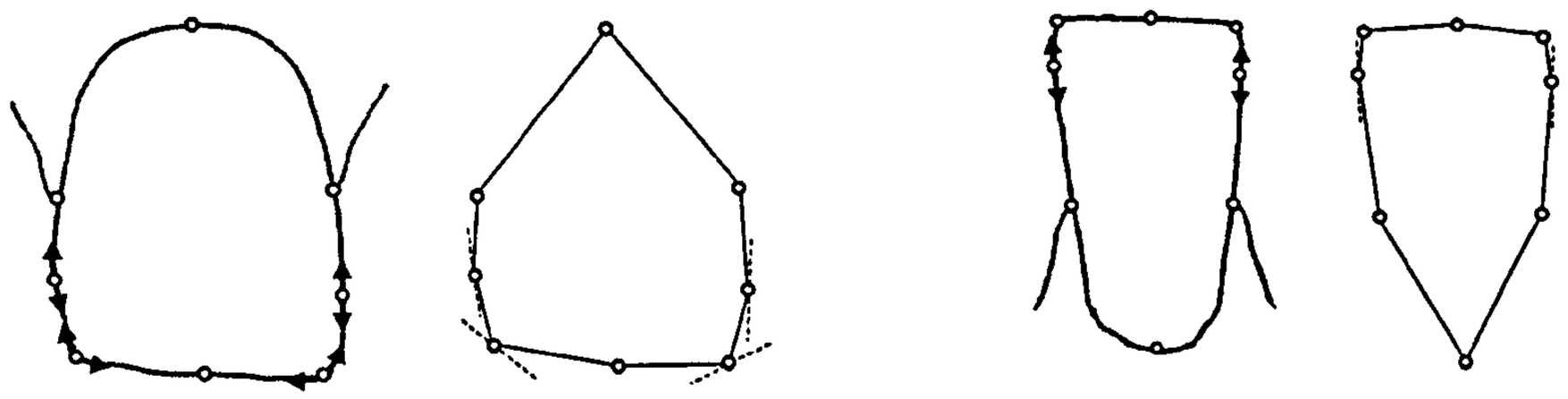


Fig 9.1: (Grey arrowed lines) Directions of landmark location inconsistencies around/along buccal outlines. (Dashed lines) unwanted variation represented as 'escribed chords'; lines through each chosen semi-landmark, parallel to that joining its immediate neighbours.

9.2.1.2 Occlusal surfaces

For the occlusal surfaces, patterns of within-case variation in the Procrustes fits suggested orientation differences in the bucco-lingual direction as a common cause of the poor reliability figures in this view. Variation in the position of the incisal or cusp edge and MD endpoints along the bucco-lingual dimension (relative to the BL endpoints) was most evident in the largest PC's of within case variation for the upper central incisor (73%) and lower canine (54%). Orientation inconsistencies were also

labial cusp width, fissure landmarks and (predominantly) the lingual and labial cusp landmarks varied with relative to the BL endpoints. The first PC for this surface also suggested operator inconsistencies in the positioning of the MD width, irrespective of orientation differences. See fig. 4.7 rows 1, 3 and 2 respectively.

Again, this variation in shape may be represented as semi-landmarks with associated chord directions, based on other (fixed or semi-) landmark positions in a configuration (see fig 9.2). Since the upper central incisor and lower canine would have the same combination of semi-landmarks and fixed landmarks, representing similar patterns of unwanted orientation inconsistencies, we shall consider only the upper central incisor data set in our subsequent investigations. In addition, the upper first premolar surface will also be considered since this has very different shape features from the incisor, as well as different numbers of fixed and semi-landmarks.

Note that for the upper central incisor, orientation differences in the bucco-lingual direction can either be represented as movement in the incisal edge landmarks, relative to fixed BL endpoints or as movement in the BL endpoints relative to a fixed incisal edge, see fig 9.2(a). The semi-landmarks and chord directions may therefore be set to represent either of the two options: as three semi-landmarks along the incisal edge, with the BL landmarks fixed; or as two semi-landmarks at the ends of the BL width, with those along the incisal edge fixed. In either case the chord movement is in directions defined by the BL endpoints. Similarly, orientation effects on the landmarks of the upper first premolar can be represented in two different ways, as shown in fig. 9.2(b). Here the cusp tips, fissures, corner landmarks and mesio-distal width all move relative to the BL landmarks (or vice-versa), as the occlusal surface is orientated differently in the BL direction. As with the incisal edge in fig. 9.2(a), these features are closer to the camera than the BL landmarks when imaging and so their locations will move more relative to the BL endpoints as the orientation changes in the BL direction. Which representation to use however, makes no difference as each set of fixed and semi-landmarks based on the same chord directions is simply a reparameterisation of the other. For each surface we shall proceed using the first of the two options described.

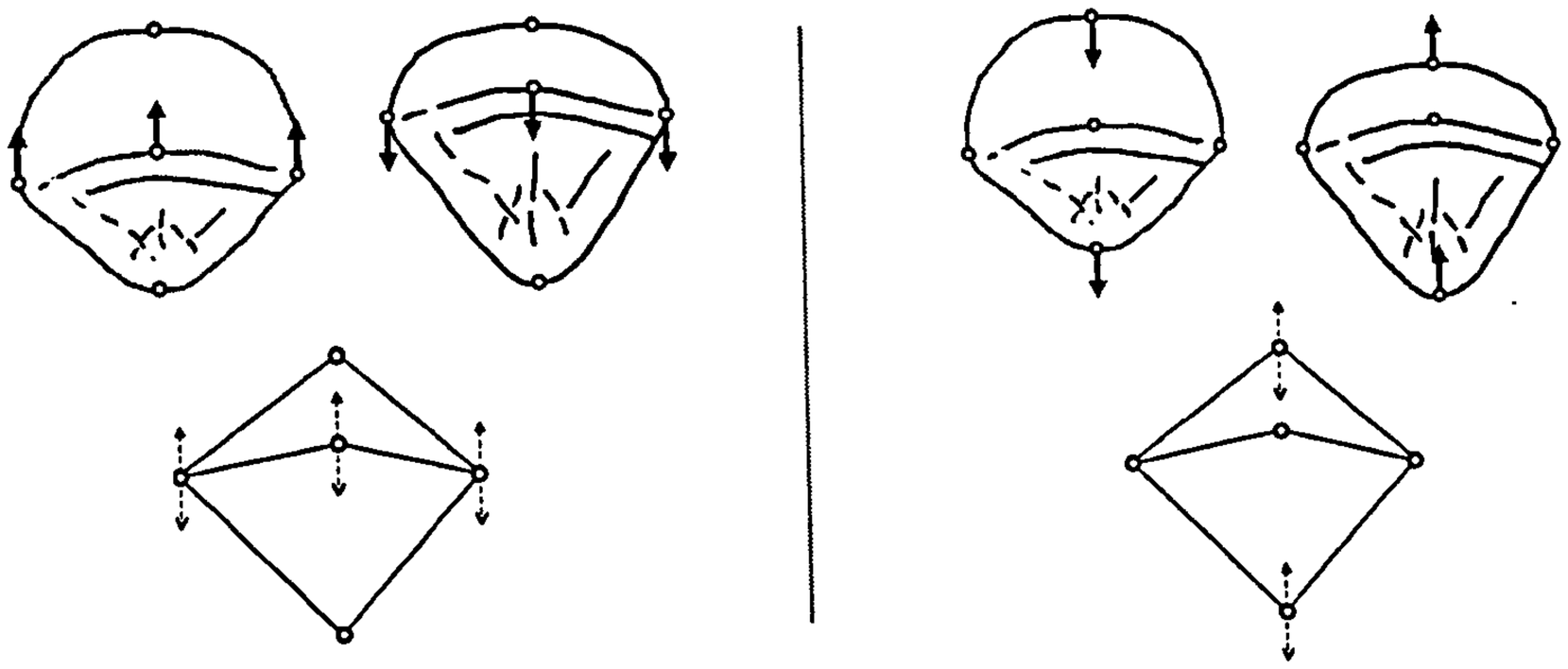


Fig 9.2(a) Alternative representations of orientation effects on occlusal landmarks of upper central incisor

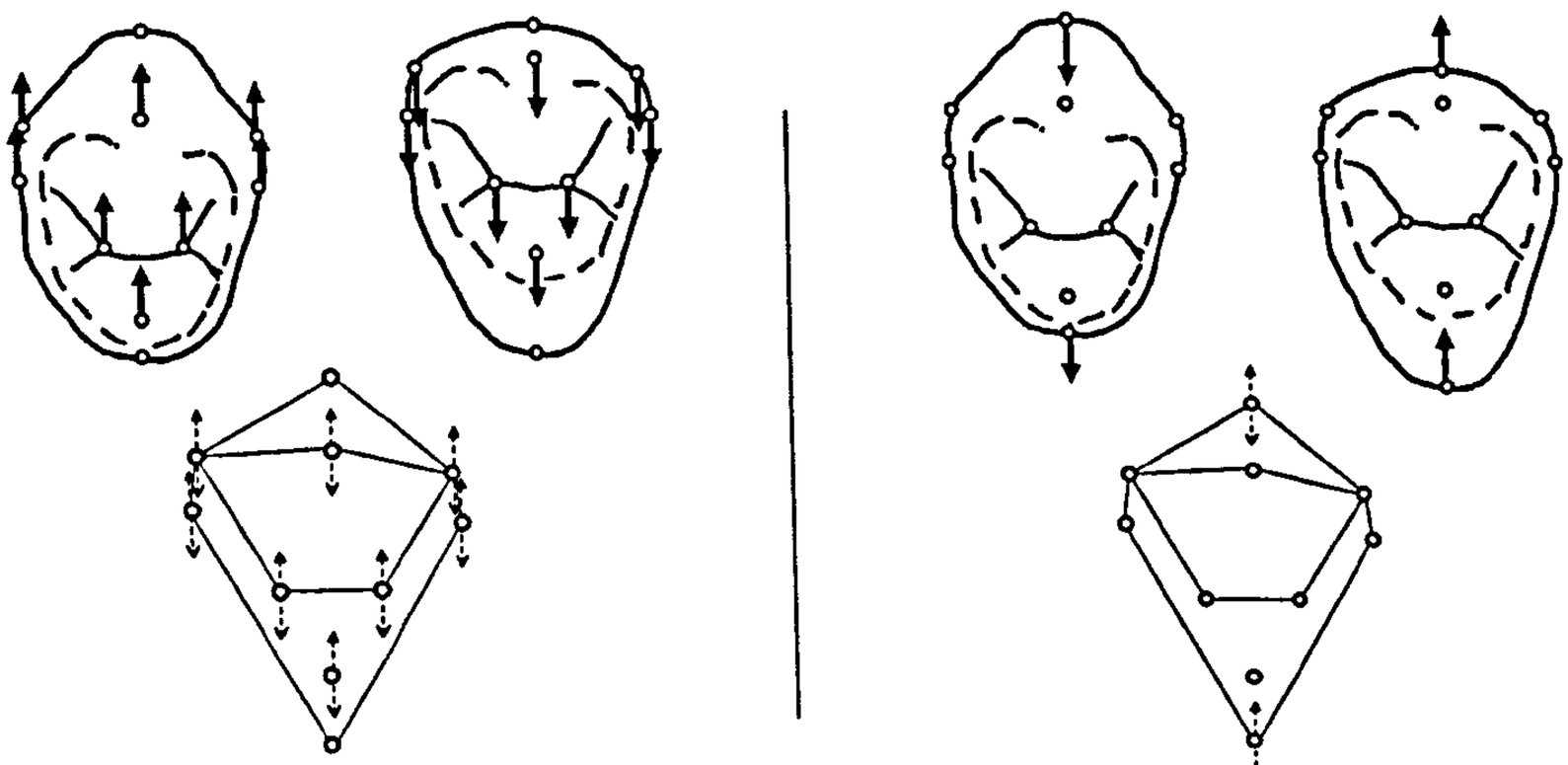


Fig 9.2(b) Alternative representations of orientation effects on occlusal landmarks of upper first premolar

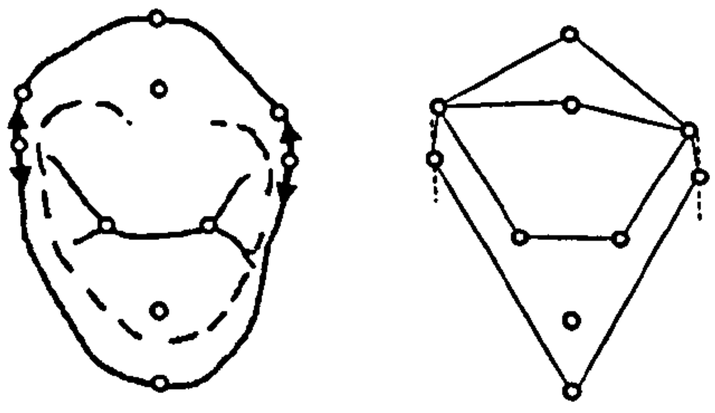


Fig 9.3: Variation in the position of the MD width expressed as semi-landmarks and chord directions of allowed movement

Aside from the orientation inconsistencies described above, the analysis of section 4.5.3.2 for the upper first molar, identified differences in the positioning of the MD endpoints as the main source of variation due to operators' inconsistencies, irrespective of orientation differences, as briefly mentioned above; the first PC accounts for 39% of within-case variation in the Procrustes fits. We shall therefore also represent this

variation by way of semi-landmarks and chords in the direction of the BL width, as shown in fig.9.3, when investigating the new semi-landmark methods on this surface.

9.2.1.3 Constrained semi-landmark movement (occlusal surface applications)

For the upper central incisor, it makes sense to constrain the movement of the semi-landmarks along the incisal edge so that the shape of the edge is always retained. One would expect that differences in the orientation of this surface in the BL direction, prior to imaging, leaves the configuration of the central, mesio and distal landmarks in fig 9.2(a) unaffected. (Recall that on this surface the ends of the MD width are at the ends of the incisal edge). This was also evident in fig 4.7 (top left), the plot of the largest principal component of within-case variation in the Procrustes fits. Any movement in the three incisal edge landmarks along the chords defined in fig 9.2(a) will therefore be constrained to be of equal amounts and in the same direction. Although allowing each landmark to be free to move on its own would allow us to remove more variation in their positions, we would lose the important information we have on the shape of incisal edge. It is only variation due to orientation effects we wish to remove, not actual differences in shape.

Recall that on each iteration, the new version of a configuration Y is always given by:

$$Y^{new} = vec^{-1}(vec(Y^0) - U\lambda), \tag{9.1}$$

where Y^0 denotes the nominal positions of the landmarks (after GPA registration to the current estimate of mean shape $\hat{\mu}$). A sublist $j_{(l)}$, l=1,..., L of the j=1,..., landmarks are then allowed to move away from these positions along pre-specified unit chord directions $u_{j_{(l)}} = (u_{jx_{(l)}}, u_{jy_{(l)}})^T$, where the $u_{j_{(l)}}$ themselves depend on positions of the landmarks in Y^0. U is a $2k \times 2$ matrix in which the $(j_{(l)}, j')$th entry is $u_{x_{(l)}}$ and the $(k+j_{(l)}, j')$th entry is $u_{y_{(l)}}$ and which is zero elsewhere. The vector $\lambda=(\lambda_1,...,\lambda_L)^T$ is a vector of unknown scalars, representing the distances moved along the directions, which are determined by the optimising the chosen objective function/criterion between $vec^{-1}(vec(Y^0) - U\lambda)$ and reference configuration T (the current mean $\hat{\mu}$).

The formulae for calculating λ and hence the new positions of the semi-landmarks, can be easily modified to accommodate the constraint described above, by re-writing (9.1) accordingly. For the occlusal surface of the upper central incisor, with landmarks 1,2 and 5 (as defined in fig 4.3(b)) chosen as the semi-landmarks, the usual vector of unknowns $(\lambda_1, \lambda_2, \lambda_3)^T$ is replaced by $(1 \quad 1 \quad 1)^T(\lambda_A)$, with only scalar λ_A to be determined. That is,

$$vec(Y^0) - U\lambda = \begin{pmatrix} y^0_{1x_{(1)}} \\ y^0_{2x_{(2)}} \\ y^0_{3x} \\ y^0_{4x} \\ y^0_{5x_{(3)}} \\ y^0_{2y_{(1)}} \\ y^0_{2y_{(2)}} \\ y^0_{3y} \\ y^0_{4y} \\ y^0_{5y_{(3)}} \end{pmatrix} - \begin{pmatrix} u_{x_{(1)}} & 0 & 0 \\ 0 & u_{x_{(2)}} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & u_{x_{(3)}} \\ u_{y_{(1)}} & 0 & 0 \\ 0 & u_{y_{(2)}} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & u_{y_{(3)}} \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix} = \begin{pmatrix} y^0_{1x_{(1)}} - u_{x_{(1)}}\lambda_1 \\ y^0_{2x_{(2)}} - u_{x_{(2)}}\lambda_2 \\ y^0_{3x} \\ y^0_{4x} \\ y^0_{5x_{(3)}} - u_{x_{(3)}}\lambda_3 \\ y^0_{2y_{(1)}} - u_{y_{(1)}}\lambda_1 \\ y^0_{2y_{(2)}} - u_{y_{(2)}}\lambda_2 \\ y^0_{3y} \\ y^0_{4y} \\ y^0_{5y_{(3)}} - u_{x_{(3)}}\lambda_3 \end{pmatrix}$$

becomes

$$\begin{pmatrix} y^0_{1x_{(1)}} \\ y^0_{2x_{(2)}} \\ y^0_{3x} \\ y^0_{4x} \\ y^0_{5x_{(3)}} \\ y^0_{2y_{(1)}} \\ y^0_{2y_{(2)}} \\ y^0_{3y} \\ y^0_{4y} \\ y^0_{5y_{(3)}} \end{pmatrix} - \begin{pmatrix} u_{x_{(1)}} & 0 & 0 \\ 0 & u_{x_{(2)}} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & u_{x_{(3)}} \\ u_{y_{(1)}} & 0 & 0 \\ 0 & u_{y_{(2)}} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & u_{y_{(3)}} \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} (\lambda_A) = \begin{pmatrix} y^0_{1x_{(1)}} - u_{x_{(1)}}\lambda_A \\ y^0_{2x_{(2)}} - u_{x_{(2)}}\lambda_A \\ y^0_{3x} \\ y^0_{4x} \\ y^0_{5x_{(3)}} - u_{x_{(3)}}\lambda_A \\ y^0_{2y_{(1)}} - u_{y_{(1)}}\lambda_A \\ y^0_{2y_{(2)}} - u_{y_{(2)}}\lambda_A \\ y^0_{3y} \\ y^0_{4y} \\ y^0_{5y_{(3)}} - u_{x_{(3)}}\lambda_A \end{pmatrix}. \qquad (9.2)$$

Hence the formulae (5.60), (5.59) and (6.81) for determining vector λ using the bending energy (BE), nearest point (NP) or full Procrustes (FP) criteria are all easily modified by simply replacing U with $U(1 \quad 1 \quad 1)^T$ and reducing the number of scalars in λ to just one, so that the semi-landmarks always all move by the same distance in the same direction.

Similarly, for the upper first pre-molar we would expect that any differences in the positions of the cusp tips, fissure pits, ends of the labial cusp diameter and MD width due to orientation differences in the BL direction would also be in the same directions. Here however we would not expect these differences to all be of the same size. For

example, as the occlusal surfaces is tilted beneath the camera in the BL direction, we would expect that the fissure pits and MD endpoints would not move by as much (relative to the BL endpoints) as the cusp tips, since these are nearer to the camera. One could also argue on the same basis that the lingual cusp will not move by as much as the labial cusp, due to their differences in height. In addition, the labial cusp diameter and cusp tip landmarks could be constrained in the same way as the incisal edge of the upper central incisor above, so that the triangulation of these landmarks is always retained.

To include these features as constraints on the semi-landmark movement, we can use a similar reformulation for obtaining Y^{new} as above, with just one scalar to be determined and a series of proportional weights instead of a column of 1's. For example, if the labial cusp tip and cusp diameter landmarks move by a distance λ_A along the unit BL direction, then the lingual cusp tip could be constrained to move by (say) 0.8 of this amount and the fissure pits by (say) 0.6 of λ_A. These landmarks are approximately 80% and 60% of the way from the base of the tooth to the labial cusp tip and so as the tooth is tilted in the BL direction (pivoted at the base), these landmarks will move roughly 0.8 and 0.6 of the distance moved by the labial cusp tip and cusp diameter landmarks in the BL direction.

Differences in the MD landmarks could also be constrained in the same way, to be some proportion of the change λ_A in the labial cusp tip along the BL direction. However, the results of the within case PCA for this surface in section 4.5.3.2 suggested that most unwanted variation in the MD endpoints was due to operator inconsistencies in their location once an image had been obtained (PC1, 39%), independent of the variation due to orientation differences affecting these and the other landmarks (PC2). It would therefore make sense to include the MD endpoints as ordinary unconstrained semi-landmarks, allowing them to move independently of the incisal edge landmarks and also of each other, since the errors in the identification of the MD endpoints on screen would not necessarily be related.

The formulations for calculating the new positions of the semi-landmarks are modified to accommodate these constraints and features as below. For the occlusal surface of the

upper first pre-molar with landmarks 1 and 2 (MD endpoints), 5 and 6 (fissure pits), 7 and 8 (lingual and labial cusp tip), 9 and 10 (end of labial cusp diameter) defined as in fig 4.3(b), chosen as the eight semi-landmarks, the usual vector of unknowns $\lambda=(\lambda_1,...,\lambda_8)^T$ is replaced by a matrix of weights and vector $\lambda=(\lambda_1,\lambda_2,\lambda_A)^T$ of three parameters to be determined, so that $vec(Y^0)-U\lambda$ becomes:

$$\begin{pmatrix} y^0_{1x_{(1)}} \\ y^0_{2x_{(2)}} \\ y^0_{3x} \\ y^0_{4x} \\ y^0_{5x_{(3)}} \\ y^0_{6x_{(4)}} \\ y^0_{7x_{(5)}} \\ y^0_{8x_{(6)}} \\ y^0_{9x_{(7)}} \\ y^0_{10x_{(8)}} \\ y^0_{1y_{(1)}} \\ y^0_{2y_{(2)}} \\ y^0_{3y} \\ y^0_{4y} \\ y^0_{5y_{(3)}} \\ y^0_{6y_{(4)}} \\ y^0_{7y_{(5)}} \\ y^0_{8y_{(6)}} \\ y^0_{9y_{(7)}} \\ y^0_{10y_{(8)}} \end{pmatrix} - \begin{pmatrix} u_{x_{(1)}} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & u_{x_{(2)}} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & u_{x_{(3)}} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & u_{x_{(4)}} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & u_{x_{(5)}} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & u_{x_{(6)}} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & u_{x_{(7)}} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & u_{x_{(8)}} \\ u_{y_{(1)}} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & u_{y_{(2)}} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & u_{y_{(3)}} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & u_{y_{(4)}} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & u_{y_{(5)}} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & u_{y_{(6)}} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & u_{y_{(7)}} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & u_{y_{(18} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0.6 \\ 0 & 0 & 0.6 \\ 0 & 0 & 0.8 \\ 0 & 0 & 1 \\ 0 & 0 & 0.8 \\ 0 & 0 & 0.8 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_A \end{pmatrix}. \tag{9.3}$$

Hence the formulae (5.60), (5.59) and (6.81) for obtaining vector λ, using the various criteria, are again easily modified by reducing the number of scalars to be determined in λ to three and replacing U with:

$$U\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0.6 \\ 0 & 0 & 0.6 \\ 0 & 0 & 0.8 \\ 0 & 0 & 1 \\ 0 & 0 & 0.8 \\ 0 & 0 & 0.8 \end{pmatrix}.$$

9.2.2 Data considerations and choice of summary criteria

As with the gum simulation, before any analysis is performed we first consider how exactly the multiple representations of different teeth should be used in the investigation and what summary measures are to be calculated and compared in order to evaluate the success of each method.

In order for us to establish whether or not we have reduced variation due to operator inconsistencies in the Procrustes fits, a reduction in the 'within-case' variation (between operators and in errors) will clearly be desirable. However, we must also be aware that doing so may also be at the expense of removing important biological variation between cases in these directions. While the removal of operator inconsistencies between representations from different cases is still desirable, it will only be beneficial if the patterns of variation removed do not also coincide with important biological differences in shape between cases. The success of using the semi-landmark techniques will therefore depend upon the extent to which this is so and how subsequent inferences and conclusions based on the final Procrustes registered configurations will be affected.

9.2.2.1 Data choice

Recall that for each of the buccal and occlusal surfaces considered in chapter 4, the data consists of the same four operators' representations of each of up to 20 teeth from different patients. As with the gum simulation we could run each method either:

a) on representations of the same tooth
b) on each single operator's set of representations, or
c) on the entire set of repeated configurations from different cases.

Since we are interested in demonstrating a reduction in variation due to operator inconsistencies so that different operators' representations of the same tooth are regarded as the same shape, option (b) is clearly less appealing, despite being closest to the situation we would have in practice. It would be difficult to calculate summary measures to show a reduction 'within-cases' if there is only one representation per case.

If we consider sets of configurations from one case at a time (option a) the resulting semi-landmarks will only have been determined and defined relative to each case mean, as noted for this option in the simulation study (section 8.5.2.1). In any application, the procedures would be run on representations from different cases and (by definition) the semi-landmark positions need to be determined with respect to the mean of the different cases so we need to ensure that different cases are used in the investigations here. As noted in 8.5.2.1, this also ensures that there is realistic variation in the fixed landmarks of the configurations, since their arrangement relative to the reference mean shape plays a major role in the determination of new semi-landmark positions with each method (in either the prior alignment of the configurations or as part of the optimisation criterion).

In light of these issues, for each tooth surface we again run each semi-landmark routine on the entire set of multiple representations of different patients' teeth (option c). By allowing landmarks on multiple representations of different cases to move relative to an overall mean, this allows variation to be removed within and between cases.

9.2.2.2 Choice of summary measures

For each application, summary measures are again based on the Euclidean sums of squares of the final Procrustes registered configurations of the m representations of each of n cases to their overall estimated mean shape $\hat{\mu}_{SL}$, denoted as $X_{im}^{P_{SL}}$, i=1,…,n, m=1,..4. Again the iterative method of registration is to be used throughout so that the variation in the Procrustes registered fits, i.e.

$$\sum_{i=1,}^{n}\sum_{m=1,}^{4}\left\|X_{im}^{P_{SL}}-\hat{\mu}_{SL}\right\|^2=\sum_{i=1,}^{n}\sum_{m=1,}^{4}d_F^2(X_{im}^{P_{SL}},\hat{\mu}_{SL})=RSS(\hat{\mu}_{SL})=RSS=EucSS(total),$$

where $\hat{\mu}_{SL}=\frac{1}{4n}\sum_{i=1,}^{n}\sum_{m=1,}^{4}X_{im}^{P_{SL}}$, may be separated into sums of Euclidean norms 'between cases' and 'within cases' (comprising variation 'between operators' and 'due to errors'), using the identity described in 4.4.2, with

$$EucSS(total)=\sum_{i=1}^{n}\sum_{m=1}^{4}\left\|\overline{X}_i^{P_{SL}}-\overline{X}^{P_{SL}}\right\|^2+\sum_{i=1}^{n}\sum_{m=1}^{4}\left\|X_{im}^{P_{SL}}-\overline{X}_i^{P_{SL}}\right\|^2=EucSS(cases)+EucSS(within)$$

and $$\sum_{i=1}^{n}\sum_{m=1}^{4}\left\|X_{im}^{P_{SL}} - \bar{X}_{i}^{P_{SL}}\right\|^2 = \sum_{i=1}^{n}\sum_{m=1}^{4}\left\|\bar{X}_{m}^{P_{SL}} - \bar{X}^{P_{SL}}\right\|^2 + \sum_{i=1}^{n}\sum_{m=1}^{4}\left\|X_{im}^{P_{SL}} - \bar{X}_{i}^{P_{SL}} - \bar{X}_{m}^{P_{SL}} + \bar{X}^{P_{SL}}\right\|^2$$

$$= EucSS(operators) + EucSS(error).$$

This also then allows calculation of overall reliability scores, using (4.5) and (4.12).

In the gum simulation, the directions of variation to be removed were uninteresting both within and between cases. Here however, we hope that most of what is removed is 'within' cases, rather than 'between' as described above. For this reason we need to examine any reduction of within-case variation relative to the reduction in total variation and/or between case variation. A recalculation of the reliability figures (or any other proportional measure) should enable us to establish whether any reduction in total variation has been mainly within cases compared to between.

As well as comparing Euclidean sums of squares and reliability measures between the final Procrustes registered data from the different semi-landmark methods, here we can also compare them to the same measures obtained from the GPA registration of the same datasets (with no semi-landmark movement) used in Chapter 4, for which the original overall reliability scores were presented in table 4.3.

Finally, note that as with the gum simulation we must again be cautious that simply minimising the within case variation in shape or improving the reliability figures may not always produce desirable results. A small within-case measure of variation in shape will result if all the representations of the same tooth end up as the same unrealistic shape (and the unrealistic case means may also result in increased variation between cases). In either instance, the proportionally larger variance between cases will give the impression of an improvement to the reliability figures. Plots of the final Procrustes fits will again be examined to ensure that realistic, plausible tooth shapes are produced.

9.2.3 Gum effects

Since we are investigating the different methods' success in addressing issues arising from the operator inconsistencies, the effect of any variation in shape produced by differences in the positions of a patient's gum tissue will initially be ignored, so that we

may isolate how well the different methods perform when dealing with these particular problems. For the reliability datasets of chapter 4 (which we use again here), any variation in the gingival landmark positions is only present 'between cases', since the data were obtained from stone casts of patients' teeth. Any variation in the gingival landmarks 'within cases' would be due only to inconsistencies in their location by different operators and this was found to be negligible. Operators reported no difficulties in identifying gingival landmarks positions and the analysis of 4.5.3.1 revealed nothing to suggest that gingival landmarks were a particular problem for operators to identify consistently.

Having obtained results for the operator inconsistency problems described in 9.2.1.1 and 9.2.1.2, we then consider including additional semi-landmarks representing the effect of differences in the position of the gingival tissue of different cases. For the buccal surface of the upper central incisor, we will consider how the new semi-landmark methods perform when we also include the three gingival landmarks as semi-landmarks, as in Chapter 8, using the same chord directions as originally described in 5.6.2. See fig. 9.4 below. Note that in the gingival simulation study, the positions of the MD endpoints and corners of the incisal edge were simulated in identical positions on each 'case'. In contrast to the current datasets there was no 'within-case' variation in these positions to be removed, only 'between cases'.

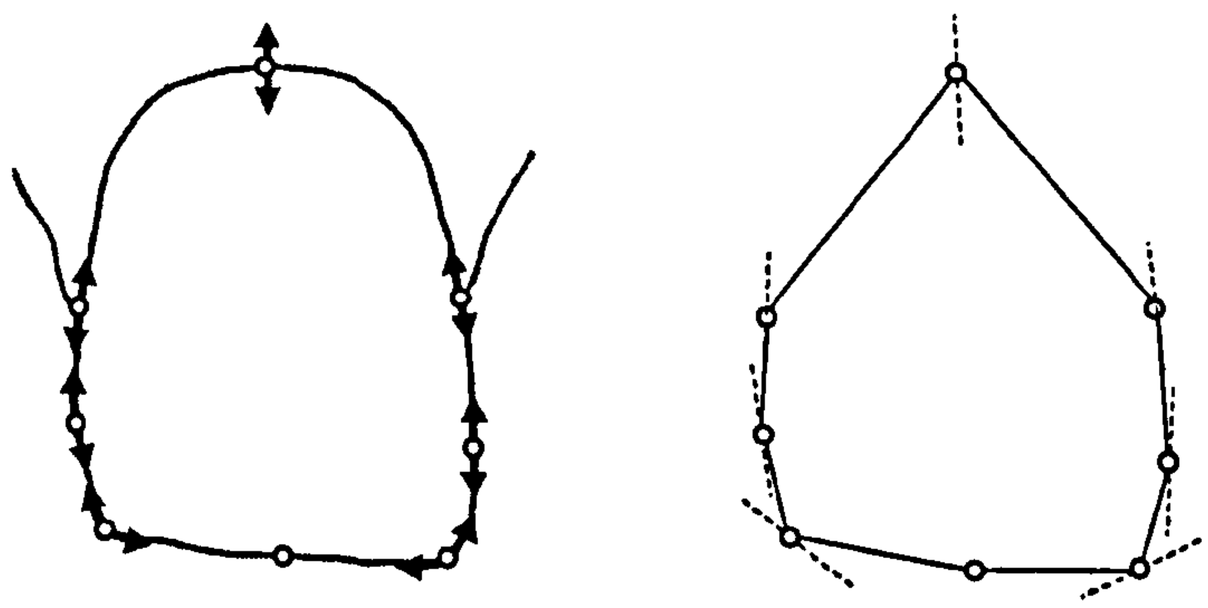


Fig 9.4: Semi-landmark representation of variation in the gingival landmarks due to differences in the position of a patient's gum as well as operator inconsistencies in the positions of the MD endpoints and incisal corners

For the occlusal surface of the upper central incisor, there is also additional variation between cases due to differences in patients' gum position at the lingual landmark at the

end of the bucco-lingual width, as shown in fig.9.5 (left) below. This may be represented as an additional semi-landmark with direction of (allowed) movement along a line through the lingual landmark, in the direction of the BL dimension as shown in fig 9.5(right). There is no similar problem at the labial end of the BL width however, since the landmark here corresponds to the most prominent part of the crown (as this was the imaging criterion used), so as a result the gingival margin is always obscured.

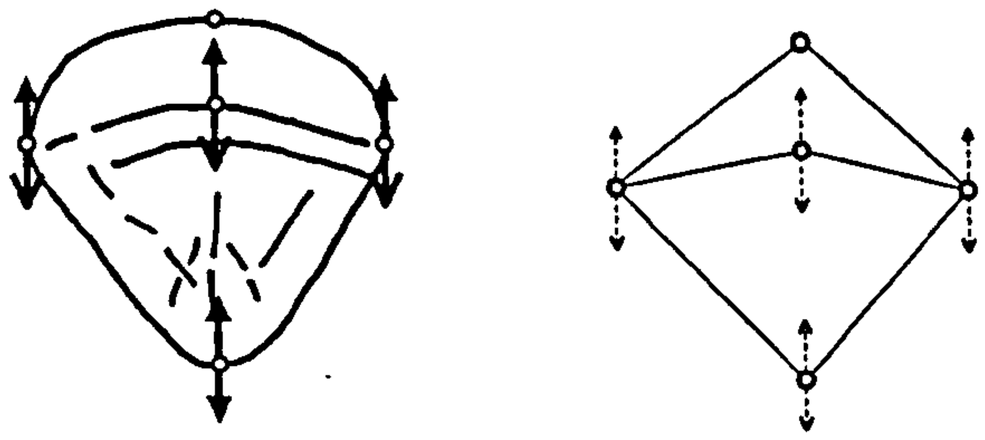


Fig 9.5: Variation in the lingual endpoint of the BL width due to differences in patients' gum position.

A further modification of equation (9.3) is required in order to accommodate the extra semi-landmark (landmark 3), as well as the semi-landmarks 1, 2 and 4 along the incisal edge. Note that any variation in the lingual BL landmark corresponding to differences in the position of a patient's gingival tissue will be unrelated to the variation in the three constrained semi-landmarks along the incisal edge. Consequently, we will require another scalar parameter to be determined, representing the distance moved along the BL direction by landmark 3 and to do this we use:

$$\begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_A \\ \lambda_B \end{pmatrix} \text{ instead of } \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} (\lambda_A)$$

in equation (9.3). For the upper first premolar there are no such problems. The curved shape of the labial and lingual sides of these teeth ensure that the gingival margins are always obscured and therefore will not affect the positions of the bucco-lingual landmarks.

9.3 Results: Operator inconsistencies

For each of the applications and tooth surfaces considered, the main table of results presented in each section details the summary measures *EucSS*(*total*), *EucSS*(*cases*), *EucSS*(*within*), *EucSS*(*operators*) and *EucSS*(*error*) listed in 9.2.2.2, based on the original (with all landmarks fixed) and final sets of Procrustes fits obtained by each of the semi-landmark methods. In the final row of each table is the overall reliability figure $R(C_{All})$, defined in 4.4.2.2 and calculated using the sums of Euclidean norms defined above. In order to examine how much each procedure has changed the mean shape, the tables also contain the full Procrustes distances $d_F(\hat{\mu}_{SL}, \hat{\mu}_0)$ between the final estimated mean shapes and the initial (original) estimates, obtained when all landmarks are fixed. This is a check that no bias is introduced by any of the methods. Finally, in addition to the measures described above, plots of the initial and final Procrustes fits obtained by each method are also presented.

The first table for each application however, details the iteration history from running each method for a fixed number of iterations, as for the gum simulation. Here we report values of *RSS* and ΔRSS for the first 15 or 20 iterations of each method. Again the number of decimal places quoted is to the decimal place still changing at the last reported iteration. As before this enables us to identify where the change in *RSS* slows, i.e. on which iteration each process enters its 'uninformative regime' or 'slow manifold', as indicated by the shaded boxes and the levels of convergence that are achievable.

Each semi-landmark process was then able to be run for a pre-specified number of iterations (corresponding to an acceptable convergence level), so as to stop the configurations becoming unrealistic from running each routine for too long. So that the results are comparable, the same convergence criterion was used for each semi-landmark method, this being the strictest possible value for which each procedure still avoids entering its 'slow manifold'. The resulting *RSS* and final ΔRSS are then indicated in bold. This meant that a change in *RSS* of less than 0.001 was usually able to used for each of the three semi-landmark methods, in each application. Based on the discussion of 5.5.6, an acceptable convergence criterion for ΔRSS when running each

semi-landmark method on $n \approx 80$ configurations would be in the region of $80 \times (2 \times 10^{-5}) = 0.0016$ and so use of $\Delta RSS < 0.001$ is clearly adequate.

For each application, the Full Procrustes (FP) criterion was again found to take fewest iterations, but much longer, in terms of time, than the bending energy (BE) or nearest point (NP) methods to attain the same level of convergence. In addition, in all applications the final *RSS* values for the Full Procrustes and nearest point methods were always almost identical.

9.3.1 Buccal surfaces

For the buccal surfaces of the upper and lower central incisors, our aim was to remove the differences in shape due to operator inconsistencies at the MD endpoints and (for the upper central incisor) at the corners of the incisal edge.

9.3.1.1 Upper central incisor

Table 9.1 below details the iteration history of the change in values of *RSS* , for each of the semi-landmark methods, with shaded/bold values indicating convergence at $\Delta RSS < 0.001$, as each process 'slows'.

Iteration	Bending energy		Full Procrustes distance		Nearest point	
	RSS	ΔRSS	*RSS*	ΔRSS	*RSS*	ΔRSS
Original	0.9231095		0.9231095		0.9231095	
1	0.6528806	0.2702290	0.5494251	0.3736845	0.5793010	0.3438086
2	0.6503416	0.0025390	0.5476594	0.0017657	0.5547829	0.0245181
3	**0.6501951**	**0.0001465**	**0.5468032**	**0.0008562**	0.5504770	0.0043059
4	0.6500662	0.0001289	0.5459452	0.0008581	0.5492745	0.0012026
5	0.6499412	0.0001250	0.5450796	0.0008656	**0.5487114**	**0.0005631**
6	0.6498156	0.0001256	0.5442063	0.0008733	0.5483099	0.0004015
7	0.6496897	0.0001259	0.5433251	0.0008812	0.5479536	0.0003563
8	0.6495633	0.0001263	0.5424358	0.0008893	0.5476110	0.0003426
9	0.6494366	0.0001267	0.5415382	0.0008976	0.5472733	0.0003377
10	0.6493096	0.0001270	0.5406322	0.0009060	0.5469382	0.0003351
11	0.6491821	0.0001275	0.5397175	0.0009147	0.5466053	0.0003330
12	0.6490541	0.0001280	0.5387940	0.0009235	0.5462743	0.0003309
13	0.6489259	0.0001282	0.5378614	0.0009326	0.5459454	0.0003289
14	0.6487971	0.0001288	0.5369196	0.0009419	0.5456186	0.0003268
15	0.6486679	0.0001291	0.5359682	0.0009514	0.5452938	0.0003248
16	0.6485386	0.0001294	0.5350071	0.0009611	0.5449711	0.0003227
17	0.6484085	0.0001300	0.5340360	0.0009711	0.5446505	0.0003206
18	0.6482783	0.0001302	0.5330546	0.0009814	0.5443319	0.0003186
19	0.6481476	0.0001307	0.5320627	0.0009919	0.5440153	0.0003166
20	0.6480164	0.0001312	0.5310600	0.0010027	0.5437007	0.0003146

Table 9.1: Iteration history of *RSS* and ΔRSS values for different semi-landmark methods.

Fig 9.6 plots the original/initial (all landmarks fixed) and final sets of GPA registered configurations after using each of the three semi-landmark criteria. Table 9.2 details the summary measures obtained for each method based on each set of $4\times(n=19)$ configurations.

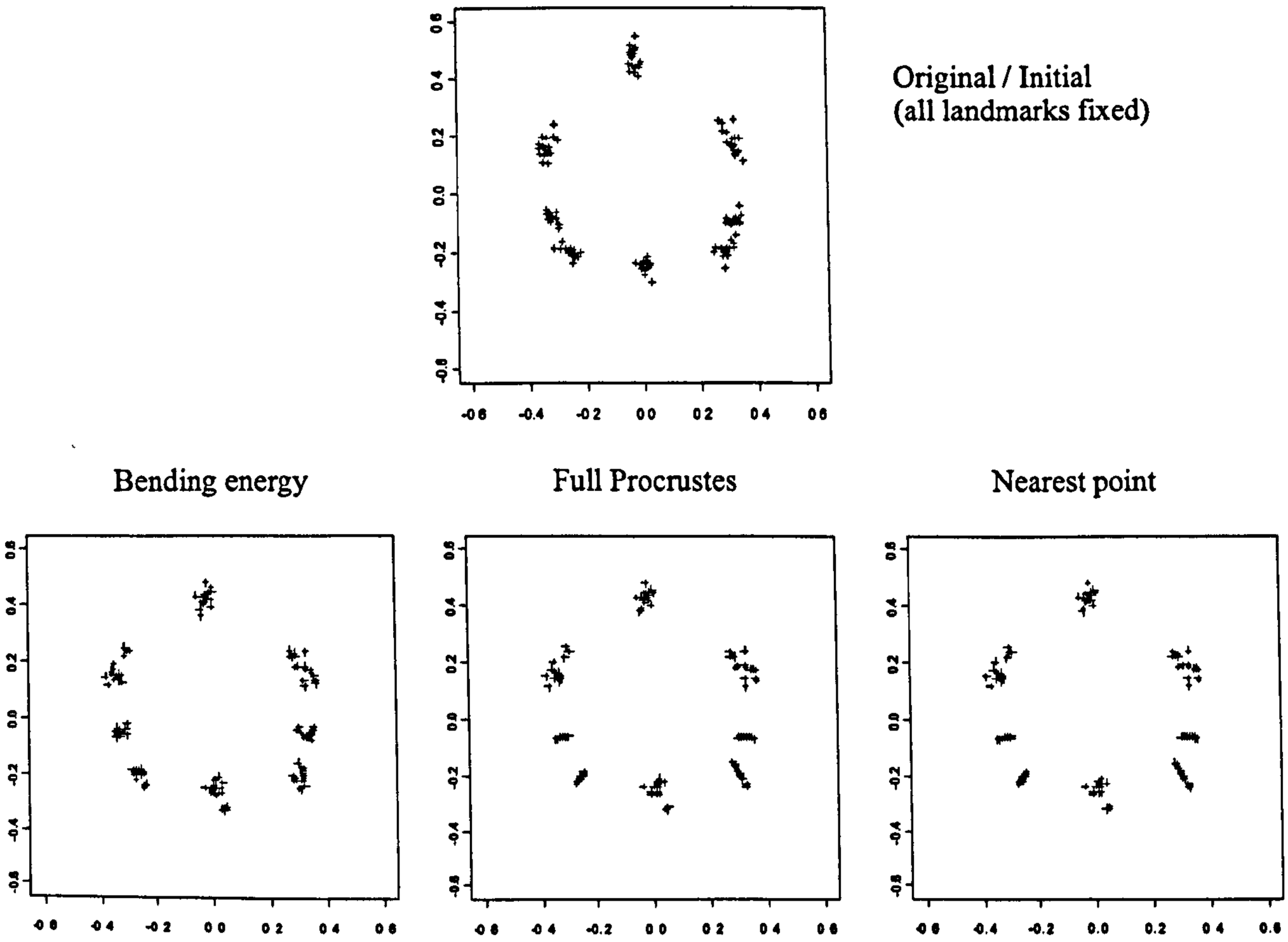


Fig 9.6: Procrustes registered configurations. (Top) Original/initial superimposition (all landmarks fixed). (Bottom) After running Semi-landmark procedure using different criteria.

	None (all fixed)	Bending energy	Full Procrustes	Nearest point
EucSS (cases)	0.564	0.493	0.412	0.414
EucSS (within)	0.359	0.157	0.135	0.135
EucSS (ops)	0.150	0.047	0.038	0.039
EucSS (errors)	0.209	0.109	0.096	0.097
EucSS (total)=*RSS*	0.923	0.650	0.547	0.549
Overall reliability	**0.522**	**0.698**	**0.690**	**0.691**
$d_F(\hat{\mu}_{SL}, \hat{\mu}_0)$		0.004	0.006	0.005

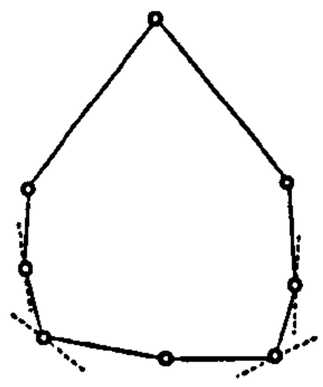

Upper central incisor

Table 9.2: Summary measures based on final Procrustes fits.

In comparison with the original GPA registration of the data, all three methods are successful in reducing the variation in the MD and corner landmarks in directions around the outline. For the FP and NP methods all remaining variation in these

positions is in directions perpendicular to the final directions of the escribed chords. The plots of the $X_{im}^{P_{SL}}$ for these methods appear almost identical, with only slight differences between the two sets of summary measures for these methods in table 9.2. On the basis of the argument in 5.5.6, the change in mean shape from the original Procrustes estimate is negligible, with all three methods ($d_F(\hat{\mu}_{SL}, \hat{\mu}_0)$<0.006), since this is of the same order as the change in a configuration's shape produced by a 1-pixel change to any one of its coordinates.

Notice in fig 9.6 however, that with each method there is a decrease in vertical variation at the gingival endpoint (top) of the LACC and increase at the incisal (bottom). There is also a tendency for the horizontal variation at each landmark to increase, albeit only slightly. We refer back to this in the discussion in 9.3.3.

Table 9.2 indicates similar reliability figures for all three of the semi-landmark methods, with an improvement of 0.18 compared to the figure of chapter 4 (i.e. in first column, obtained with all landmarks fixed). The reduction in within-case variation is slightly larger for the FP and NP methods, but this is accompanied by a larger decrease in the variation between cases (and overall) and so in the final Procrustes fits, the proportion of variation attributable to actual variation between cases is similar for all methods.

Fig 9.7 shows the original and final Procrustes fits obtained using each of the semi-landmark methods for operators' representations of three of the 19 cases. The three rows are plots for the three cases which had the highest total variance in component scores on the first two PC's of the within-case sums of squares and products (SSP) matrix, obtained from the original Procrustes fits, in 4.5.3. That is, the operators' representations of these three cases were found to have the largest within case variance along the first two PC's shown in the top row of fig 4.6 (along the directions which the semi-landmarks represent). Note that in each plot, each configuration is registered to the overall grand mean, not its case mean (since we are removing variation about the overall mean, not each case mean). Comparing the results of the semi-landmark methods with the original/initial GPA registration of each case, it is clear that the variance in the positions of the MD and corner landmarks on representations of the same case has been decreased in directions around the outline of each tooth as required,

with little to choose between the results of the different methods. Notice however, for the case in the third row, this is always at the expense of an increase in variation in the gingival landmark position at the top of the LACC, where the configurations matched to the overall mean shape, although the within-case *RSS* for this case has decreased.

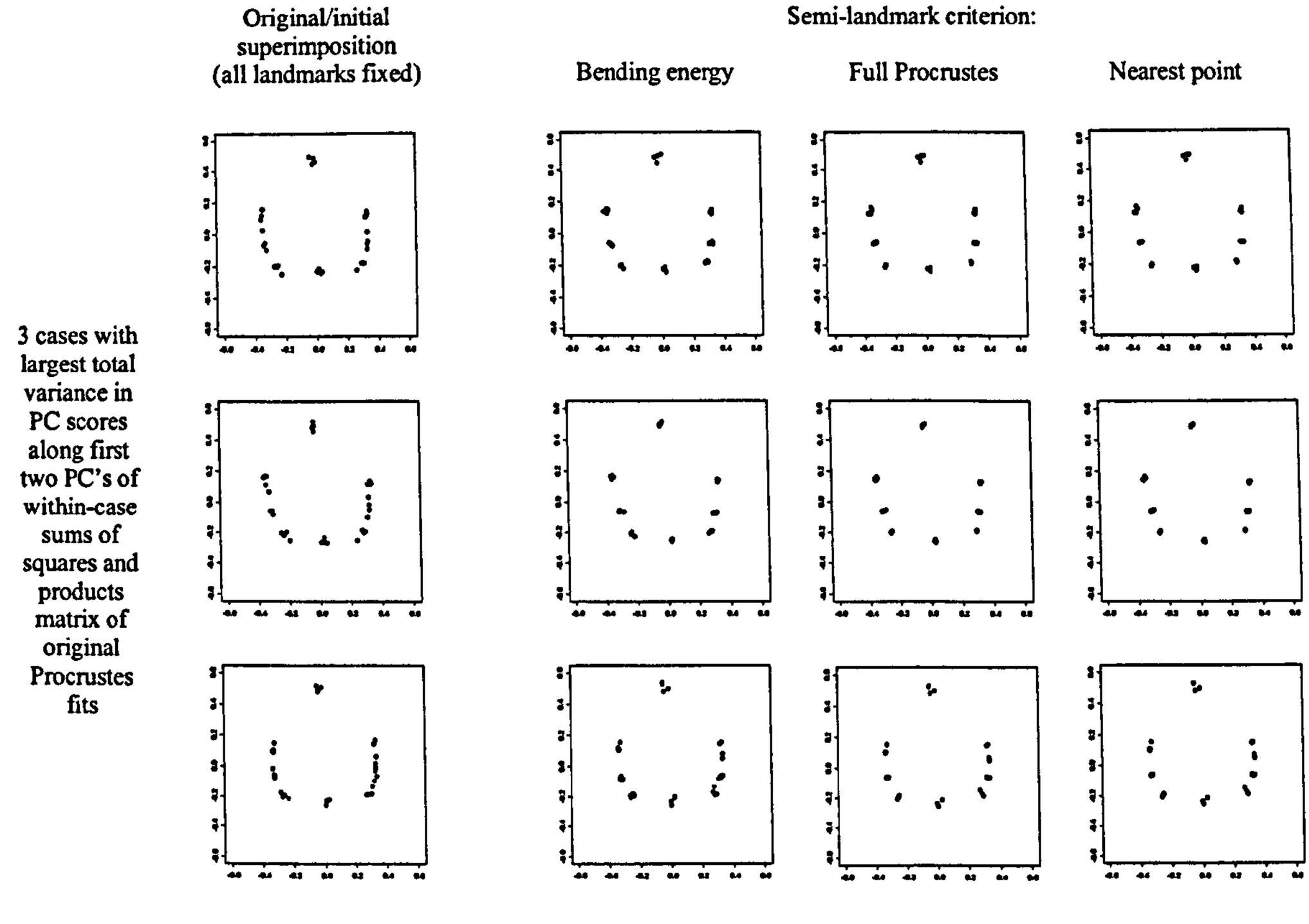


Fig 9.7: Procrustes fits of different operators' representations of the buccal surfaces of four upper central incisors, obtained using ordinary GPA registration and each of the semi-landmark methods.

9.3.1.2 Lower central incisor

Table 9.3 details the iteration history for each of the semi-landmark methods. Again, a criterion of $\Delta RSS < 0.001$ stops each process before it enters its 'uninformative regime', indicated by the bold/shaded values. Fig. 9.8 plots the original and final sets of GPA registered configurations after using each of the semi-landmark methods and table 9.4 details the summary measures based on each set of $4\times(n=19)$ configurations.

The resulting scatters of the final Procrustes fits in fig. 9.6 are again almost identical for the FP and NP criteria, with all remaining variation in shape at the semi-landmarks (MD endpoints) normal to the final directions of each configuration's escribed chords. The results using the BE criterion are again also similar, although there is clearly some

remaining variation in the directions of the final escribed chords as well as normal to them, which we discuss in 9.3.3. Notice also that since the changes in shape produced by the semi-landmark methods are small here, any effect on the patterns of variance at the other fixed landmarks is less noticeable than in the previous example.

Iteration	Bending energy		Full Procrustes		Nearest point	
	RSS	ΔRSS	RSS	ΔRSS	RSS	ΔRSS
Original	0.831048939		0.831048939		0.831048939	
1	0.683246295	0.147802644	0.655757520	0.175291418	0.667621480	0.163427458
2	**0.683046499**	**0.000199796**	**0.655392965**	**0.000364555**	0.656839173	0.010782308
3	0.683067860	-0.000021361	0.655025455	0.000367510	**0.656109281**	**0.000729892**
4	0.683090190	-0.000022331	0.654654197	0.000371258	0.656064735	0.000044546
5	0.683112556	-0.000022366	0.654279157	0.000375040	0.656063272	0.000001464
6	0.683134944	-0.000022389	0.653900303	0.000378855	0.656063263	0.000000008
7	0.683157358	-0.000022414	0.653517600	0.000382703	0.656062910	0.000000353
8	0.683179799	-0.000022440	0.653131016	0.000386584	0.656062410	0.000000500
9	0.683202263	-0.000022464	0.652740516	0.000390500	0.656061865	0.000000545
10	0.683224752	-0.000022489	0.652346067	0.000394449	0.656061306	0.000000559
11	0.683247266	-0.000022515	0.651947635	0.000398432	0.656060742	0.000000564
12	0.683269806	-0.000022540	0.651545186	0.000402449	0.656060174	0.000000567
13	0.683292371	-0.000022565	0.651138685	0.000406501	0.656059605	0.000000570
14	0.683314961	-0.000022590	0.650728098	0.000410587	0.656059033	0.000000572
15	0.683337577	-0.000022616	0.650313391	0.000414708	0.656058459	0.000000574

Table 9.3: Iteration history of RSS and ΔRSS values for different semi-landmark methods.

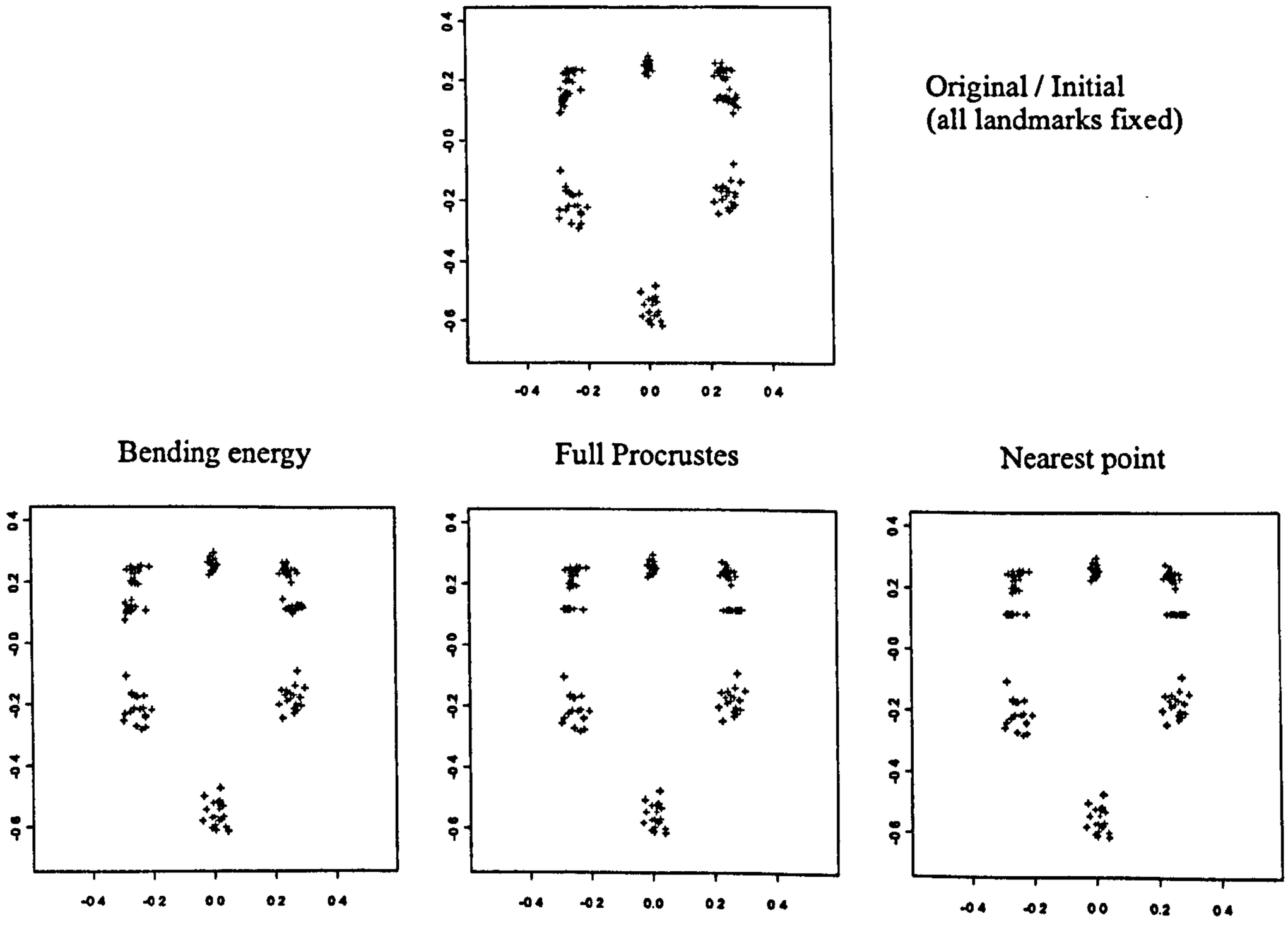


Fig 9.8: Procrustes registered configurations. (Top) Original/initial superimposition (all landmarks fixed). (Bottom) After running Semi-landmark procedure using different criteria.

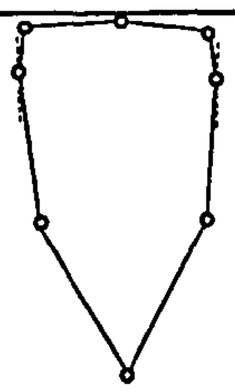
		None (all fixed)	Bending energy	Full Procrustes	Nearest point
 Lower central incisor	EucSS (cases)	0.595	0.524	0.524	0.524
	EucSS (within)	0.236	0.132	0.132	0.132
	EucSS (ops)	0.080	0.032	0.032	0.032
	EucSS (errors)	0.156	0.100	0.099	0.100
	EucSS (total)=*RSS*	0.831	0.656	0.655	0.656
	Overall reliability	**0.646**	**0.747**	**0.747**	**0.747**
	$d_F(\hat{\mu}_{SL}, \hat{\mu}_0)$		0.001	0.004	0.002

Table 9.4: Summary measures based on final Procrustes fits.

The summary measures in table 9.4 are almost identical for each method, with an improvement in the reliability figures of 0.1, with each method used, indicating that of the small amount of variation removed variation removed (i.e. the decrease in *RSS*), most of this has been 'within-cases'. Also, note again that the change in mean shape from the initial/original registration of the data is negligible ($d_F(\hat{\mu}_{SL}, \hat{\mu}_0)$<0.004).

Fig 9.9 shows the original and final Procrustes fits, obtained using each of the semi-landmark methods, for representations of the case with the highest variance in component scores on the first PC of the within-case SSP matrix obtained from the original Procrustes fits in 4.5.3, i.e. this case had the largest variance along the directions shown in the second row of fig 4.6. Clearly, the within-case variation in the positions of the MD endpoints is reduced for this case, with all three of the semi-landmark methods, although this is always at the expense of a slight increase in the variation of the gingival endpoint of the LACC.

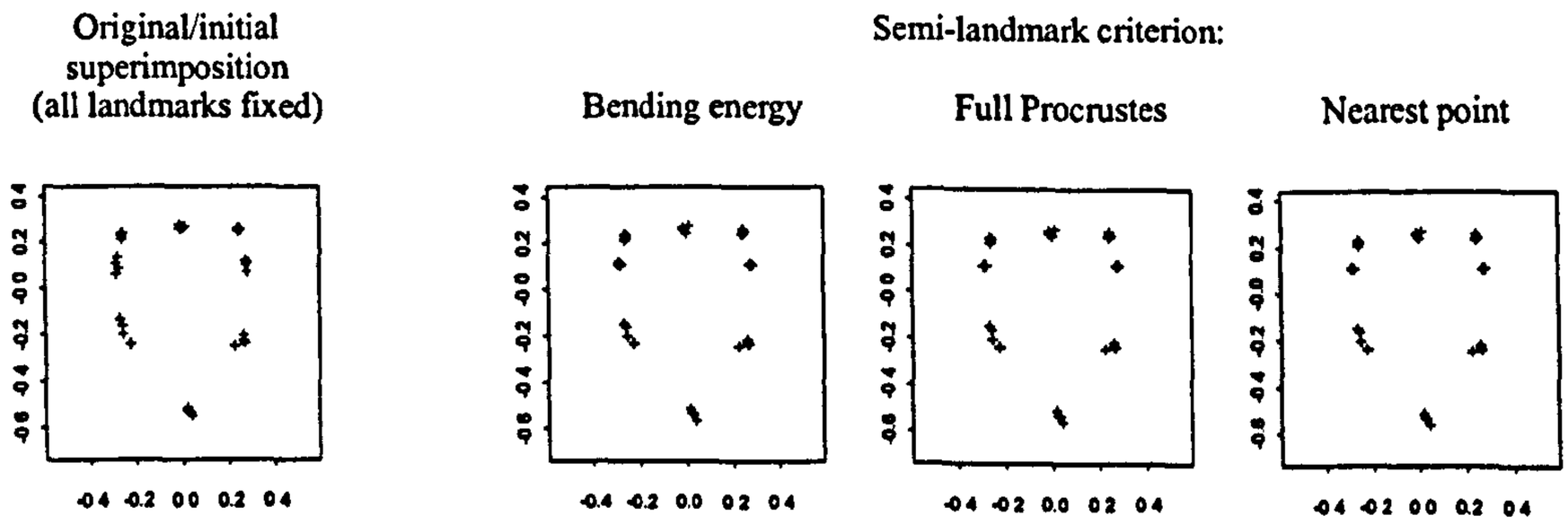


Fig 9.9: Procrustes fits of different operators representations of the lower central incisor with highest variation in component scores on first PC of the within case sums of squares and products matrix of the original Procrustes fits.

9.3.2 Occlusal surfaces

For the occlusal surfaces of the upper central incisor and upper first pre-molar our aim was to remove differences in shape due to operator inconsistencies in the orientation of the surfaces prior to imaging. (For the latter we also wish to remove operator differences in the position of the MD endpoints). Recall that here certain landmarks have been constrained to move by the same (or a proportion of the) distances moved by other semi-landmarks in the same direction, as described in 9.1.3.

9.3.2.1 Upper central incisor

In this application, the semi-landmarks along the incisal edge were constrained to move by the same distance in the BL direction, using equation (9.2) to determine their new positions. Table 9.5 details the iteration history of the change in values of RSS, for each of the semi-landmark methods. Again, the values in bold indicate the final RSS and ΔRSS when using a convergence criterion of $\Delta RSS < 0.001$.

Iteration	Bending energy		Full Procrustes		Nearest point	
	RSS	ΔRSS	RSS	RSS	ΔRSS	RSS
Original	0.676043331	0.000000000	0.676043331	0.000000000	0.676043331	0.000000000
1	0.314026993	0.362016338	0.297910933	0.378132398	0.438475319	0.237568012
2	**0.314018439**	**0.000008554**	**0.297886902**	**0.000024032**	0.350084250	0.088391069
3	0.314013049	0.000005390	0.297862689	0.000024212	0.317280749	0.032803500
4	0.314007655	0.000005394	0.297838293	0.000024396	0.305115740	0.012165009
5	0.314002259	0.000005396	0.297813712	0.000024581	0.300606539	0.004509201
6	0.313996862	0.000005397	0.297788944	0.000024768	0.298936988	0.001669552
7	0.313991463	0.000005399	0.297763988	0.000024956	**0.298320781**	**0.000616207**
8	0.313986062	0.000005401	0.297738842	0.000025146	0.298095318	0.000225463
9	0.313980660	0.000005402	0.297713505	0.000025337	0.298014792	0.000080526
10	0.313975256	0.000005404	0.297687976	0.000025529	0.297988015	0.000026777
11	0.313969851	0.000005406	0.297662253	0.000025723	0.297981162	0.000006854
12	0.313964444	0.000005407	0.297636335	0.000025918	0.297981686	-0.000000524
13	0.313959034	0.000005409	0.297610220	0.000026115	0.297984938	-0.000003252
14	0.313953624	0.000005410	0.297583907	0.000026313	0.297989195	-0.000004257
15	0.313948213	0.000005412	0.297557394	0.000026513	0.297993819	-0.000004624
16	0.313942798	0.000005414	0.297530679	0.000026714	0.297998576	-0.000004757
17	0.313937383	0.000005415	0.297503762	0.000026917	0.298003378	-0.000004802
18	0.313931966	0.000005417	0.297476640	0.000027122	0.298008193	-0.000004816
19	0.313926548	0.000005418	0.297449313	0.000027327	0.298013011	-0.000004818
20	0.313921128	0.000005420	0.297421778	0.000027535	0.298017828	-0.000004816

Table 9.5: Iteration history of RSS and ΔRSS values for different semi-landmark methods

Fig 9.11 plots the original and final sets of Procrustes registered configurations after using each of the semi-landmark methods. The scatters of the final Procrustes fits are almost identical for the different methods, with each showing a reduction in variance at

all landmark positions. In each set of final configurations, the position of the incisal edge along the BL dimension has been standardised, whilst retaining the shape of the configuration of points along the edge and angle between and relative dimensions of the BL and MD widths. The only visible difference between the methods is a slightly smaller/larger scatter of points at the buccal/lingual ends of the BL width for the BE method, due to slight differences in the final versions of the configurations being registered (see 9.3.3), although this is hardly noticeable.

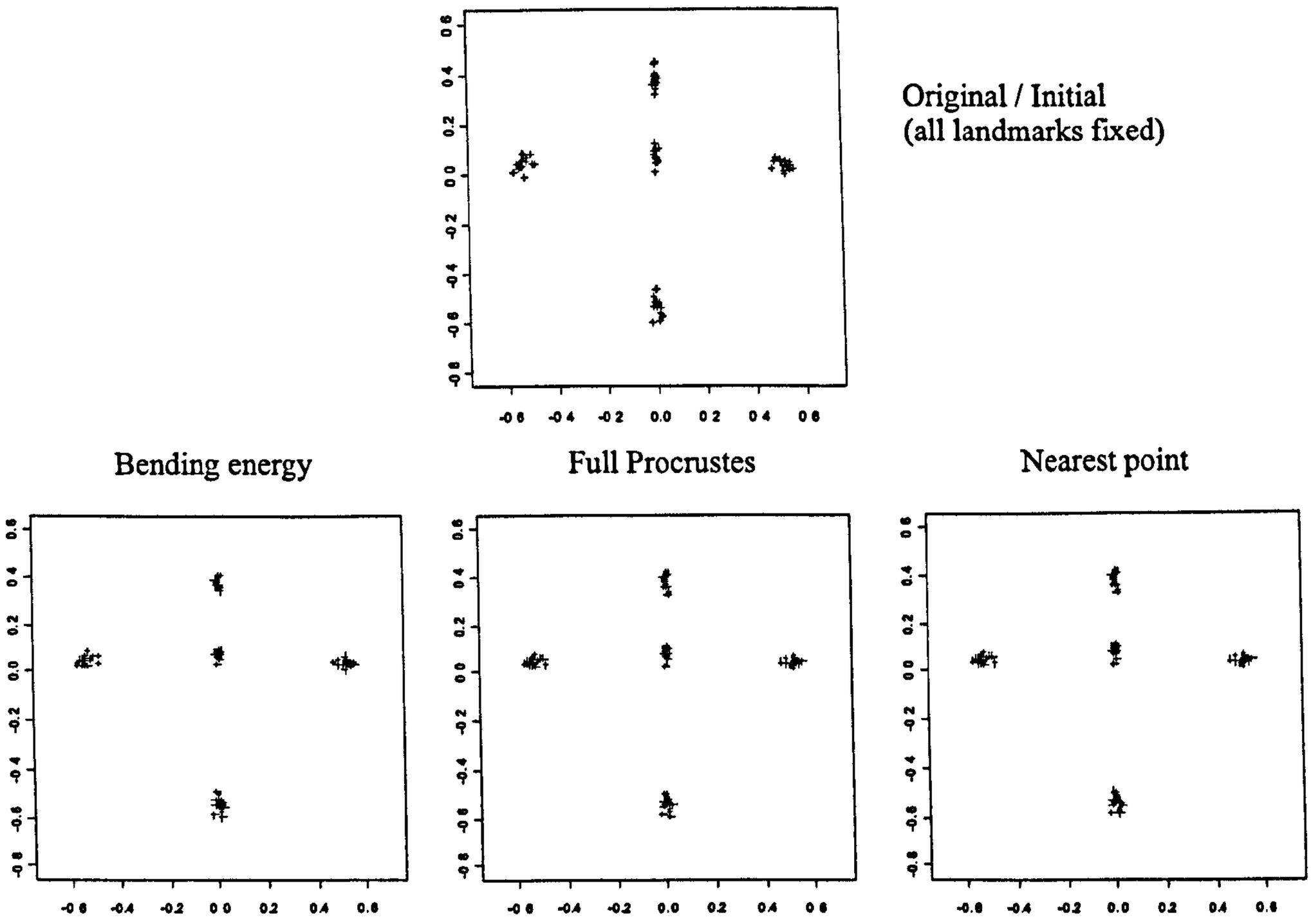
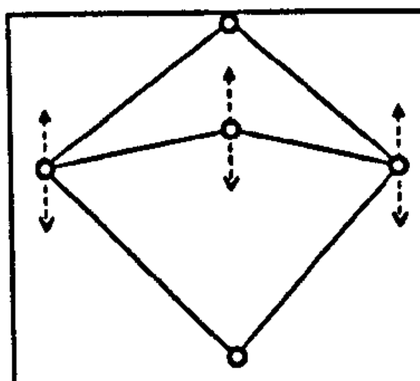


Fig 9.11: Procrustes registered configurations. (Top) Original/initial superimposition (all landmarks fixed). (Bottom) After running Semi-landmark procedure using different criteria.



Upper central incisor

	None (all fixed)	Bending energy	Full Procrustes	Nearest point
EucSS (cases)	0.273	0.196	0.186	0.186
EucSS (within)	0.403	0.118	0.112	0.113
EucSS (ops)	0.230	0.017	0.017	0.017
EucSS (errors)	0.173	0.101	0.096	0.096
EucSS (total)=*RSS*	0.676	0.314	0.298	0.298
Overall reliability	**0.297**	**0.522**	**0.520**	**0.520**
$d_F(\hat{\mu}_{SL}, \hat{\mu}_0)$		0.000	0.001	0.000

Table 9.6: Summary measures based on final Procrustes fits.

Table 9.6 details the summary measures obtained for each method from the final sets of 4×(n=19) configurations. Here there is essentially no difference between the measures and so interpretation is the same for each method. With over 50% of the variation in shape removed, each column of table 9.6 indicates that most of this was within cases (a reduction from 0.40 to 0.11 or 0.12) with the remaining variation between cases now accounting for over 50% of the variation in shape, rather than 30%. Note also from table 9.6 that the change in mean shape from the original/initial estimate is again negligible, with each $d_F(\hat{\mu}_{SL}, \hat{\mu}_0)$<0.001.

Fig. 9.12 shows the original and final Procrustes configurations obtained using each of the semi-landmark methods for each operator's representations of the three cases with the highest total variance in component scores on the first two PC's of the within-case SSP matrix, obtained from the original Procrustes fits in 4.5.3 (the three cases with the largest variance along the directions shown in the first row of fig. 4.7). Clearly the within-case variation in Procrustes fits has been reduced for each of the cases at all landmarks (not just the semi-landmarks), using each of the methods.

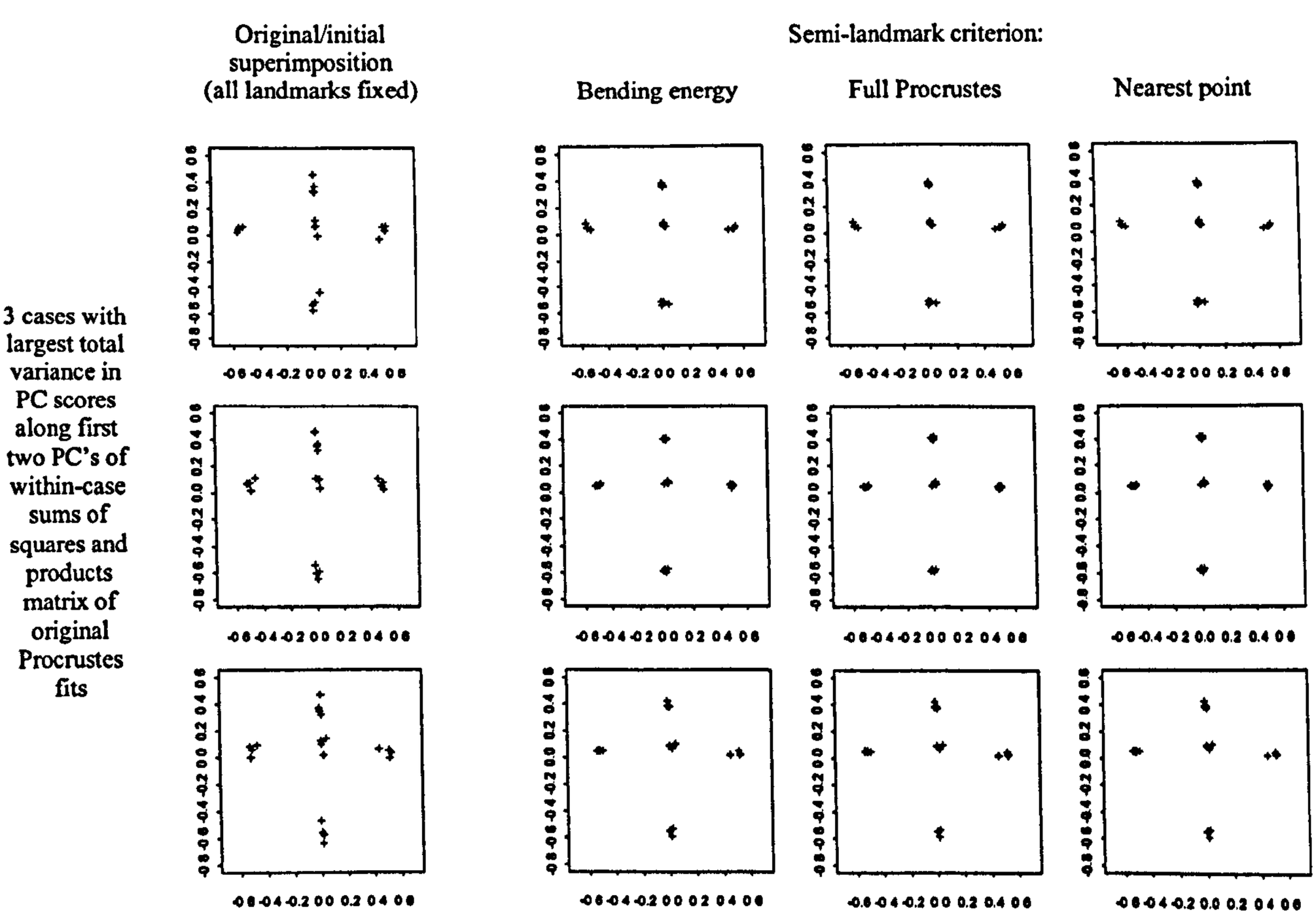


Fig 9.12: Procrustes fits of different operators representations of the occlusal surfaces of three upper central incisors, obtained using ordinary GPA registration and each of the semi-landmark methods.

Note that better results (in terms of the reliability figures) can be obtained if the semi-landmarks are unconstrained, since this allows greater freedom for where they are able to move to when reducing the variation along each direction. However, one should not be swayed by this. It still makes more sense to ensure that the shape of the landmarks along the incisal edge is always retained since this is what happens in reality.

9.3.2.2 Upper first pre-molar

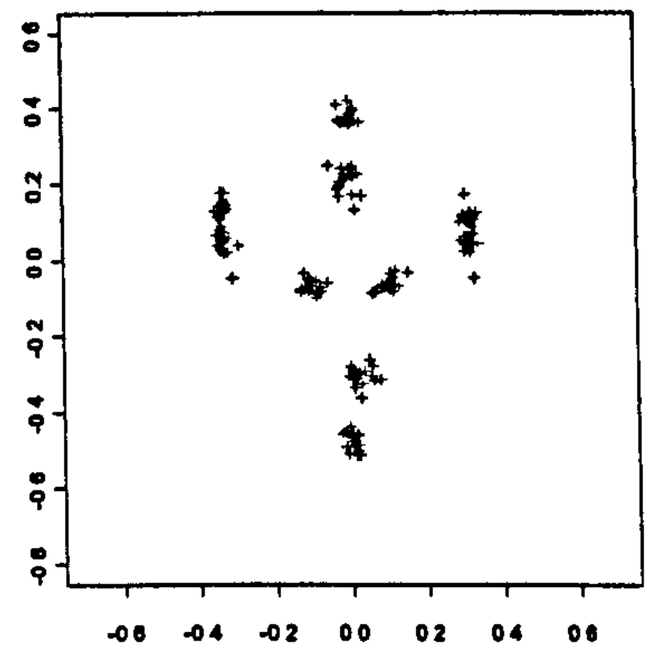
In this application, a series of weighted proportions were used to constrain the movement of some of the semi-landmarks in the BL direction, with the MD landmarks free to move independently along this direction.

Table 9.7 shows the iteration history for each method, with bold/shaded values denoting convergence at ΔRSS<0.001 and when each process begins to 'slow'. For the BE method, the process was stopped on the second iteration when a negative value of ΔRSS occurred (as in the gingival simulation study).

Iteration	Bending energy		Full Procrustes		Nearest point	
	RSS	ΔRSS	*RSS*	*RSS*	ΔRSS	*RSS*
Original	0.91375520		0.913755202	0.000000000	0.91375520	
1	0.67675521	0.23699999	0.575076847	0.338678355	0.64525176	0.26850344
2	**0.67696355**	**-0.00020834**	**0.574870961**	**0.000205886**	0.61773735	0.02751441
3	0.67696308	0.00000047	0.574662031	0.000208930	0.60135615	0.01638120
4	0.67696180	0.00000128	0.574450007	0.000212024	0.59130788	0.01004826
5	0.67696048	0.00000132	0.574234846	0.000215161	0.58512704	0.00618085
6	0.67695906	0.00000142	0.574016504	0.000218342	0.58132186	0.00380518
7	0.67695770	0.00000136	0.573794936	0.000221568	0.57897738	0.00234448
8	0.67695620	0.00000150	0.573570098	0.000224838	0.57753149	0.00144589
9	0.67695470	0.00000150	0.573341944	0.000228154	**0.57663865**	**0.00089284**
10	0.67695311	0.00000159	0.573110428	0.000231516	0.57608630	0.00055235
11	0.67695150	0.00000161	0.572875504	0.000234924	0.57574365	0.00034265
12	0.67694985	0.00000165	0.572637124	0.000238380	0.57553017	0.00021348
13	0.67694821	0.00000164	0.572395241	0.000241883	0.57539627	0.00013390
14	0.67694643	0.00000178	0.572149807	0.000245434	0.57531140	0.00008486
15	0.67694457	0.00000186	0.571900772	0.000249035	0.57525675	0.00005465
16	0.67694270	0.00000186	0.571648088	0.000252684	0.57522072	0.00003604
17	0.67694080	0.00000191	0.571391704	0.000256384	0.57519614	0.00002457
18	0.67693883	0.00000197	0.571131570	0.000260134	0.57517863	0.00001751
19	0.67693679	0.00000204	0.570867635	0.000263935	0.57516546	0.00001317
20	0.67693476	0.00000202	0.570599847	0.000267788	0.57515496	0.00001050

Table 9.7: Iteration history of *RSS* and ΔRSS values for different semi-landmark methods.

Fig. 9.14 plots the original and final sets of Procrustes registered configurations after using each of the semi-landmark methods and table 9.7 details the summary measures based on each set of 4×(n=19) final Procrustes fits.



Original / Initial
(all landmarks fixed)

Bending energy

Full Procrustes

Nearest point

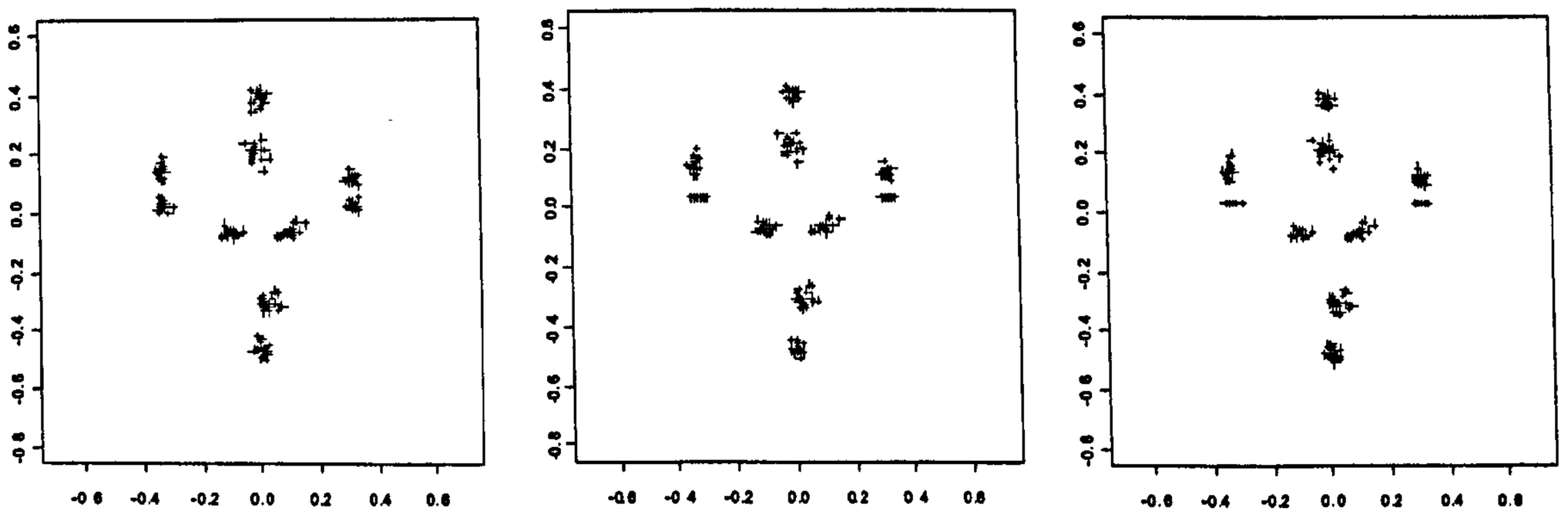
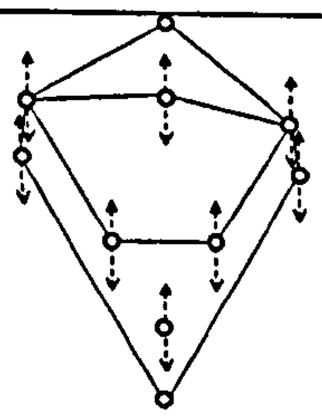


Fig 9.14: Procrustes registered configurations. (Top) Original/initial superimposition (all landmarks fixed). (Bottom) After running Semi-landmark procedure using different criteria.



Upper first pre-molar

	None (all fixed)	Bending energy	Full Procrustes	Nearest point
EucSS (cases)	0.289	0.279	0.234	0.234
EucSS (within)	0.624	0.398	0.341	0.342
EucSS (ops)	0.262	0.103	0.093	0.093
EucSS (errors)	0.362	0.295	0.248	0.249
EucSS (total)=*RSS*	0.914	0.677	0.575	0.577
Overall reliability	**0.176**	**0.265**	**0.260**	**0.260**
$d_F(\hat{\mu}_{SL}, \hat{\mu}_0)$		0.001	0.002	0.001

Table 9.7: Summary measures based on final Procrustes fits.

Compared to the original/initial superimposition of the data, use of the FP or NP method removes all variation at the MD landmarks in the BL direction and reduces the variance all other landmarks in this direction too. However, in table 9.7 the results are all the same, with no significant improvement in reliability offered by the use of either method. Of the within-case variation examined in 4.5.3, only 19% of this was found to be along directions consistent with orientation effects which we have attempted to represent here and so the lack of a real improvement to the figures is unsurprising.

9.3.3 Further examination of results

In this section we consider the results of the applications above in more detail and examine some of the similarities and differences between the final sets of configurations produced by the different methods. We explain why, in contrast to the gingival simulation study, in these applications the BE method produces more comparable results to those obtained by the FP or NP methods without the problem of unrealistic shapes being produced.

9.3.3.1 Changes in variance at fixed landmarks

As well as a reduction in variance at the positions of the semi-landmarks, in each of the applications it was found that after running each method changes also occurred in the patterns of variance at the positions of the other, fixed landmarks. This was most noticeable for the upper central incisor, where with each method a decrease in vertical variation at the gingival endpoint of the LACC was observed along with an increase in variation at the incisal endpoint and slight increase in the horizontal variation at all fixed landmarks (see fig. 9.6).

By removing the variance at the semi-landmarks, this unwanted variation is no longer distributed about the rest of the configuration by GPA, allowing the actual variance at these positions to be represented more effectively. Sections 7.4 and 8.3.3.3 described how, when variation at the semi-landmark positions is reduced and the configurations re-registered to an (essentially) unchanged mean, then up until convergence at a suitably chosen level, GPA distributes the reduction in variance at the semi-landmarks around the rest of the landmarks, resulting in an overall reduction in the contribution to *RSS* for each case.

However, how the variance at each individual landmark changes from how it was previously affected, will depend on the choice of semi-landmarks, chord directions and distribution of the landmarks in the configurations in question. (In the gingival simulation study, a reduction in variance was noted at all fixed landmarks, after using both the FP and NP methods). Here for example, in fig 9.6 (top), the initial GPA registration of each configuration to the initial estimated mean $\hat{\mu}_0$, means that the 5

landmarks below (0,0) are usually matched more closely to those in $\hat{\mu}_0$ than the 3 gingival landmarks above (0,0). As each configuration is centred (to (0,0)), rotated and scaled (in all directions from (0,0)), optimisation of the OLS distances between corresponding landmarks is usually achieved by concentrating the registration at the 5 non-gingival landmarks. Consequently the variation in fig 9.6 (top) is initially greater at the three gingival landmarks than the others, particularly in the occlusal-gingival (vertical) direction, since the non-gingival landmarks do not vary as much in this direction. When each new semi-landmark method successfully reduces the variation at the four semi-landmark positions, at the MD positions and either side of the incisal edge, the GPA registration at the end of each iteration is able to register each new version of a configuration to an (almost unchanged) mean with a reduction in residual sums of squares between all corresponding landmarks, as described above. Here however, having removed some of the variation in these semi-landmarks, the gingival landmarks can now be matched more closely than before (especially in the *y*-direction), but this is at the expense of an increase in variation at the (fixed) incisal landmark at the bottom of the LACC and a slight increase in variation in the *x*-direction of all landmarks of the Procrustes fits.

9.3.3.2 Similarity of results under different criteria

In contrast to the gingival simulation study, in each of the applications of this chapter, the BE method produces more similar results to those obtained using the NP or FP methods. In this section we look at the reasons why.

Using the BE method, recall how the new positions of the semi-landmarks in $Y^{new} = X^{new} = vec^{-1}(vec(X^P) - U\lambda)$ are determined as those which produce the PTPS mapping from mean $T = \hat{\mu}$ which gives the closest possible 'affine fit', $Q\Gamma^{21}vec^{-1}(vec(X^P) - U\lambda)$, to the fixed landmarks and positions along the semi-landmark's chords. (Matrices Q, and Γ^{21} are as defined in equations (5.31) and (5.32)). In other words, the semi-landmarks move so as to minimise the sum of squared corresponding non-affine components of the mapping, added to the 'affine fit' to ensure that $\hat{\mu}$ is mapped to $vec^{-1}(vec(X^P) - U\lambda)$ exactly. They can also be thought of as the positions which produce the smallest possible variation in the shape of the grid cells of a

deformation grid, representing the PTPS transformation from $T=\hat{\mu}$ to $vec^{-1}(vec(X^P)-U\lambda)$. As explained in 5.6.3, large movements of the semi-landmarks will occur when a notable change in the affine mapping of $\hat{\mu}$ (from that in the original transformation to X^P) is able to occur in the directions of the semi-landmarks' chords, determined by the configuration of landmarks effectively fixed in these directions. This requires that there are no nearby fixed landmarks, or other semi-landmarks whose movement is restricted to more perpendicular directions, to stop them moving in accordance with such a mapping.

In each of the applications here, there are fixed landmarks in the vicinity of the semi-landmarks, on either side of their chord directions, which must be matched to exactly by the PTPS transformation from $\hat{\mu}$ to $vec^{-1}(vec(X^P)-U\lambda)$. This means that the affine component of the mapping and hence shape of the configuration cannot change too much, whereas in the gingival simulation study there were no fixed landmarks in such positions or differences in chord directions to constrain the semi-landmarks' movement.

For example, with the lower central incisors (section 9.3.1.2) the semi-landmarks (the MD endpoints) will always remain between the incisal corners and papillae landmarks as they move to positions in accordance with the vertical affine component of the mapping between the fixed landmarks, as illustrated in fig 9.10. The top row displays the PTPS mapping from $\hat{\mu}$ to initial Procrustes fit X^P of one of the lower central incisor configurations (buccal surface), as a deformation grid, indicating how a square grid placed over $\hat{\mu}$ is deformed so that the landmarks are mapped exactly to those of X^P. The right side panel shows the affine mapping of $\hat{\mu}$ to X^P as a deformation grid, as described in 5.3.3, with diagonal crosses indicating the fitted points $Q\Gamma^{21}X^P$ (to which residual, non affine components must be added, producing the variation in the shape of the grid cells in the middle panel, so that $\hat{\mu}$ is mapped to X^P exactly). The bottom row displays the PTPS mapping from $\hat{\mu}$ to X^{new}, after the semi-landmarks have been allowed to move along their chords. The right side panel shows the affine fit of $\hat{\mu}$ to X^{new}, here given by $Q\Gamma^{21}X^{new}$, which differs from that from $\hat{\mu}$ to X^P, in the panel above by just a slight increase in the scale of the vertical component where the transformation is now able to achieve an improved affine fit to the landmarks and

chords of $vec^{-1}(vec(X^P) - U\lambda)$ (see landmarks around the gum and along the incisal edge).

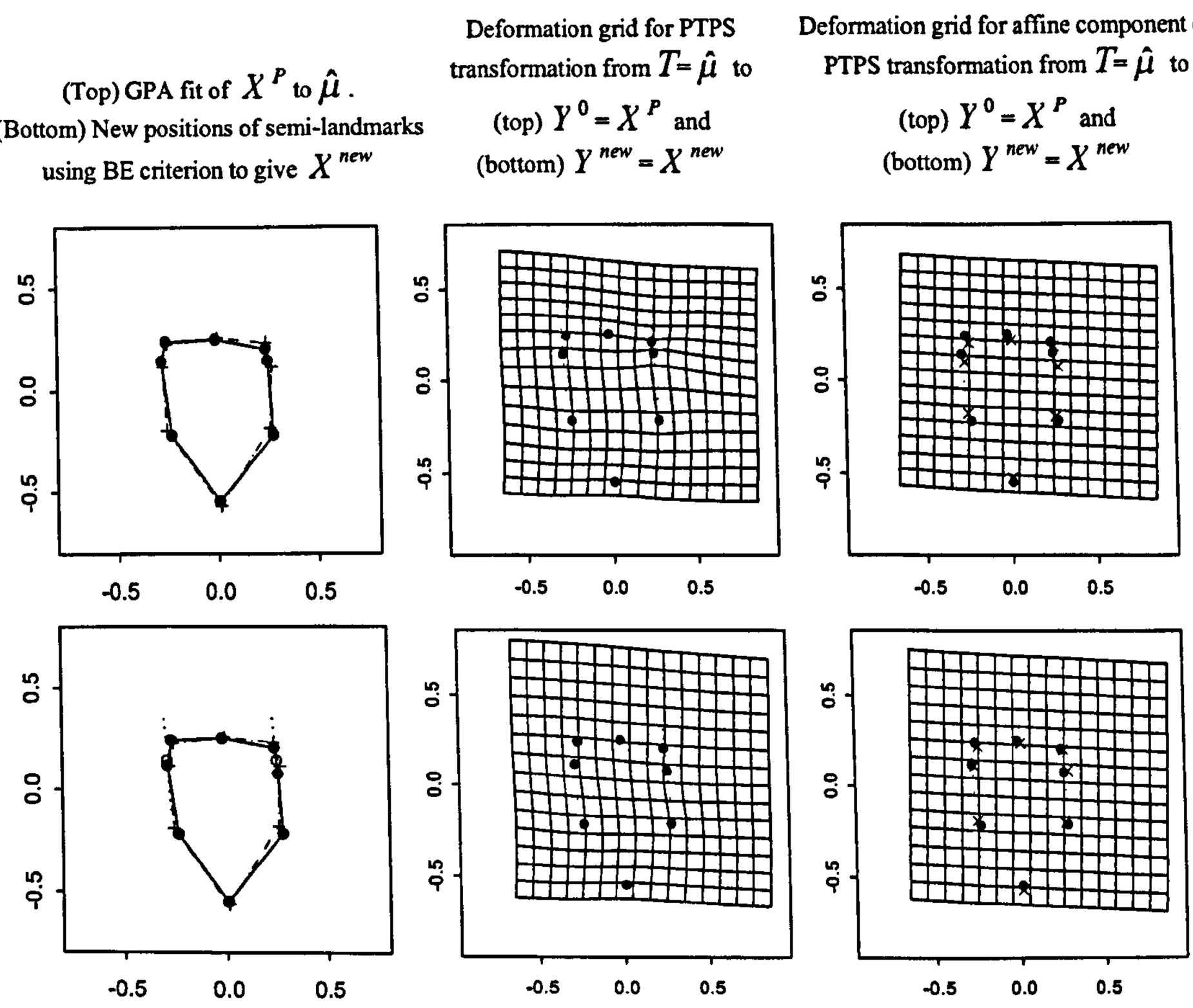


Fig 9.10: (Top left) Procrustes fit X^P (filled circles, solid lines) to $\hat{\mu}$ (plusses, dashed lines). (Bottom left) New configuration X^{new} (filled circles, solid lines) after semi-landmarks move from nominal positions (empty circles) along escribed chords according to minimum bending energy criterion. (Middle column) as described at top. (Right column) Affine mapping of $\hat{\mu}$ to X^P (top) and X^{new} (bottom) (filled circles), with fitted values (diagonal crosses) $Q\Gamma^{21}X^P$ (top) and $Q\Gamma^{21}X^{new}$ (bottom).

Similarly, on the upper central incisor (section 9.3.1.1), the fixed landmarks at the ends of the LACC ensure that this dimension is retained relative to the fixed landmarks at the papilla locations in the transformation from $\hat{\mu}$ to $vec^{-1}(vec(X^P) - U\lambda)$. In addition there is typically an angle of around 90 to 135 degrees between each of the two semi-landmarks' chord directions on each side of the tooth (see fig. 9.1, left), which the landmarks in $\hat{\mu}$ must be mapped onto exactly. This also restricts the possibility of large or unrealistic movements occurring along these directions.

With both the FP and NP methods however, the semi-landmarks are unable to move large distances since, when using the FP criterion, it is required that the fit of the entire

configuration to $\hat{\mu}$ is optimised and with the NP method, prior GPA registration ensures that the semi-landmarks need not move too far in order to find positions along their chords at nearest Euclidean distance to corresponding landmarks in the mean.

Even though the presence of nearby fixed landmarks and/or differences in chord directions restricts the possibility of large unrealistic movements occurring when using the BE method, differences will still remain between the results of the individual steps and in the final semi-landmark positions, compared to the FP and NP techniques. For example, fig 9.11 illustrates how on the first iteration, the BE criterion does not quite result in the same new semi-landmark positions on the first movement step as the FP method. Here we describe conditions affecting the degree of dissimilarity in the results.

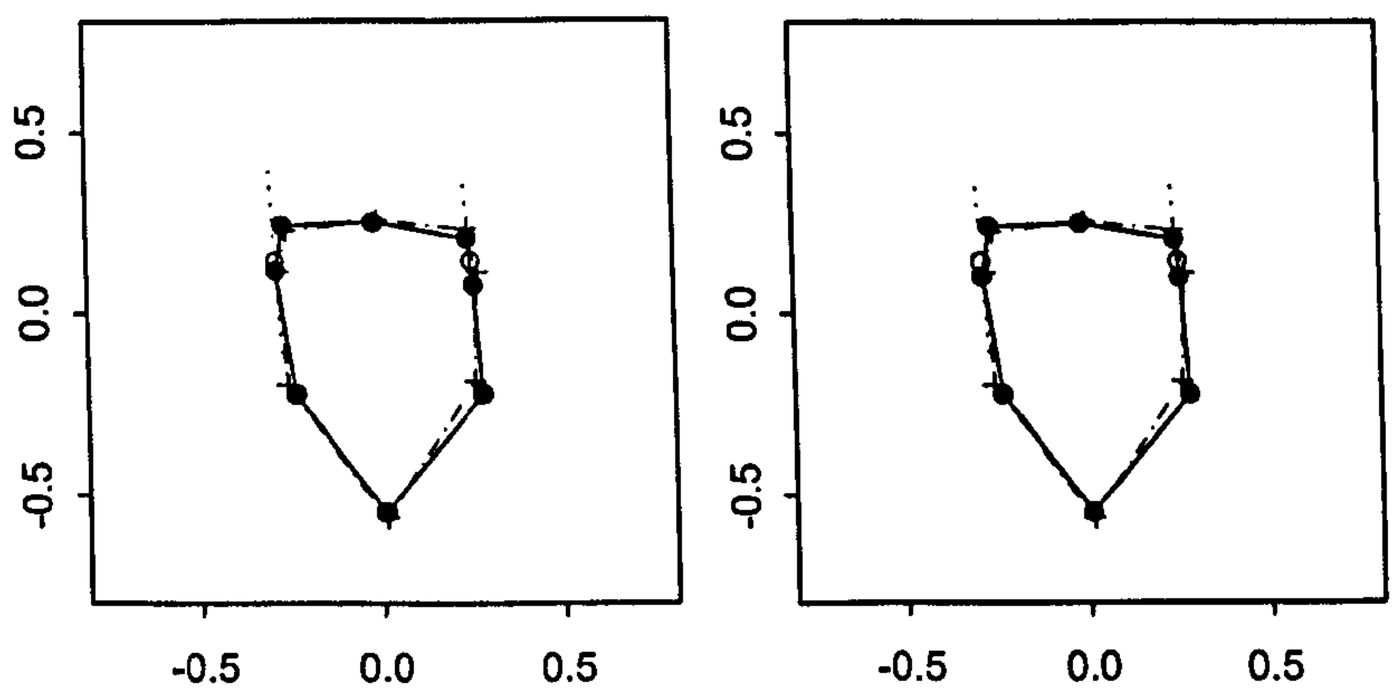


Fig 9.11: New semi-landmark positions for a lower central incisor (buccal surface) configuration using (left) bending energy and (right) full Procrustes criteria. (Filled circles, solid lines) New version of configuration. (Empty circles) Nominal semi-landmark positions. (Dotted lines) chord directions. (Plusses, dashed lines) Means shape.

In each of the applications so far in this chapter, the chord directions do not change between iterations as these are defined by the fixed landmarks. As there is also very little change in $\hat{\mu}$ between iterations the optimal positions of the semi-landmarks determined on the first movement step using the BE or FP criteria hardly change. Therefore differences in the results of any single step for these methods, are likely to be the similar to those observed in the final sets of Procrustes fits from each method. Over the course of the iteration, the final semi-landmark positions of the NP technique are usually identical to those obtained using the Full Procrustes criterion, as previously noted in 7.3 and 8.3.3.1, so here we concentrate on comparing the BE and FP methods.

Recall from equations (5.61) and (5.62) that with the BE method, the new semi-landmark positions are given by:

$$\min_{B,\lambda} tr(vec^{-1}(vec(X^P)-U\lambda)-(1_k:\hat{\mu})B)^T S^{-1}(vec^{-1}(vec(X^P)-U\lambda)-(1_k:\hat{\mu})B) \quad (9.4)$$

where B is a $k\times 2$ matrix of parameters and S is as defined in (5.31). For the FP criterion, the new semi-landmark positions at each movement stage are instead given by finding λ minimising:

$$\min_{SG,\lambda} \left\| SG(vec^{-1}(vec(X^P)-U\lambda))-\hat{\mu}\right\|^2 \text{ or } \min_{SG,\lambda} \left\| vec^{-1}(vec(X^P)-U\lambda)-SG(\hat{\mu})\right\|^2$$

$$=\min_{SG,\lambda} tr(vec^{-1}(vec(X^P)-U\lambda)-SG(\hat{\mu}))^T(vec^{-1}(vec(X^P)-U\lambda)-SG(\hat{\mu})). \quad (9.5)$$

With the NP method, there is no transformation of $\hat{\mu}$ involved; $vec^{-1}(vec(X^P)-U\lambda)$ is matched to $\hat{\mu}$ (or vice-versa) by finding λ minimising:

$$\min_{\lambda} tr(vec^{-1}(vec(X^P)-U\lambda)-\hat{\mu})^T(vec^{-1}(vec(X^P)-U\lambda)-\hat{\mu})$$

which, in contrast to (9.4) and (9.5), depends on the nominal positions of X^P and $\hat{\mu}$.

By comparison of (9.4) and (9.5) above, the BE criterion in (9.4) allows separate scaling in either the x or y direction and shearing, in addition to similarity transformations when matching $\hat{\mu}$ to $vec^{-1}(vec(X^P)-U\lambda)$, whereas in (9.5) the transformation of $\hat{\mu}$ is restricted to rigid transformations of location, scale and rotation only. In addition (9.4) is a minimisation of a sum of generalised squared residuals, i.e. the correlated non-affine components, with weighting matrix S, whereas (9.5) is a minimisation of an ordinary sum of squared residuals with no weighting matrix.

Based on the first point above, the new positions of the semi-landmarks will always be similar for the BE and FP methods, when the affine component resembles a similarity transformation, as was seen in fig 9.10 (bottom right). However, even when this is the case, differences in the new semi-landmark positions will still exist, due to the fact that the non-affine components, completing the PTPS mapping onto the fixed landmarks and

chords of $vec^{-1}(vec(X^P)-U\lambda)$ are correlated, weighted displacements from the affine fit, and not OLS residuals. i.e. due to the difference in how the 'residuals' of the Procrustes and affine fits of $\hat{\mu}$ to $vec^{-1}(vec(X^P)-U\lambda)$ are defined. For example, fig 9.10 (bottom left) shows how with a similarity-like affine transformation of $\hat{\mu}$, the semi-landmarks move to positions on their chords which give the smallest set of non-affine components between the fitted affine points, $Q\Gamma^{21}vec^{-1}(vec(X^P)-U\lambda)$ and $vec^{-1}(vec(X^P)-U\lambda)$. Despite their appearance however, these chord positions are not those at least Euclidean distance to those in $Q\Gamma^{21}vec^{-1}(vec(X^P)-U\lambda)$. On the other hand, with the FP method, the semi-landmarks would move precisely to positions at least Euclidean distances to the rotated, scaled and translated version of $\hat{\mu}$.

This also explains why, in each of the examples considered here and in the simulation study of chapter 8, when using the FP and NP methods with unconstrained semi-landmarks, any remaining variation in shape at the final positions of the semi-landmarks is entirely in directions perpendicular to the final chords directions. When the BE method is successful in removing variation along the specified directions, what is left appears more isotropic than with the new methods, with some variation still remaining along the final chord directions, rather than only perpendicular to them.

9.4 Results: Inclusion of gum effects

9.4.1 Buccal surface of upper central incisor

In addition to the semi-landmarks at the MD endpoints and incisal corners, here we re-run the analysis of 9.3.1.1 to also include the three gingival landmarks as semi-landmarks, with chord directions as defined in fig. 9.4. In contrast to the gingival simulation study, here there is very little or no variation in the position of the gingival landmarks 'within-cases' since the data were obtained from stone casts and their positions were found to be consistent between different operators' representations. However we still want remove any variation 'between cases' due to differences in the position of the patient's gum, as this will be of no interest in any subsequent analysis. Additionally, in chapter 8, the MD endpoints and incisal corners were simulated at the

same position on the same case (and so did not vary within cases, only between). They were also treated as fixed landmarks during the simulation study as our only interest was the performance of each method in removing variation in the positions of the gingival landmarks. In practice however, variation exists in these positions due to operator inconsistencies in their placement around the lower outline.

9.4.1.1 Main findings using reliability data

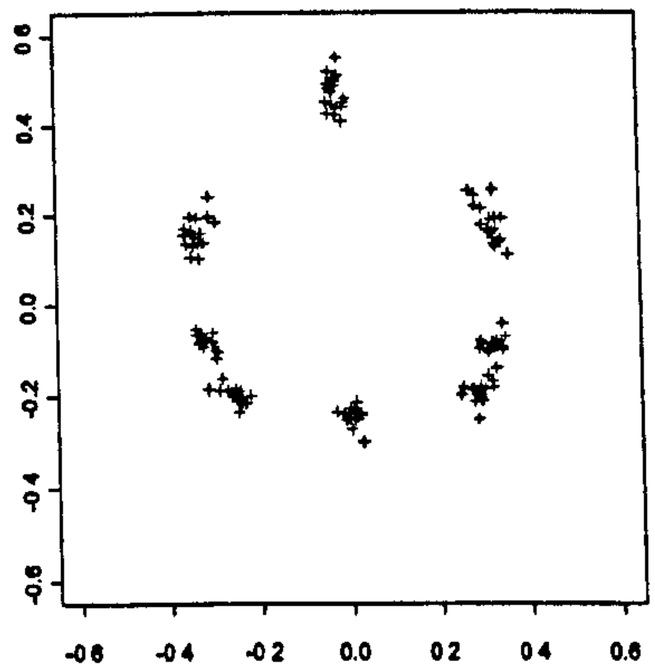
Table 9.8 details the iteration history of the change in values of RSS, for each of the semi-landmark methods. For the FP and NP methods, similar final values of RSS were obtained using the usual criterion of $\Delta RSS < 0.001$. For the BE method however, the process soon produces unwanted results. A large increase in RSS occurs on the first (and second) iterations, when most movement in the semi-landmarks occur, the reasons for which, we discuss in detail in 9.4.3. Following this RSS does decrease but it was not until the 28th iteration that the change in RSS was less than 0.01. By the 60th iteration, ΔRSS was still decreasing but had still not reached the 0.001 level.

Iteration	Bending energy		Full Procrustes		Nearest point	
	RSS	*ΔRSS*	*RSS*	*RSS*	*ΔRSS*	*RSS*
Original	0.923110		0.923110		0.923110	
1	1.943434	-1.020325	0.069321	0.853789	0.158607	0.764503
2	2.400833	-0.457399	**0.068593**	**0.000728**	0.131140	0.027466
3	2.001942	0.398891	0.068363	0.000230	0.114014	0.017126
4	1.664637	0.337304	0.068136	0.000227	0.101918	0.012096
5	1.414340	0.250297	0.067912	0.000224	0.093181	0.008737
6	1.235849	0.178492	0.067689	0.000222	0.086769	0.006413
7	1.094980	0.140868	0.067469	0.000220	0.082001	0.004768
8	0.950557	0.144424	0.067251	0.000218	0.078417	0.003584
9	0.843384	0.107172	0.067035	0.000216	0.075697	0.002720
10	0.742467	0.100917	0.066821	0.000214	0.073615	0.002083
11	0.677628	0.064839	0.066609	0.000212	0.072005	0.001610
12	0.611467	0.066161	0.066399	0.000210	0.070749	0.001256
13	0.550927	0.060540	0.066190	0.000208	**0.069758**	**0.000991**
14	0.507788	0.043139	0.065984	0.000206	0.068965	0.000792
15	0.459007	0.048780	0.065780	0.000204	0.068324	0.000642
16	0.418182	0.040825	0.065578	0.000202	0.067796	0.000528
17	0.380096	0.038086	0.065377	0.000200	0.067355	0.000441
18	0.349298	0.030798	0.065179	0.000199	0.066979	0.000376
19	0.325050	0.024248	0.064982	0.000197	0.066653	0.000326
20	0.299431	0.025619	0.064787	0.000195	0.066365	0.000288

Table 9.8: Iteration history of RSS and ΔRSS values for different semi-landmark methods.

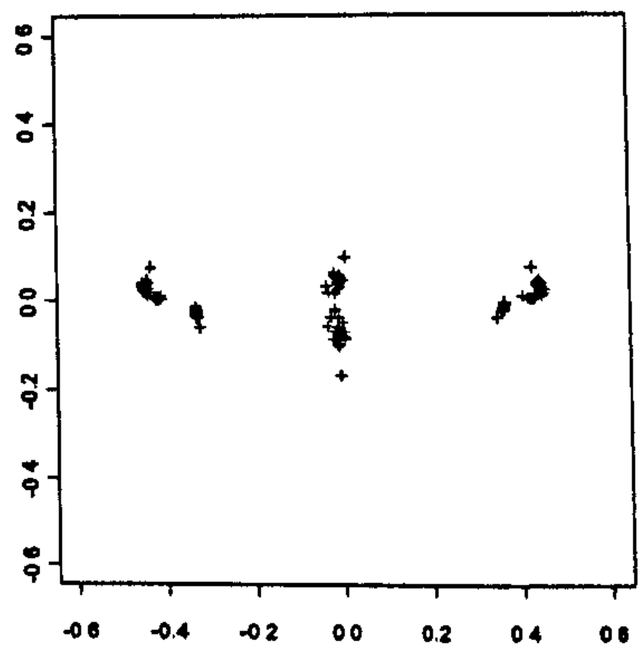
Fig 9.12 plots the original (with all landmarks fixed) and final Procrustes registered configurations after using each of the semi-landmark methods and table 9.9 details the summary measures obtained from the final sets of $4\times(n=19)$ configurations. For

comparison, the results of the BE method, based on running the process to the 28th iteration (ΔRSS <0.01) have been included (although it is questionable as to whether the process should have been run beyond the first iteration).

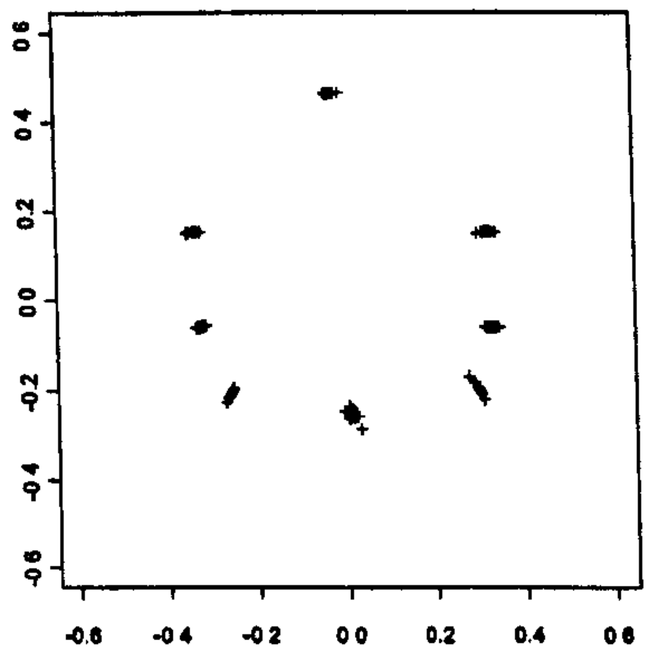


Original / Initial
(all landmarks fixed)

Bending energy



Full Procrustes



Nearest point

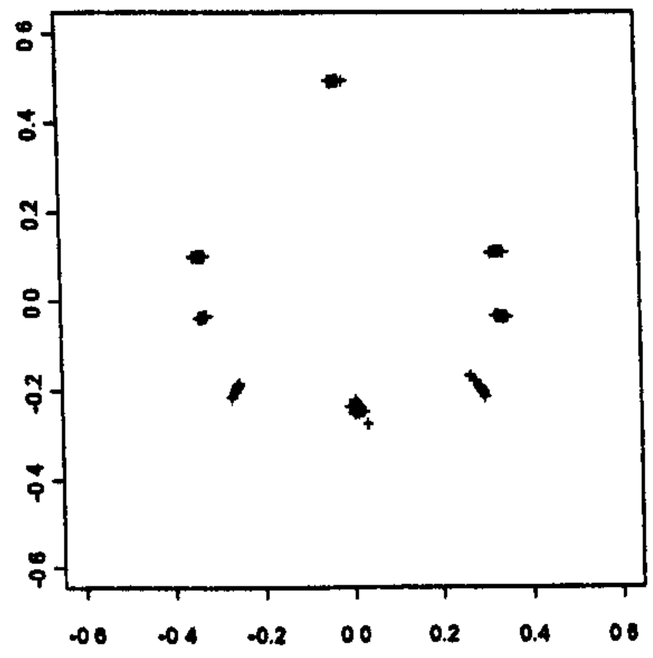
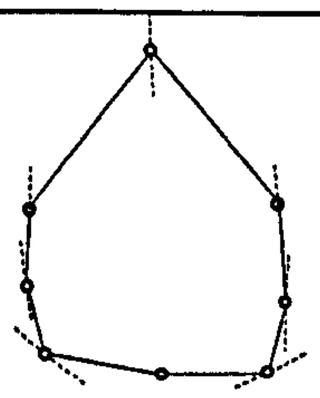


Fig 9.12: Procrustes registered configurations. (Top) Original/initial superimposition (all landmarks fixed). (Bottom) After running semi-landmark procedure using different criteria.



Upper central incisor

	None (all fixed)	Bending energy*	Full Procrustes	Nearest point
EucSS (cases)	0.564	0.113	0.047	0.049
EucSS (within)	0.359	0.077	0.022	0.021
EucSS (ops)	0.150	0.041	0.004	0.004
EucSS (errors)	0.209	0.037	0.017	0.017
EucSS (total)=*RSS*	0.923	0.190	0.069	0.070
Overall reliability	**0.522**	**0.507**	**0.600**	**0.617**
$d_F(\hat{\mu}_{SL}, \hat{\mu}_0)$		0.581	0.006	0.014

Table 9.9: Summary measures based on final Procrustes fits. * ΔRSS <0.01 used for convergence.

As usual the FP and NP methods achieve almost identical results. Fig 9.12 shows that both methods are successful in reducing the variance of the chosen landmarks along the specified directions, with all remaining variation at the semi-landmarks perpendicular to their final chord directions. Only the incisal end of the LACC was a fixed landmark

here and the variation in this position is reduced as a result of the variation at all other landmarks being reduced. For the BE method however, it is clearly evident that something has gone wrong, with all configurations collapsing to unrealistic but similar shapes. The large change in mean shape in table 9.9, is consistent with the observation that the configurations have all notably changed in some systematic way. For the FP and NP methods the change in mean is again negligible (although slightly larger for the nearest point method here).

Note that since we are allowing the gingival landmarks to move, in order to reduce variation between cases, the reliability measures in table 9.9 may not be particularly informative here (as in the gum simulation) since this will naturally decrease reliability values. Despite this however, an increase in the values was found with use of both the FP and NP methods, when compared to the original GPA registration of the data. With both methods the total variation in shape (*RSS*) has been reduced from 0.92 to 0.07 and that 'between cases' from 0.56 to 0.05. This leaves less than 5% of the original variation between cases on which to base any subsequent analysis. (With just the MD endpoints and incisal corners as semi-landmarks in section 9.4.1.1, over 70% of the original variation between cases remained). It is clear from these figures that a large proportion of what is originally observed 'between cases' and regarded as actual variation in shape is due to differences in the positions of the patients gingival profiles. However, as the example in 9.4.1.2 below illustrates, important differences in shape between populations are still able to be detected.

9.4.1.2 Analysis example: Comparison of two populations

Two samples of n=20 landmark configurations were obtained from the same images of buccal surfaces of upper left central incisors, used as the example data in 3.4.2. The first set of configurations were obtained from images of study casts of 20 'control' patients. Here, all eight landmarks were identified on each image, as defined in fig. 1.3 (left). The second sample comprised 20 corresponding configurations from patients with moderate/severe hypodontia (three or more congenitally missing teeth). In 3.4.2, all corresponding landmarks except for the mesio-distal (MD) endpoints were identified on each image, as these are generally difficult to locate precisely along the sides of the teeth in this population. Here however, approximate positions were identified

corresponding to the operator's best guess of where landmarks corresponding to the MD endpoints in the normal sample should be. It is hoped that use of the new semi-landmark methods, with semi-landmarks and chord directions as defined in fig. 9.4 and used in the application above, will remove any differences due to the location of these positions along the sides of the teeth, while allowing information on the relative dimensions of the sides of the hypodontia teeth to be incorporated into the registration procedure and subsequent comparison of mean shapes In addition, we also wish that differences in the locations of the three gingival landmarks due to differences in the position of patients' gingival profiles be accounted for and removed as well as any unwanted variation which may result from the poor operator reliability in locating the landmarks at the MD endpoints and corners of the incisal edge.

Fig 9.13 (top) shows the Procrustes fits of both samples to an overall estimated pooled mean shape, before and after using the semi-landmark method with the FP criterion determining the new positions of the semi-landmarks at each step and a convergence criterion of $\Delta RSS < 0.001$. Almost identical results were found using the NP method. Fig 9.13 (bottom) shows the (arithmetic) mean shape obtained from the control configurations Procrustes matched to the (arithmetic) mean shape of the hypodontia sample, both before and after using the FP method. In the bottom left panel, differences in the mean locations of the MD endpoints in each sample, result in the appearance of a large difference in mean shape. In the hypodontia population, there is a tendency to locate the endpoints of the maximal buccal more gingivally (towards the top) of the contact region, rather than at positions corresponding to where the teeth would contact with neighbouring teeth, if they had normal shape. As the right panels show, the semi-landmark procedure accounts for such differences, removing this variation (along with that due to differences in the position of patients' gingival tissue and along directions in which errors in the locations of landmarks at the corners of the incisal edge may occur) allowing a more sensible comparison of mean shape to be made. Although the difference in mean shape is smaller (as measured by the Euclidean sums of squares between the configuration in the bottom right panel), use of Hotelling's two-sample T^2 test, as described in (3.29), gives a p-value of 0.002, providing strong evidence of a difference between the hypodontia and control mean incisor shapes. Fig. 9.13 (bottom right) suggests that the mean shapes differ in the shape of the sides of the teeth, with the hypodontia cases being more 'tapered' in shape.

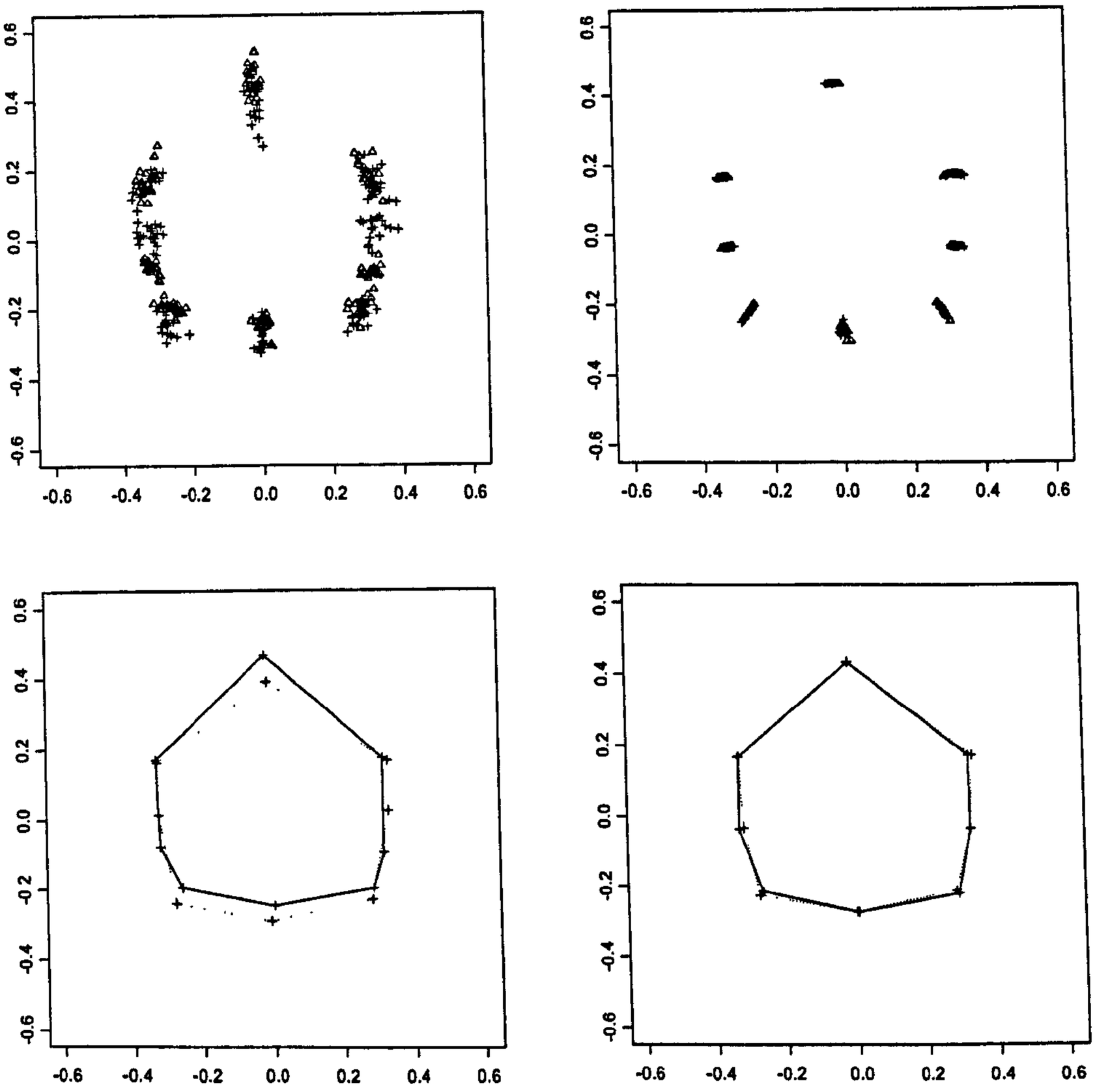


Fig. 9.13: (Top) Procrustes fits using a pooled sample (Δ = control, + = hypodontia). (Left) Original/Initial GPA registration. (Right) Following iterative semi-landmark procedure using Full Procrustes criterion. (Bottom) Overlaid estimated mean shapes of control (solid) and hypodontia (dashed) upper left central incisors.

9.4.2 Occlusal surface of upper central incisor

In this section we re-run the analysis of 9.3.2.1, to also allow for differences between patients due to the position of their lingual gingival tissue, in addition to any variation attributable to orientation differences at the imaging stage. As well as the three constrained semi-landmarks along the incisal edge, we now include the lingual landmark at the end of the BL width as an additional semi-landmark. This is also free to move in the BL direction but independent of the positions of the semi-landmarks along the incisal edge. Consequently, two parameters are determined at each semi-landmark stage, using the modification of equation (9.3) described in 9.2.3. As in the previous example, there will be little or no variation in the position of this additional

semi-landmark to remove 'within-cases', only 'between', since the data were obtained from the same stone casts from each patient and the landmark itself is easy to locate.

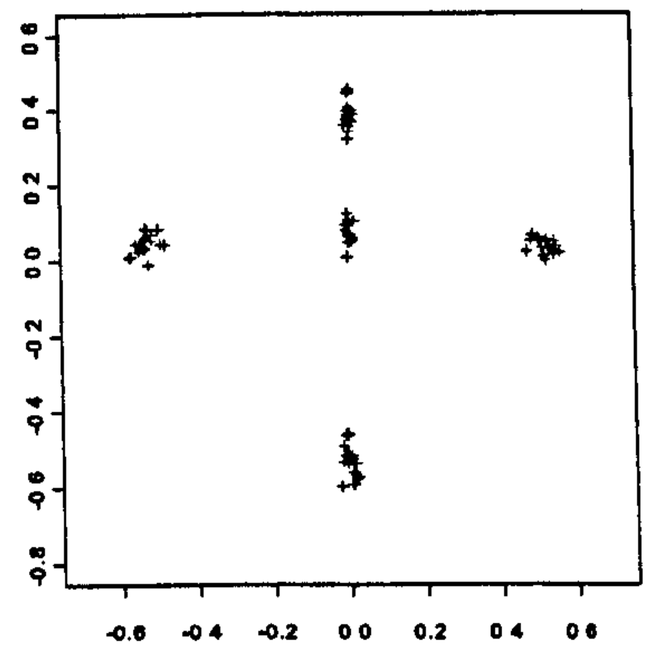
9.4.2.1 Main findings using reliability data

Table 9.10 details the iteration history of the change in values of *RSS* for each of the semi-landmark methods. For the BE method, there are again problems, with a ten-fold increase in *RSS* occurring on the first iteration, and (with the exception of the second iteration) additional, variable increases in *RSS* occurring from then on.

Iteration	Bending energy		Full Procrustes		Nearest point	
	RSS	ΔRSS	*RSS*	*RSS*	ΔRSS	*RSS*
Original	0.6760433		0.6760433		0.6760433	
1	6.9252550	-6.2492117	0.1303294	0.5457140	0.2672252	0.4088181
2	6.9193269	0.0059282	**0.1301726**	**0.0001568**	0.2136984	0.0535268
3	6.9209825	-0.0016556	0.1300151	0.0001575	0.1907041	0.0229943
4	6.9219948	-0.0010123	0.1298570	0.0001581	0.1748090	0.0158951
5	6.9232860	-0.0012912	0.1296982	0.0001588	0.1631265	0.0116825
6	6.9243556	-0.0010696	0.1295386	0.0001595	0.1545107	0.0086158
7	6.9260214	-0.0016658	0.1293784	0.0001602	0.1481677	0.0063430
8	6.9267609	-0.0007395	0.1292175	0.0001609	0.1435078	0.0046598
9	6.9282433	-0.0014824	0.1290559	0.0001616	0.1400921	0.0034158
10	6.9293100	-0.0010667	0.1288936	0.0001623	0.1375940	0.0024981
11	6.9305978	-0.0012878	0.1287306	0.0001630	0.1357716	0.0018224
12	6.9316735	-0.0010756	0.1285669	0.0001637	0.1344460	0.0013256
13	6.9333517	-0.0016782	0.1284024	0.0001644	**0.1334850**	**0.0009610**
14	6.9339009	-0.0005492	0.1282373	0.0001651	0.1327914	0.0006936
15	6.9355990	-0.0016981	0.1280715	0.0001658	0.1322935	0.0004979
16	6.9363314	-0.0007324	0.1279049	0.0001665	0.1319389	0.0003547
17	6.9374837	-0.0011524	0.1277377	0.0001673	0.1316889	0.0002500
18	6.9385888	-0.0011051	0.1275697	0.0001680	0.1315154	0.0001735
19	6.9401798	-0.0015909	0.1274010	0.0001687	0.1313976	0.0001177
20	6.9408795	-0.0006998	0.1272316	0.0001694	0.1313207	0.0000770

Table 9.10: Iteration history of *RSS* and ΔRSS values for different semi-landmark methods.

Fig. 9.14 plots the original (all landmarks fixed) and final Procrustes registered configurations after using each of the semi-landmark methods. Table 9.11 details the usual summary measures obtained from the final sets of 4×(n=19) configurations. For comparison, results of running the BE method to the second iteration have been included (although, as in 9.4.3.1, it is questionable as to whether the process should have been run beyond the first iteration). On this occasion slightly different final values of *RSS* were obtained using a convergence of ΔRSS<0.001 with the FP and NP criteria. However, any visual difference between the plots of the final sets of configurations is barely noticeable and had a stricter criterion of ΔRSS<0.0001 been used, the results would have been even more similar (see table 9.10).



Original / Initial
(all landmarks fixed)

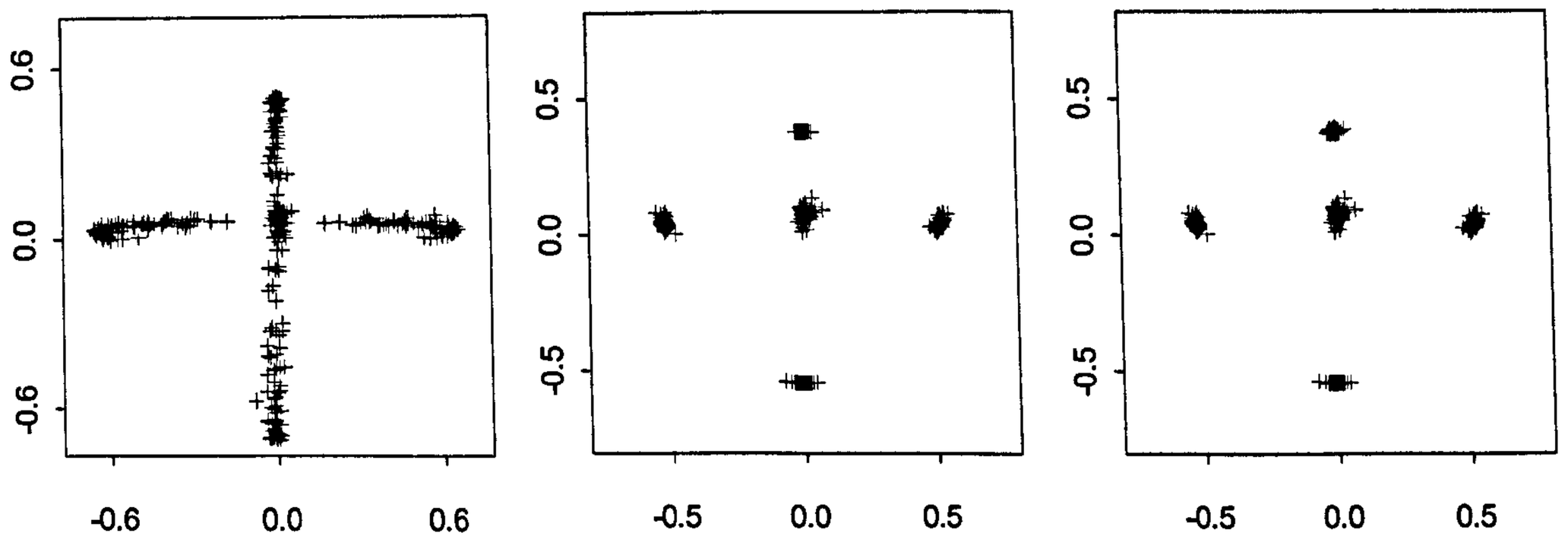
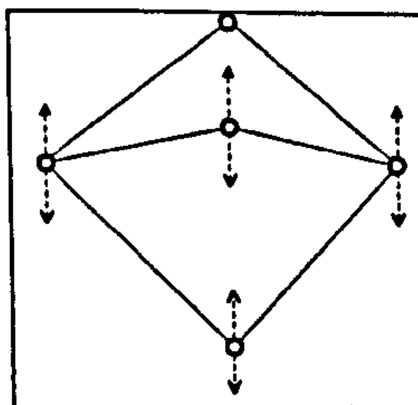


Fig 9.14: Procrustes registered configurations. (Top) Original/initial superimposition (all landmarks fixed). (Bottom) After running semi-landmark procedure using different criteria.



Upper central incisor

	None (all fixed)	Bending energy*	Full Procrustes	Nearest point
EucSS (cases)	0.273	2.200	0.043	0.045
EucSS (within)	0.403	4.719	0.087	0.089
EucSS (ops)	0.230	0.363	0.010	0.011
EucSS (errors)	0.173	4.356	0.077	0.078
EucSS (total)=*RSS*	0.676	6.919	0.130	0.133
Overall reliability	**0.297**	**0.112**	**0.135**	**0.142**
$d_F(\hat{\mu}_{SL}, \hat{\mu}_0)$		0.003	0.002	0.004

Table 9.11: Summary measures based on final Procrustes fits (landmarks along incisal edge and lingual endpoint of BL width as semi-landmarks). * ΔRSS <0.006 used for convergence.

9.4.2.2 Subsequent analysis

Using either the Full Procrustes or nearest point methods any variation in the position of the labial endpoint in the BL direction is eliminated. This means that information on the relative dimensions of the MD and BL dimensions is lost and the remaining variation in shape reduced to just the shape of the three landmarks along the incisal edge

and its angle to the BL direction. This begs the question as to whether it is still more efficient to perform any subsequent analysis of shape on the positions of the five Procrustes coordinates after using either of the new semi-landmark methods (10 variables, with dimension 7), or to the two angular and one ratio variables shown in fig. 9.15 below, which may be used to represent the same remaining variation in shape.

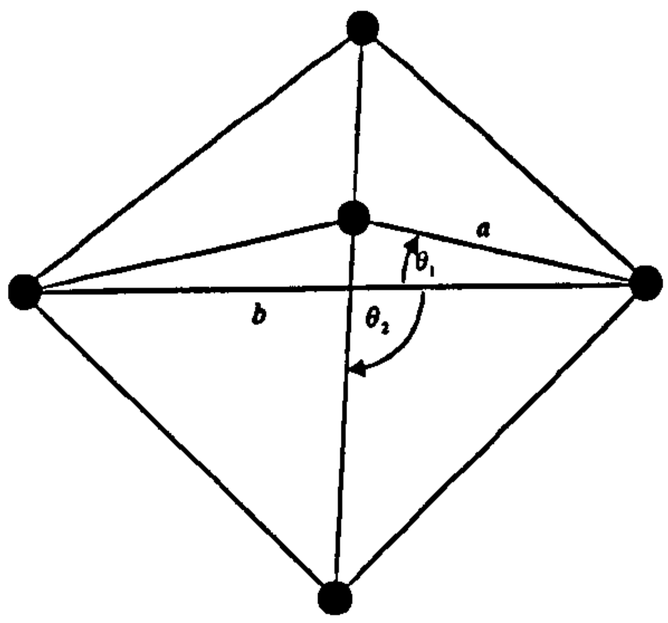


Fig 9.15: Remaining variation in shape after using nearest point or Full Procrustes semi-landmark methods, represented as angles θ_1, θ_2 and ratio *a/b*.

However, as table 9.11 shows, the final variation in shape within cases is still much larger than that between them. Of the variation that does remain, most is due to operator inconsistency, rather than actual differences between patients. Although the reliability measures will be reduced by the fact that we set out to remove variation between patients, the overall value of 0.14 for the FP and NP methods, represents a substantial decrease from the value of 0.52 in table 9.6, achieved when just the landmarks along the incisal edge were allowed to move in section 9.3.2.1.

Many previous studies (for example, Peck & Peck, 1972) have shown the ratio of the BL and MD dimensions to be an important variable for comparing the shape of teeth between populations, despite the effect of any variation on the lingual BL endpoint. Here however, this information is lost as soon we allow the variation in this landmark, in the BL direction, to be removed. Since there is no documented material on the extent of variation in the lingual gingival tissue and its effect on the proportion of the occulsal surface which may be obscured, we have no way of knowing exactly how important this issue may be; only that it has some potential to affect any measures of shape involving the BL dimension. Given the success with which the MD/BL ratio has been used in the past, perhaps the best way to proceed would be as in 9.4.2.1, where just the orientation inconsistencies in the position of the incisal edge along the BL width are removed. The BL endpoint would then still be regarded as fixed and the MD to BL ratio retained.

This leaves the configurations with a reasonable amount of reliability and in addition to the MD/BL ratio, information such as the triangulation of landmarks along the incisal edge and the angle between the MD and BL dimensions would also be represented.

9.4.3 Further examination of results using bending energy method

As in the gingival simulation study, with the data sets and applications in 9.4.1 and 9.4.2, use of the BE method again results in unrealistic shapes being produced on the first iteration, causing a large, unwanted increase in the value of *RSS*. Here we again explore these results in more detail.

In 9.4.1, the gingival and the MD landmarks were free to move in a predominantly vertical direction and those at the incisal corners in a more horizontal or diagonal direction. In 9.4.2 the landmarks along the incisal edge and at the lingual endpoint of the BL width were free to move along chords in the BL direction, subject to the constraint that the triangulation of landmarks along the incisal edge is retained. Figs. 9.16 and 9.17 consider what happens to two of the configurations from each data set in each application, on the first iteration.

Recall from 5.6.3 or 9.3.3.2.1 that the semi-landmarks in $vec^{-1}(vec(X^P)-U\lambda)$ move to positions which produce the PTPS mapping from mean $\hat{\mu}$ with closest possible 'affine fit', to the fixed landmarks and positions along the semi-landmark's chords. The 'fit' is measured by the generalised sum of squared residual non-affine components or BE, so that there is the smallest possible variation in the shape of the grid cells of the deformation grid representing the PTPS mapping. In each of figures 9.16 and 9.17, the first and third columns display the transformation from $\hat{\mu}$ to the initial Procrustes fit X^P of each configuration, as a deformation grid. The bottom panel shows the affine mapping of $\hat{\mu}$ to X^P with diagonal crosses indicating the fitted points (to which residual, non affine components must be added, so that $\hat{\mu}$ is mapped to X^P exactly). The second and fourth columns display the PTPS mapping from $\hat{\mu}$ to $X^{new}=vec^{-1}(vec(X^P)-U\lambda)$, after the semi-landmarks have been allowed to move along their chords, with the bottom panels showing the affine map of $\hat{\mu}$ to X^{new}.

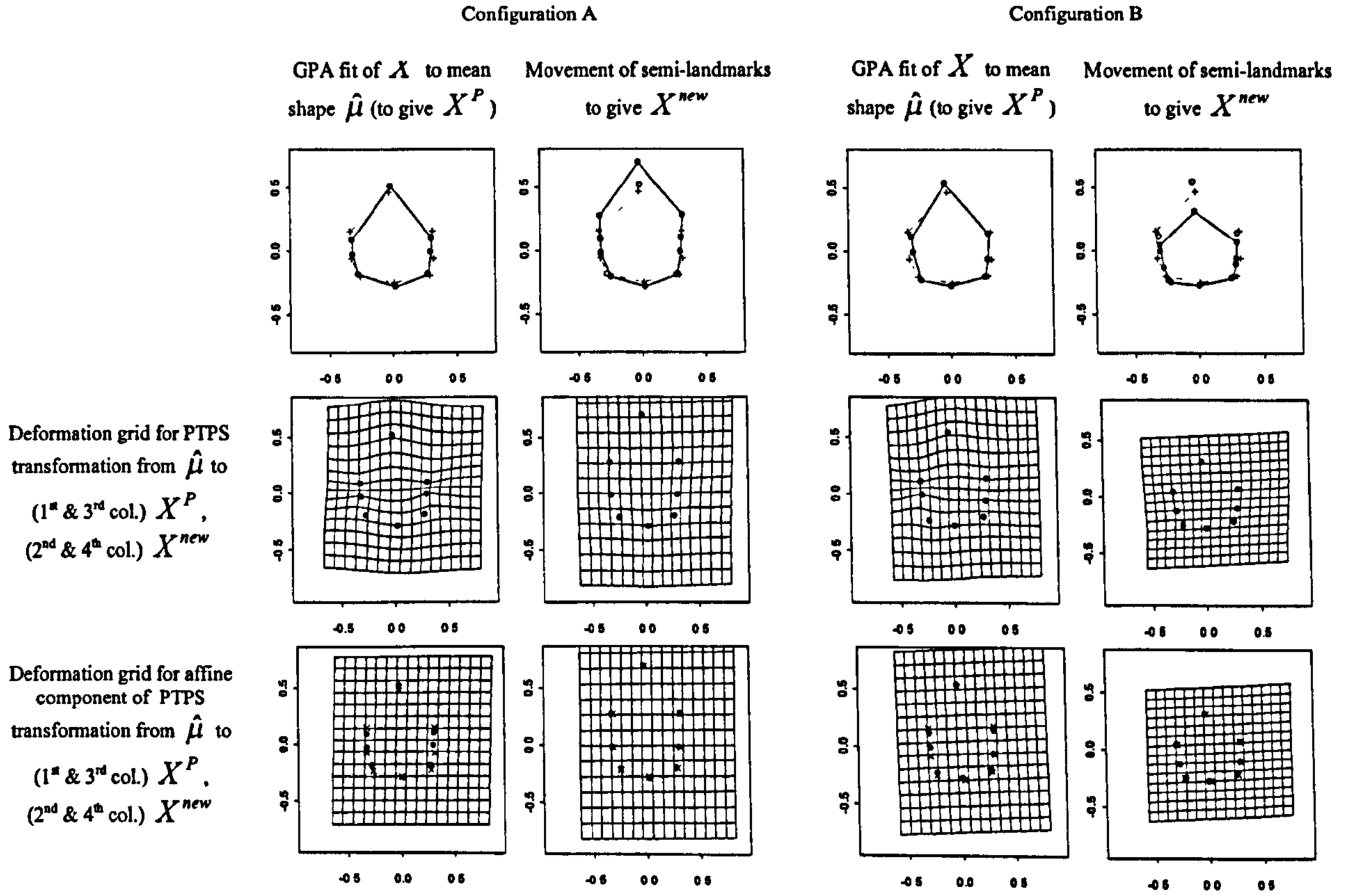


Fig 9.16: First movement step of min. BE method for two configurations in application of 9.4.1.

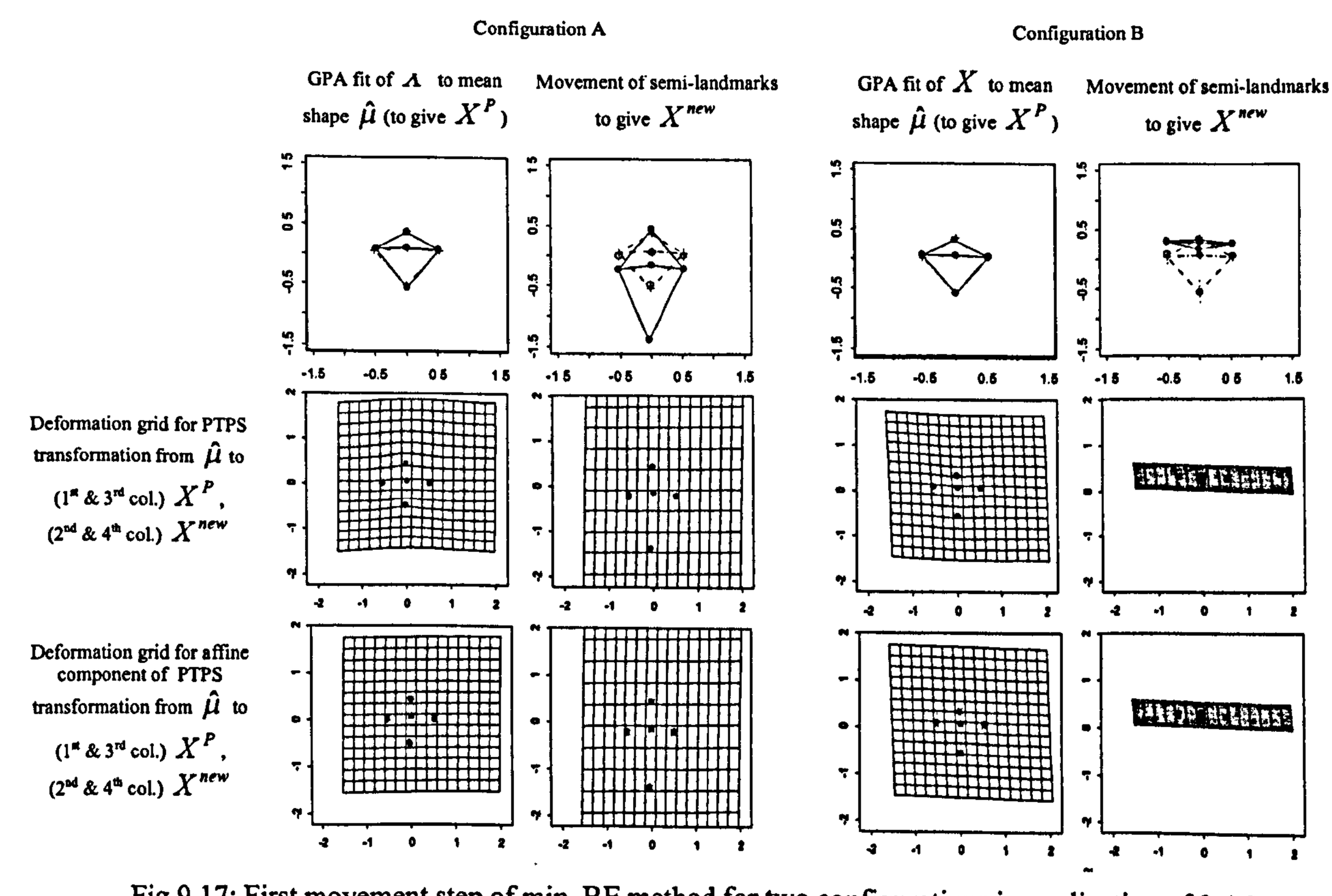


Fig 9.17: First movement step of min. BE method for two configurations in application of 9.4.2.

Key for both figs: (Top row) (1st and 3rd cols.) Procrustes fit X^P (filled circles, solid lines) to $\hat{\mu}$ (plusses, dashed lines), (2nd and 4th cols.) new configuration X^{new} (filled circles, solid lines) after semi-landmarks move from nominal positions (empty circles) along escribed chords according to minimum bending energy criterion. (Bottom row) (1st and 3rd cols.) Affine mapping of $\hat{\mu}$ to X^P and (2nd and 4th rows) X^{new} (filled circles), with fitted values shown as diagonal crosses.

In fig. 9.16, the optimal positions for each semi-landmark to adopt are again in accordance with the PTPS transformation from $\hat{\mu}$ in the direction of each chord, determined by the mapping between landmarks effectively 'fixed' in each direction. Since the chords of the gingival and MD semi-landmarks are in roughly vertical directions, the positions determining the vertical mapping and therefore any movement of the semi-landmarks, are the two corner semi-landmarks at the ends of the incisal edge (effectively constrained in the vertical direction) and the fixed landmark at the incisal end of the LACC. Similarly, a change in the positions of the two corner landmarks, only results if the new horizontal affine component of the PTPS transformation provides an improved fit in this direction to the (vertical) chords of the gingival and MD semi-landmarks and the other fixed landmark.

For configuration *A* in fig. 9.16, a 'better fitting' vertical affine component (lower BE PTPS mapping from $\hat{\mu}$) is able to be produced by the MD and gingival semi-landmarks adopting positions in accordance with the mapping to just the three landmarks along the incisal edge in $vec^{-1}(vec(X^P)-U\lambda)$. Here the vertical distance between the central (incisal end of the LACC) landmark and the two corner landmarks is relatively smaller in $\hat{\mu}$ (when considered alongside the other vertical distances between landmarks), than in X^P. This would require that the PTPS transformation (and in particular it's affine component) 'stretch' $\hat{\mu}$ in the vertical direction to map between these positions. Since there are no nearby fixed landmarks or other semi-landmarks with different chord directions, the gingival and MD semi-landmarks are able to move cervically (upward), so that a similar mapping may be used to map all landmarks in $\hat{\mu}$ to $vec^{-1}(vec(X^P)-U\lambda)$ (as this would require less bending). If there were such constraining features, the affine component would need to be different in this region, thereby requiring more BE in the mapping, i.e. more variation in the 'affine derivative' or shape of the grids cells in a corresponding deformation grid. In other words, applying the transformation required for the landmarks along the incisal edge to $\hat{\mu}$, would map the gingival and MD landmarks to chord positions cervical (above) their nominal positions in $vec^{-1}(vec(X^P)-U\lambda)$ and so these semi-landmarks move upwards in order to minimise the BE of the mapping (GLS residuals of the affine fit). Conversely in configuration *B* in fig.9.16, the vertical distances between the central and

corner landmarks on the incisal edge are larger (relative the other vertical distances between landmarks) in $\hat{\mu}$, compared to X^P and so the MD and gingival semi-landmarks move downwards to give the 'least bending' PTPS mapping from $\hat{\mu}$.

In contrast to the gum simulation, where the configurations of each set of non-gingival landmarks were always the same within representations of each case, here there are differences in the positions of these landmarks. Consequently, the resulting new positions for the semi-landmarks can also differ considerably between representations of the same case. In fact, the two configurations in fig. 9.16 were different operators' representations of the same case.

For the application of 9.4.2, fig 9.17 shows how with the BE method, the large increase in *RSS* on the first iteration occurs because of the differences in the triangulation of landmarks along the incisal edge. Because the four semi-landmarks are all free to move along chords in the (vertical) BL direction, relative to the one fixed landmark at the labial end of the BL width, the only constraint on where the semi-landmarks move to is that the triangulation of landmarks along the incisal edge is retained. This means that there are effectively only three landmarks to be matched in the vertical (BL) direction, and so the vertical affine component of the PTPS transformation from $\hat{\mu}$ is able to map to the y-coordinates in $vec^{-1}(vec(X^P) - U\lambda)$ exactly. In the horizontal direction, there are five positions to be matched, the horizontal coordinates along the four chords and that of the fixed landmark, but these are hardly able to change due to the chord directions all being roughly vertical and so the horizontal component is essentially unchanged. Consequently, those configurations where the central angle formed by the three landmarks along the incisal edge is smaller/larger compared to the configuration of these landmarks in $\hat{\mu}$ (with a larger/smaller vertical distance between the central landmark and those at the MD endpoints, relative to the MD width), require that $\hat{\mu}$ be stretched/shrunk in the vertical (BL) direction in order to match between these positions exactly. The semi-landmark at the labial end of the BL width is then free to move along its chord to where the corresponding landmark in $\hat{\mu}$ is sent to by this transformation.

In configuration *A* in fig 9.17, the incisal edge is more curved, compared to the mean, and so the PTPS transformation of $\hat{\mu}$ must stretch $\hat{\mu}$ vertically in order to map to the

vertical positions of these landmarks in $vec^{-1}(vec(X^P)-U\lambda)$ (see third row). The semi-landmark at the end of the BL width moves accordingly, here away from the incisal edge, so that there is no variation in the affine component of this transformation required and to ensure that the non-affine/residual components of the mapping of $\hat{\mu}$ is minimised. Conversely, in configuration B the incisal edge is almost flat and so the PTPS transformation squashes $\hat{\mu}$ vertically instead in order to map the incisal edge landmarks to these positions. As can be seen in the second row of fig 9.17, in the optimal PTPS transformation to the new versions of both configurations, no non-affine (bending) terms are required between the fitted affine part and $vec^{-1}(vec(X^P)-U\lambda)$ in the vertical direction and hardly any in the horizontal direction.

9.4.4 Inclusion of gum effects as two-stage registration procedure

Instead of running the procedures as above, with one set of semi-landmarks representing gingival variation and operator inconsistency, the iterative processes could have instead been run in two stages, using two sets of semi-landmarks.

1) with just the landmarks affected by operator inconsistencies as semi-landmarks
2) on the final configurations from 1) with the gingival landmarks as semi-landmarks

or *vice versa*.

On the occlusal surface of the upper central incisor, this would first allow the orientation inconsistencies to be filtered out by standardising the position of the incisal edge along the BL dimension, while retaining the relative dimensions of MD and BL widths as in 9.3.2.1. Variation in the lingual (gum) endpoint of the BL dimension could then be removed while leaving the relative positions of the incisal edge and buccal (labial) endpoint of the BL dimension unchanged. When considering all semi-landmarks at once, this information is lost.

Intuitively, ordering the stages in which the variation in each set of landmarks is removed as described above would make sense. Variation in the position of the lingual endpoint of the BL width should not affect orientation, since this requires that the most prominent part of the crown be visible and so the landmarks along the incisal edge will not be influenced by the position of this landmark. However, treating this landmark as

fixed in stage 1, as above, means that any unwanted variation between patients in this position, will influence the new positions of the semi-landmarks along the incisal edge, since the BE and FP criteria are dependent on the locations of all landmarks, in each configuration and the NP method on the prior registration of the reference and target configurations (which also depends on all landmarks). Conversely, if we removed the variation in BL endpoint (gum position) first, the new positions would by influenced by the positions of the landmarks along the incisal edge, which while being treated as fixed, would still contain errors due to orientation inconsistencies.

For the buccal surface of the upper central incisor, a similar two-stage procedure on the gingival, MD and incisal corner landmarks would also be affected in this way. Since the placement of the landmarks around the lower outline will not be affected by the position of the gum, and *vice versa*, a two stage registration procedure would appear sensible. However, as was seen in chapter 8 and 9.4.1, the locations to which the gingival landmarks move will be influenced by the configuration of landmarks around the lower outline, either at the movement stage (BE and FP methods) or during GPA registration (NP method) and here these landmark are unreliable. Conversely, where the MD and incisal corner landmarks move to would be influenced by the differences between patients in the position of their gingival landmarks.

In the majority of situations, it is therefore always likely to be preferable to consider all semi-landmarks together as a single set and determine their new positions simultaneously, allowing for the uncertainty at all semi-landmarks at the same time.

9.5 Summary and discussion

In this chapter we have investigated how well the original and new semi-landmark methods perform in addressing some of the main reliability problems identified in section 4.5.3, when using a standard Procrustes analysis, for the study of tooth shape. Applications involved making use of the data obtained by different operators from the buccal and/or the occlusal surfaces of upper and lower central incisors and upper first pre-molars, from which the reliability figures were produced in chapter 4. For the buccal surfaces, operator inconsistencies in the locations of landmarks along the sides of

the teeth and at the corners of their incisal edges were found to be a common problem. For the occlusal surfaces, unwanted variation in the location of landmarks along the incisal edge and cusps tips of teeth appeared to result from orientation differences in the bucco-lingual (BL) direction, at the imaging stage of data collection and (on the pre-molars) from operator inconsistencies in locating the MD endpoints. Having defined suitable chord directions to represent the directions along which lack of landmark homology occurs, our aim in each application was to filter out any differences in shape due to these inconsistencies, by allowing the affected landmarks to move iteratively along their chords during the semi-landmark procedure, with new positions determined by either the BE, FP, or NP criteria. For the occlusal surface applications, we also introduced the idea of constraining a subset of the semi-landmarks to always move by inter-dependent proportions along their chords, representing the effect of tilting the three dimensional surface in the BL direction. For both surfaces of the upper central incisor, semi-landmarks representing differences between patients in the positions of their gingival tissue were then also included and the analysis repeated. However, an assessment of reliability measures was regarded as misleading here since the variation to be removed occurs between representations of different cases rather than within.

For the reliability investigations, all three methods were successful in reducing the variation at the semi-landmark positions along the chosen chords directions. For both surfaces of the upper central incisor, a marked improvement on the original reliability figures of chapter 4 was produced by each method, indicating that most of the variation removed was 'within-cases' (due to operator inconsistencies) rather than between. However, for the upper first pre-molar, no real improvement to the reliability figures was found, although variation due to orientation inconsistencies and in the positions of the MD landmarks, accounted for only a small proportion of the within case variation previously observed in section 4.5.3.

As was found in chapter 8, with the FP and NP methods, the resulting final scatters of Procrustes fits were almost identical in each application (including when gingival semi-landmarks were added). All remaining variation in shape at the semi-landmarks was normal to the final chord directions of each configuration and almost identical summary measures were produced. Note again however, that despite taking more iterations, the NP method was always found to be notably faster in terms of processing time.

As also noted in chapter 8, with both new methods, reducing the unwanted variation at the semi-landmark positions means that this is no longer distributed around the rest of the configuration by GPA, allowing the actual variation at the fixed landmarks to be represented more accurately. By reducing the variance at the semi-landmark positions on each movement step, this always results in a decrease in the residual sums of squares of each configuration to the overall mean on the next GPA step. (GPA similarly distributes any reduction in variance at the semi-landmarks about the rest of the configuration). However, how the variance at each landmark is affected, depends on the configurations in question and the combination of semi-landamrks, chord directions and fixed landmarks being used. For example, in the gingival simulation study (with only gingival landmarks as semi-landmarks), a reduction in variance was observed at all landmarks of the upper central incisor configurations, whereas when the MD landmarks and those at the corners of the incisal edge are free to move, the reduction in variance at these positions was found to be at the expense of a slight increase and change in pattern of variance at some of the fixed landmarks.

In contrast to the gingival simulation study, in each of the reliability applications, use of the original BE criterion was found to produce more comparable results to the NP and FP methods, without the problem of unrealistic shapes being produced. This is because there are fixed landmarks in the vicinity of the semi-landmarks or the directions along which other semi-landmarks may move are non-parallel. Since the mean shape must be mapped to these positions exactly, the semi-landmarks are unable to move large distances, as the PTPS transformation from the mean cannot change too much from that to the nominal landmark positions of a configuration. In many such situations, the affine component of the PTPS mapping resembles a similarity transformation, with little stretching or sheering and so the new positions of the semi-landmarks are very similar to those determined by FP criteria. However, in these instances, some variation always remains along the directions of the escribed chords, as well as normal to them, due to the fact that the BE method uses a genralised least squares criterion, with correlated residual or non-affine components to determine where the landmarks move to (in relation to the affine component), rather than an ordinary least square criterion, as is used for the residuals of a full Procrustes superimposition.

When additional semi-landmarks were included, to represent differences between patients in the position of their gum as well as operator inconsistencies, the BE method produced unrealistic shapes on the first iteration, as in the gingival simulation study. In each application, a lower bending energy PTPS transformation is able to be produced by semi-landmarks adopting chord positions, in accordance with the mapping to those landmarks effectively fixed (by definition or as a result of their chord directions) in the direction of their chords. For some semi-landmarks this results in large movements occurring, so as to minimise the change in affine derivative or change in affine component (BE) in these directions, where in contrast to the reliability applications, there are no nearby fixed landmarks or semi-landmarks with different chord directions that neighbouring landmarks must be mapped to, to stop this occurring

With use of the FP and NP methods in these applications, however, more promising results were produced, with both methods successfully reducing the variation at each of the semi-landmarks, leaving only that perpendicular to their final chord directions, as described above. For the buccal surface configurations, with semi-landmarks representing both operator inconsistencies and differences between patients in the position of their gum, a significant and more meaningful difference in mean shape was demonstrated between upper central incisor configurations obtained from hypodontia and control patients (despite the variation between cases being reduced to around 5% of its value when using an standard GPA). When using these methods on the occlusal surface configurations of the same teeth (with a gingival semi-landmark included in addition to those along the incisal edge, representing orientation inconsistencies), all variation in the BL direction is eliminated, but this means that important information on the relative dimensions of the MD and BL dimensions (frequently used in previous studies) is lost. Therefore, perhaps the best strategy for the analysis of such data, may be to ignore the effect of gingival variation and proceed using only constrained semi-landmarks representing orientation inconsistencies, as illustrated in 9.4.2.1.

In the following, penultimate chapter, we consider how the different semi-landmark methods perform in filtering out simulated patterns of variation on configurations generated from the simple geometric forms. In contrast to the applications considered so far, the configurations in each dataset are manufactured distortions of the same known shape, thereby providing an 'absolute' baseline, by which to assess the performance of each method.

Chapter 10

Simulated distortions of simple geometric shapes

10.1 Introduction

In this chapter we investigate use of the new and original semi-landmark methods with configurations generated from simple geometric shapes. By simulating both the fixed and semi-landmarks to vary in specific controlled ways, we are able to monitor and assess more effectively how the different methods perform when trying to remove particular patterns of variation. Here the simulated shapes are distortions of the same 'true' shape, which we would hope the various methods would be able to recognise and recover.

In the first investigation, eight equally spaced landmarks around the circumference of a unit radius circle were used as the base shape from which to generate different simulations. In the second, a unit sided square was used with landmarks at each of the corners and four others placed exactly half way along each of the sides. Simulated configurations were then generated as perturbations of all (in the case of the circle) or some (in the case of the square) of the base shape landmarks around the perimeter of the circle/square, according to some probability distribution. Since their landmarks still all lay on the outline of the circle/square, the configurations were still regarded as representations of the same base shape. Noise might then also be added to the positions of these landmarks based on a second probability distribution, so that they no longer lay exactly on the outline.

The nearest point, full Procrustes and bending energy methods were then investigated to see whether the simulated variation between the configurations, around the perimeter of the base shapes can be removed, so that they end up being regarded as the same shape

(the shape of the base configuration from which they were generated). Different parameter values, producing different ranges and patterns of variation in the positions around the perimeter and when adding noise, were considered to investigate the possible effect on the success of the different semi-landmark methods.

10.2 Materials and methods

10.2.1 Simulations on a circle

The base configuration from which the samples of configurations were produced corresponds to eight equally spaced landmarks around the circumference of a unit radius circle (the base shape) centred on (0,0). The positions were at angles 0,45,90,..,315°, from the positive x-axis as shown in fig 10.1.

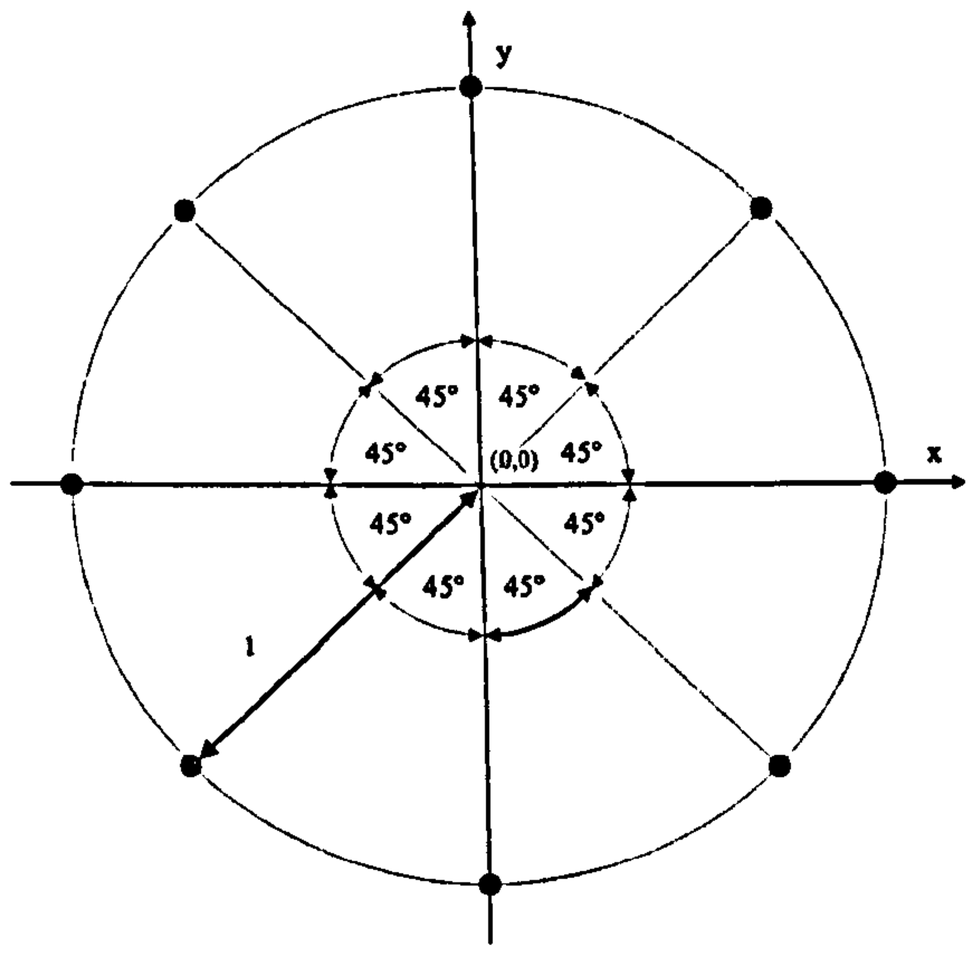


Fig 10.1: Unit radius circle, centre (0,0) with landmarks at 45° intervals from *x*-axis.

10.2.1.1 Simulation procedure

Perturbations of these landmarks were produced by sampling sets of eight independent angular values from a Uniform distribution with range (-*w*,*w*) and adding these to each of the angular positions of the landmarks in the base configuration. Positions around the circumference of a unit radius circle, centred on (0,0) and measured in terms of degrees from the positive x-axis, may then be converted to Cartesian (*x*,*y*) coordinates by using:

$$(\cos(2\pi\theta/360), \sin(2\pi\theta/360)) \qquad (10.1)$$

where multiplication by $2\pi/360$ first converts θ from degrees into radians. One hundred configurations of coordinates were generated in this way. Noise was then added to each set of landmarks, by generating pairs of values (e_{jx}, e_{jy}) from a bivariate normal distribution with mean (0,0) common standard deviation σ and zero correlation.

10.2.1.2 Values of w and σ to be investigated

In determining suitable values of w and σ, the main consideration was that the order of the landmarks as they appear around the outline should not change. One thing this prevents is 'folding' when using PTPS transformations. Consequently, if landmark j in the base configuration is originally 45° anti-clockwise (say) to landmark j', then the clockwise perturbation of landmark j around the outline, plus 'noise', should not exceed the maximum possible anti-clockwise displacement of landmark j', including noise, in the clockwise direction. This required that the displacement of any of the base landmarks around the outline, from their original positions plus any additional noise, did not exceed 22.5°.

By trial and error, a maximum value of $\sigma = 0.04$ was chosen for the standard deviations of the bivariate normal distributions. The value was regarded as large enough to allow a large range of angular displacements, produced by the uniform sampling, to be considered (without the problem described above occurring) yet still allow a sizeable scatter of noise around each point to be produced (while also ensuring that the configurations still resemble circles). For every value e_{jx} (or e_{jy}) sampled, we would expect 99.7% of these will be within 3 standard deviations of zero. This means that in 100 values of e_{jx} (or e_{jy}), the most that a point would be displaced along the circumference of the circle would be unlikely to exceed 0.12 units. (A check was made on each sampled pair, with the instruction that new values be sampled if this happened, but this proved to be unnecessary). By using simple trigonometry, a displacement of this magnitude in a direction around the outline of a circle of unit radius, corresponds approximately to an angular displacement of 7°. Therefore, in order that the base landmarks be displaced no more than 22.5° in either direction around the outline, a

(rounded down) maximum range of (-*w*,*w*)=(-15,15) was used for the uniform sampling of the angular displacements. See fig 10.2.

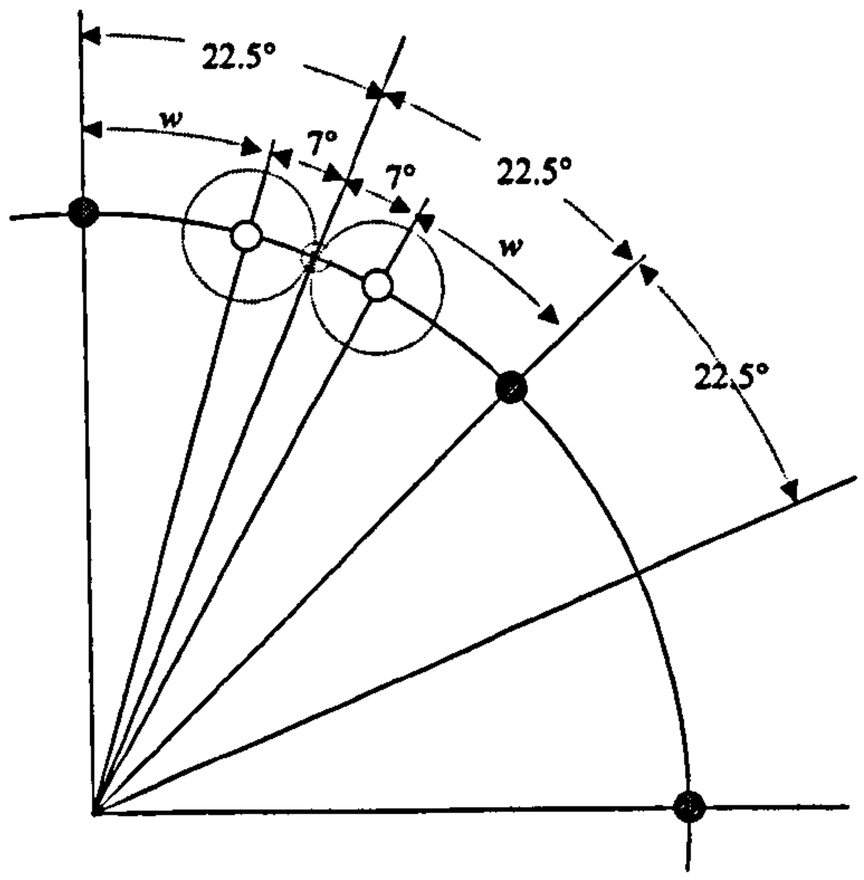


Fig 10.2: Choice of maximum range and standard deviation parameters w and σ to be used for Uniform and bivariate Normal sampling

In order to investigate a variety of different values and combinations of w and σ (within the ranges chosen above), the following eight sets of 100 simulations each were produced using the procedures described above. In fig 10.3(left) overleaf, the simulated configurations are shown in the space of the original base shape, around the outline of a circle of unit radius, centred on (0,0). In fig 10.3(right) each set of 100 configurations has been registered by GPA to their estimated mean shape. We again use the iterative technique, so that the Procrustes estimate of mean shape is the arithmetic mean of the Procrustes fits, with residual sums of squares equal to the minimised value of (3.6). The three semi-landmark methods were then run on each of the eight datasets in fig. 10.3. Each landmark was allowed to move along the escribed chord between its two neighbouring landmarks, approximating the directions of variation around the circumference of the circle, as shown in fig 10.4 below.

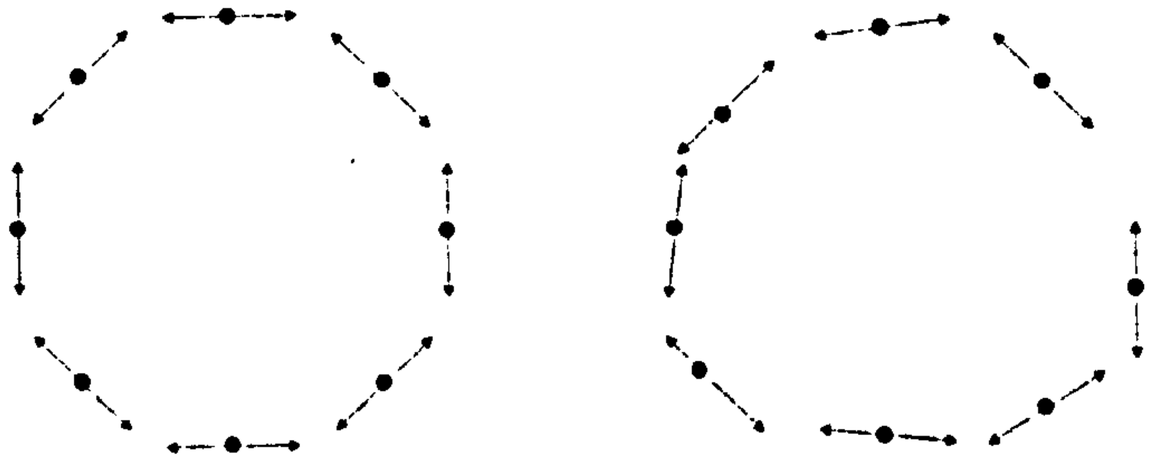


Fig 10.4: (Left) Semi-landmarks and escribed chord with no angular perturbations or added noise (Right) simulated with uniform angular perturbations and bivariate normal noise

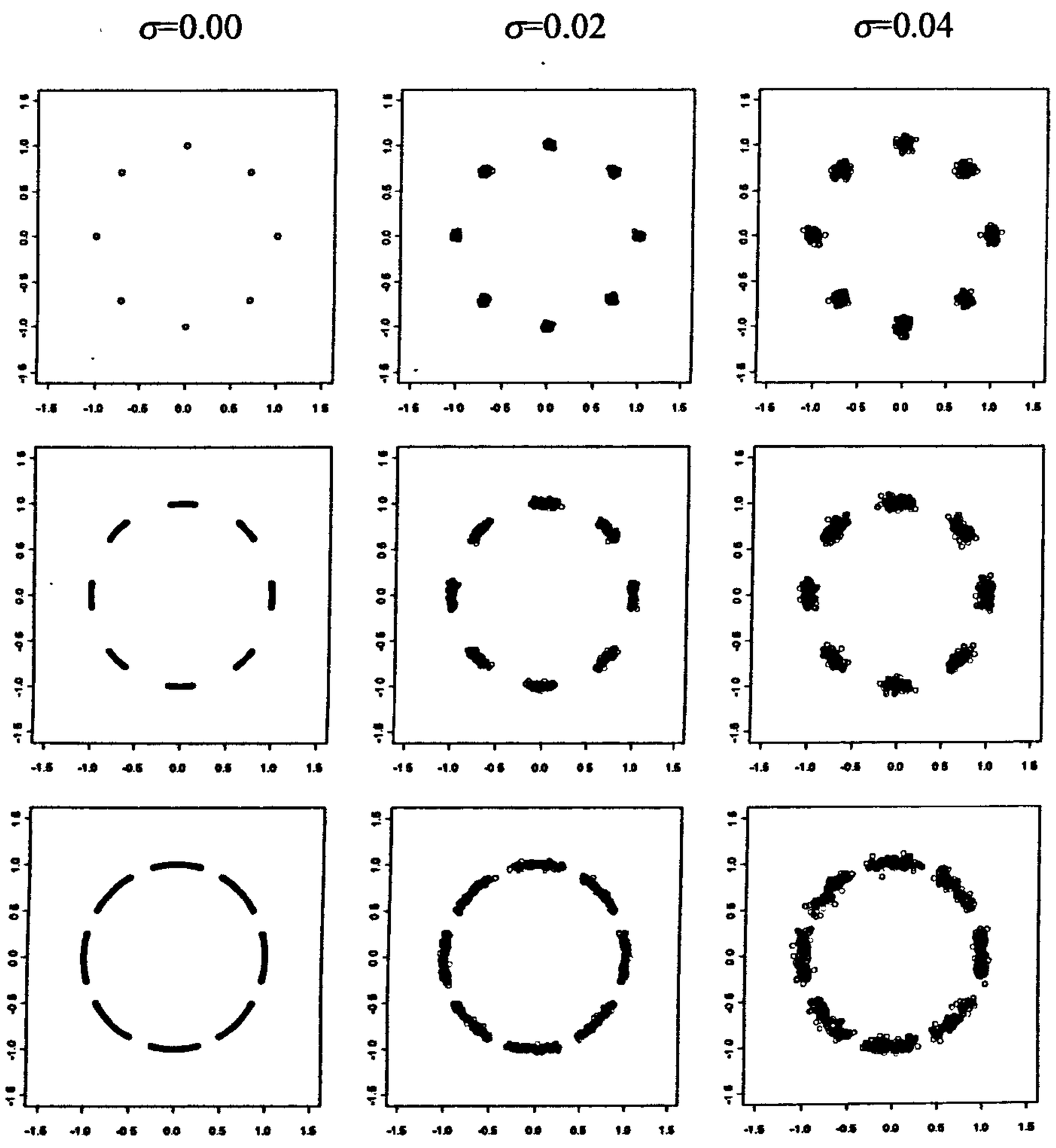


Fig 10.3(left): Sets of 100 configurations, simulated from eight equally spaced points on a unit radius circle (top left), centred on (0,0), using different values of w and σ .

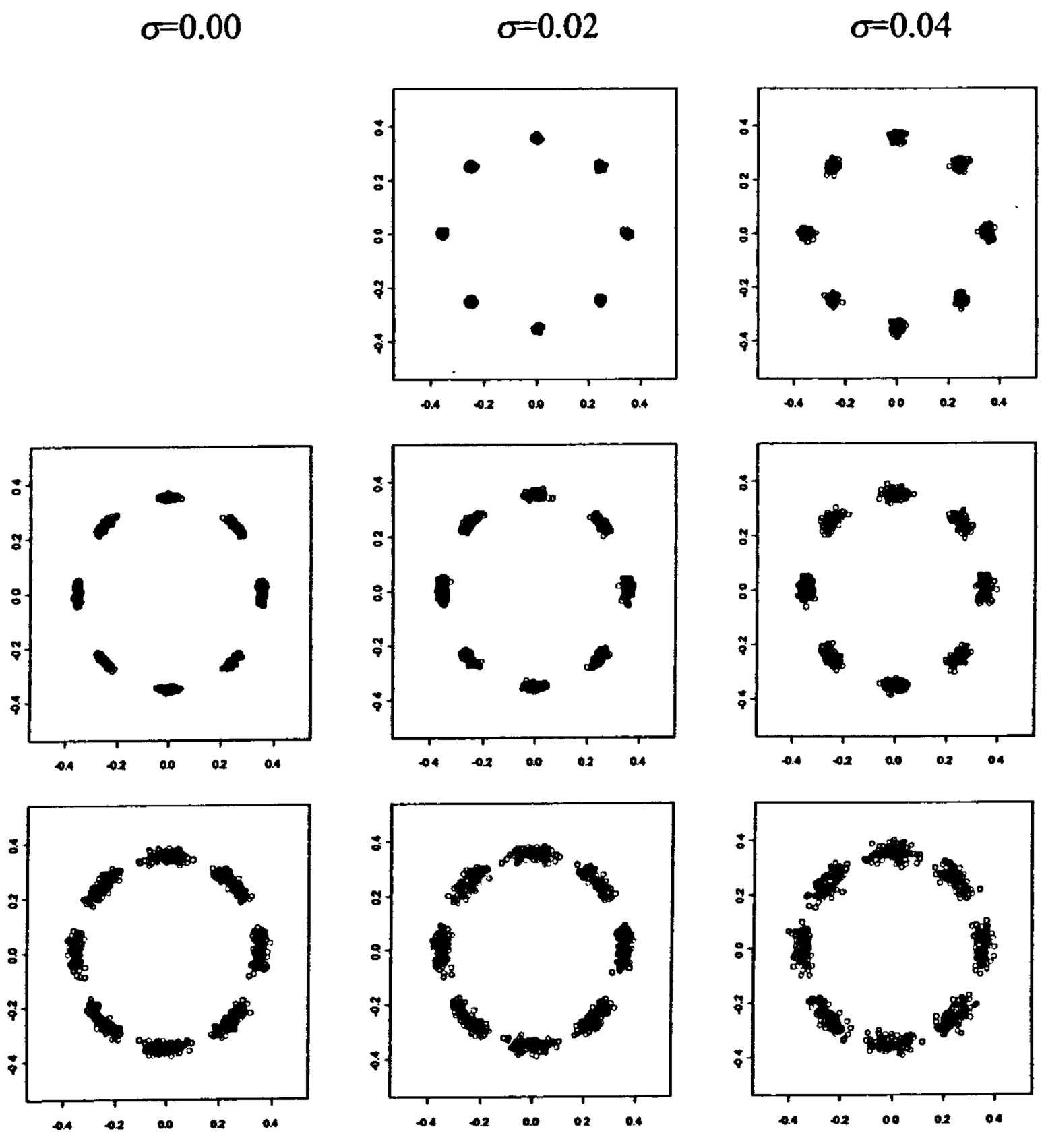
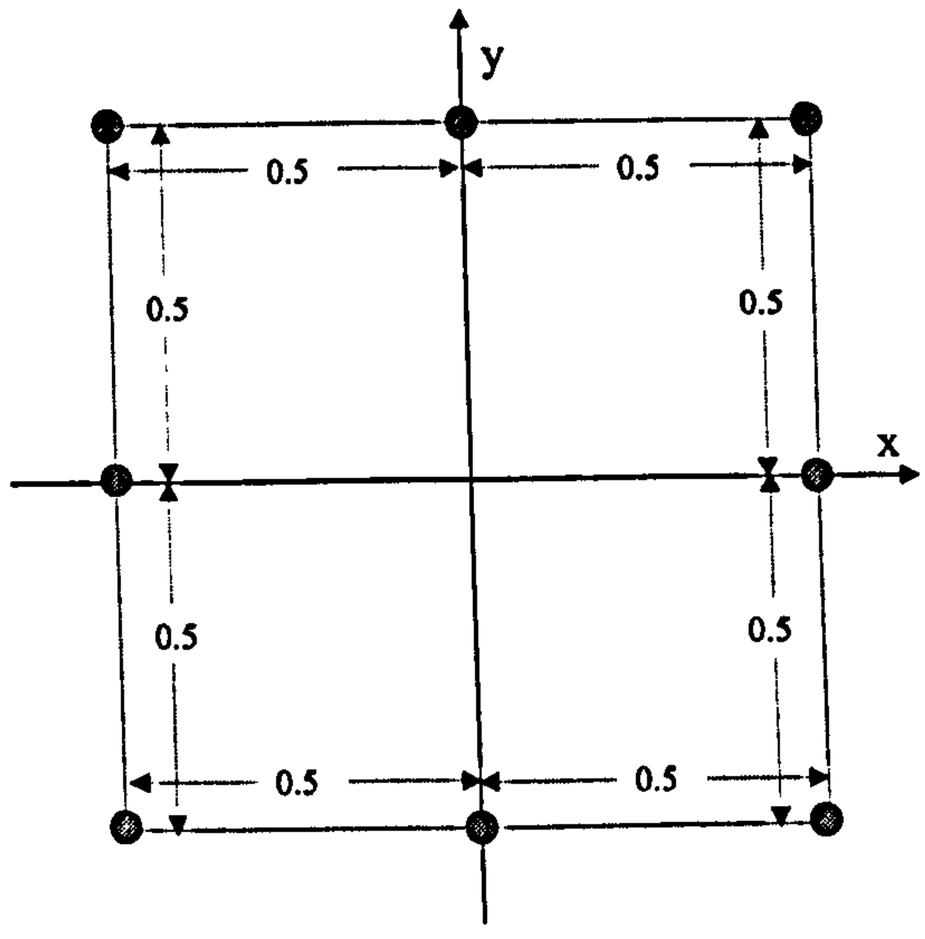


Fig 10.3(right): Procrustes fits following GPA of each set of simulated configurations in fig. 10.3(left), for each combination of values of w and σ .

10.2.2 Simulations on a square

The base configuration, from which the samples of configurations were generated, comprised four landmarks located at each of the corners of a unit sided square, centred on (0,0), at positions (-0.5,0.5), (0.5,0.5), (0.5,-0.5) and (-0.5,-0.5) and four others exactly half-way along each of the sides of the square at positions (-0.5,0), (0,0.5), (0.5,0) and (0,-0.5). See fig 10.5.

Fig 10.5: Unit sided square, centre (0,0) with landmarks at corners and mid-way along each side.



10.2.2.1 Simulation procedure

Perturbations of the four landmarks along the sides of the square were produced by again randomly sampling sets of values from a Uniform distribution with range (-w,w). These values were added to either the x or y coordinate of the side landmarks in the base configuration, depending on whether the landmark was on the top/bottom side of the square or left/right side respectively. Again n=100 configurations were generated. Noise was then added to all landmarks in each simulated configuration by again sampling sets of eight pairs of values (e_{ix}, e_{iy}) from a bivariate normal distribution with mean (0,0) common standard deviation σ and zero correlation.

10.2.2.2 Values of w and σ to be investigated

In determining suitable values of w and σ, our main consideration was again to produce a wide range of different positions for the generated landmarks while ensuring that the order of the landmarks around the outline was maintained. For example, consider the

base configuration landmark half way along the right side of the square. The perturbation of this landmark by adding a value sampled from a Uniform (-*w*,*w*) distribution to the *y*-coordinate, plus any additional bivariate 'noise' should not result in a *y*-direction displacement that exceeds the maximum possible displacement of the corner landmarks towards it by noise.

By trial and error, a maximum value of $\sigma = 0.025$ was chosen for the standard deviations of the bivariate normal distribution from which noise values were sampled. Again this value was regarded as large enough to allow wide range of perturbations along the sides of the square to be produced by the uniform sampling at the same time as a sizeable scatter of noise around each point, while also ensuring that the configurations still resemble squares. Consider again the base landmark half way along the right side of the square, after already being displaced by a uniform value. For every noise value e_{iy} sampled, we would expect 99.7% of these to be within 3 standard deviations of zero. This means that in 100 values of e_{iy}, the most we would expect the side or corner landmarks of the base configuration to be displaced by noise in the *y*-direction would be unlikely to exceed 0.075 units. (Again, during simulation a check confirmed that the e_{iy} (or e_{ix}) sampled did not exceed this value). Therefore in order for the side landmarks not to be displaced to a position that would change the order of the landmarks around the outline, a maximum range of (-*w*,*w*), with w=0.5-2×0.075, rounded down to 0.3, was used for the uniform sampling distribution. See fig. 10.6.

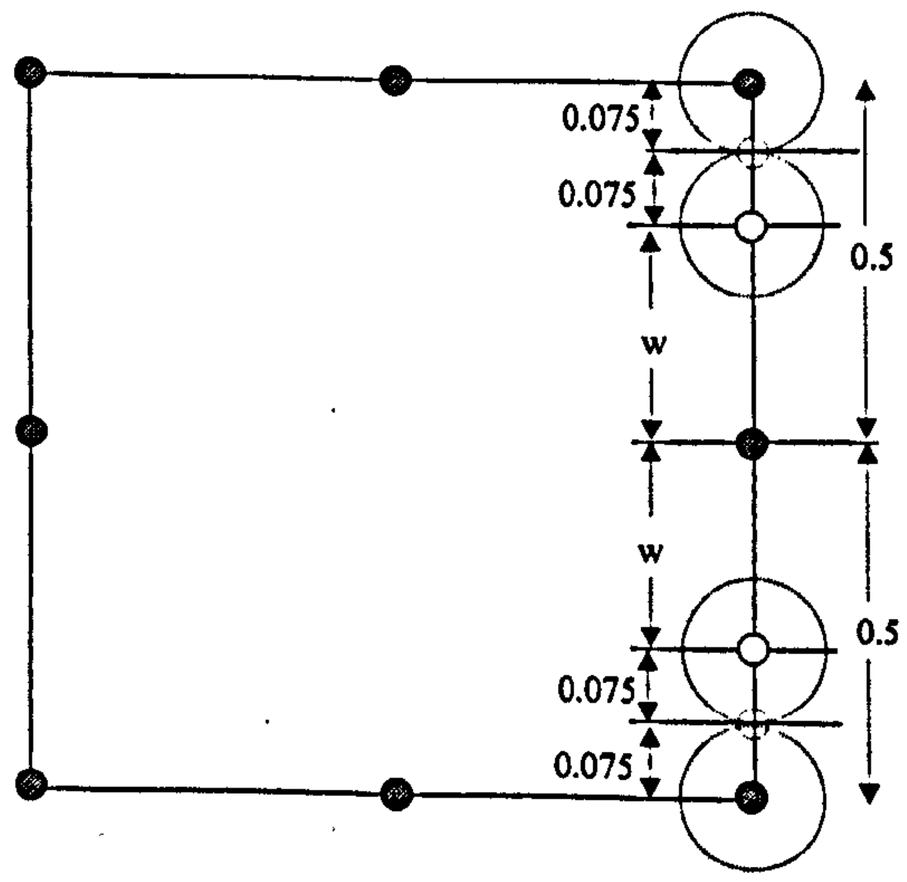


Fig 10.6: Choice of maximum range and standard deviation parameters w and σ to be used for Uniform and Bivariate Normal sampling.

In order to investigate a range of possible combinations of values of w and σ (within the maximum values chosen above) the following eight sets of 100 simulations each were produced using the procedures described above. In fig 10.7(left) overleaf, the simulated configurations are shown in the space of the original base shape, around the outline of a unit side square, centred on (0,0). In fig 10.7(right) below, each set of 100 configurations has been registered by GPA to their estimated mean shape (again using the iterative GPA technique, for the reasons described in 10.2.1.2). As with the simulations from a circle, each of the three semi-landmark methods was run on each of the eight datasets above. However, here only the landmarks generated along the sides of the base shape are allowed to move, along chords between their adjacent corners, approximating the sides of the square. See fig 10.8 below.

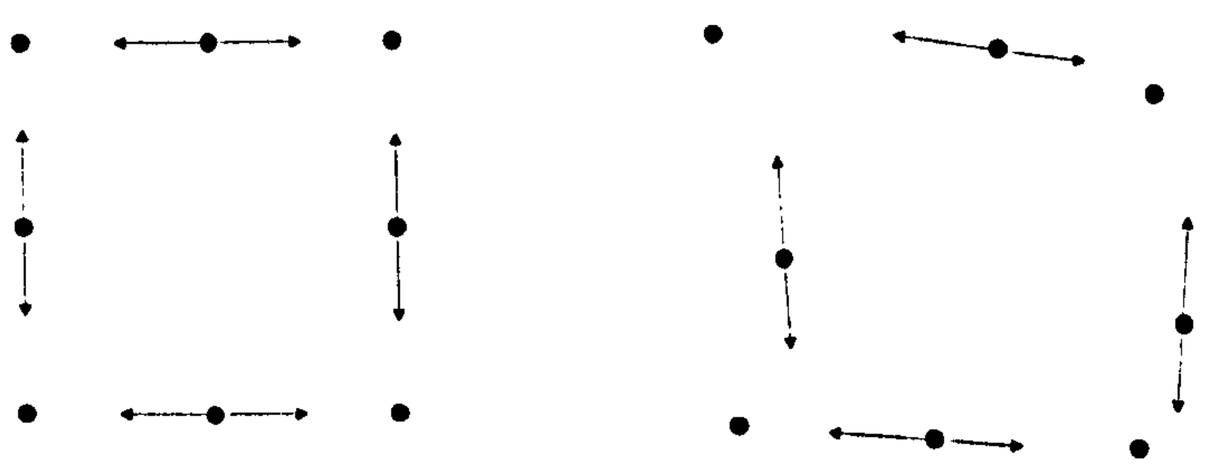


Fig. 10.8: (Left) Semi-landmarks and escribed chord with no uniform perturbations or added noise (Right) Example with uniform perturbations and simulated bivariate normal noise.

10.2.3 Choice of summary measures

The simulations on a circle are perturbations of the landmarks of a regular octagon to which circular disturbances and isotropic noise have been applied. By setting all landmarks as semi-landmarks, with escribed chord directions approximating the outline of a circle, a successful implementation of any semi-landmark methods would result in the variance along these directions (and possibly in others) being reduced as much as possible. Conversely, the simulations on a square are linear perturbations of the four landmarks at each of the midpoints along its sides and (of all landmarks) by isotropic noise. In contrast to the above, only the landmarks along the sides of the configurations are allowed to move, along escribed chords approximating the sides of a square.

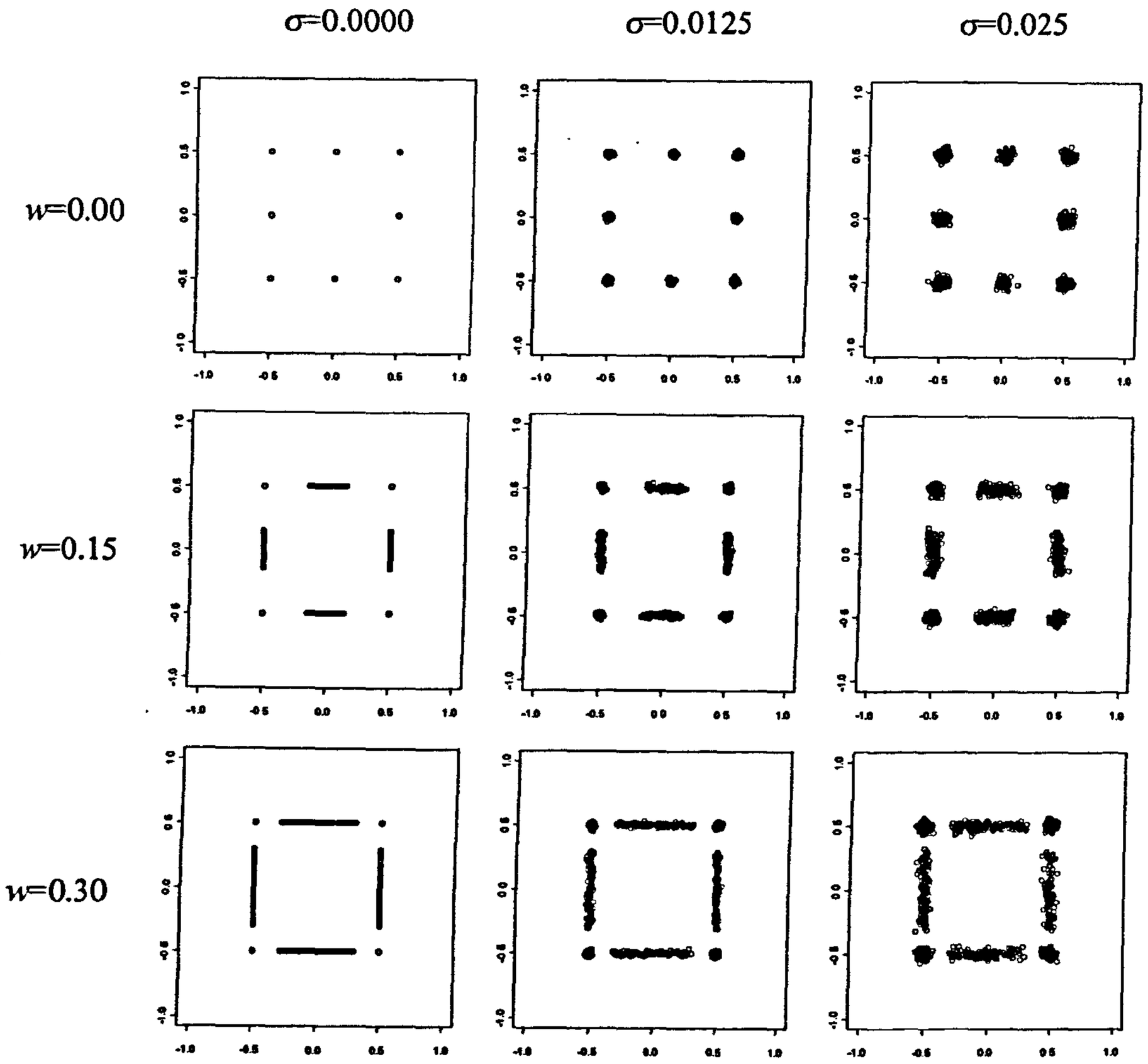


Fig 10.7(left): Sets of 100 configurations, simulated from eight points on a unit sided square, centred on (0,0), using different values of w and σ

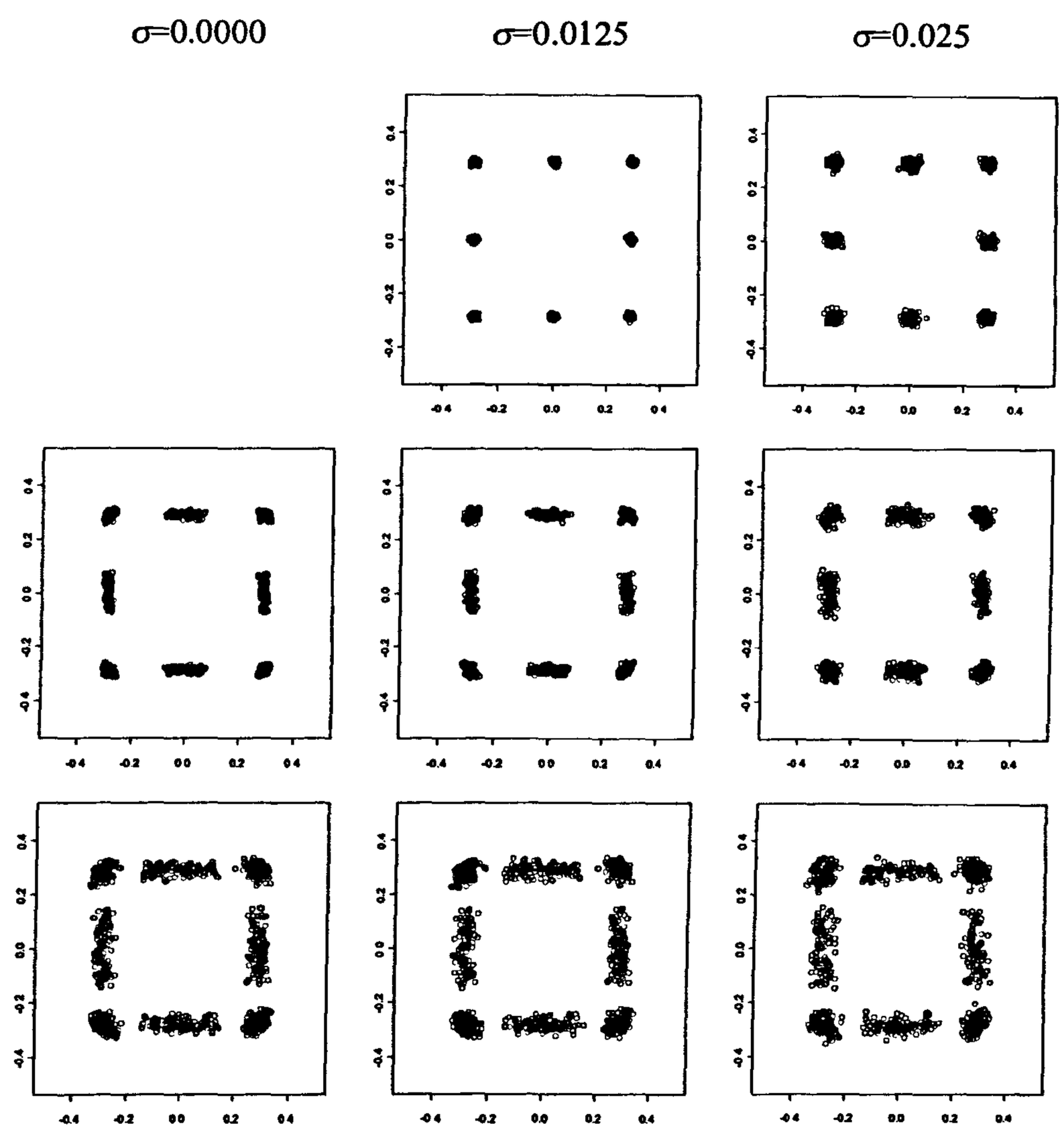


Fig 10.7(right): Procrustes fits following GPA of each set of simulated configurations in fig. 10.7(left), for each combination of values of w and σ .

A successful implementation of each semi-landmark method would again result in the variance along these directions (and possibly at the fixed landmarks and normal to the chords) being reduced as much as possible. With both sets of data, we would ideally hope that all configurations end up the same shape, with

$$RSS(\hat{\mu}_{SL}) = \sum_{i=1}^{100} \left\| X_i^{P_{SL}} - \hat{\mu}_{SL} \right\|^2 = \sum_{i=1}^{100} d_F^2(X_i^{P_{SL}}, \hat{\mu}_{SL}) \qquad (10.1)$$

as small as possible and final estimated mean the same shape as that of the base configuration. However, the extent to which this is possible may be affected by the level of noise. Since the corner landmarks on the simulations from a square are treated as fixed and noise had been added in directions perpendicular to the semi-landmarks' chords, it may not be possible to eliminate the variance at these positions entirely.

Equation (10.1) is the same value with both GPA methods, however, in keeping with these previous investigations, we again use the iterative approach, with $\hat{\mu}_{SL} = \overline{X}^{P_{SL}}$ and base summary measures on $RSS\ (= RSS(\hat{\mu}_{SL}))$. In contrast to the applications in chapters 8 and 9, however, there is no grouping structure within each simulated dataset and so a partitioning of RSS into separate sums of Euclidean norms will not be required.

In examining whether or not an elimination or reduction in variance has been achieved with each of the methods, each final RSS value must be considered in relation to the original starting value for each dataset, obtained following GPA with all landmarks treated as fixed. As can be seen in figs 10.3(right) and 10.7(right), with larger disturbances and amounts of noise used to simulate the different datasets, the original values of RSS will all be different. As w or σ is increased, so is the initial value of RSS. As well as differences in the values of RSS it may therefore also be beneficial to examine the percentage change in RSS from these starting values.

Also considered are the sum of squared Full Procrustes distances of the final configurations to the base configuration B from which they were generated. If B is first centred and scaled to unit size, then:

$$\sum_{n=1}^{100} d_F^2(X_i^{P_{SL}}, B) = \sum_{i=1}^{100} \min_{SG_i} \left\| \frac{B}{\|B\|} - SG_i(X_i^{P_{SL}}) \right\|^2 = RSS_B \qquad (10.2)$$

where *SG* denotes the similarity group of transformations of scale, location and rotation. Note also that

$$\sum_{n=1}^{100} d_F^2(X_i^{P_{SL}}, B) = RSS_B \geq RSS = \sum_{i=1}^{100} d_F^2(X_i^{P_{SL}}, \hat{\mu}_{SL}),$$

with equality only when the Full Procrustes mean shape of the fits, $\hat{\mu}_{SL}$, is identical to the base configuration *B*. This is because $\hat{\mu}_{SL}$ is determined (by GPA) as the configuration for which the sum of squared full Procrustes distances from the $X_i^{P_{SL}}$ is minimised, whereas in (10.2) each $X_i^{P_{SL}}$ is simply superimposed to pre-specified, unit sized *B*, by OPA. Again, as well as the differences in the actual values of RSS_B we also examine the percentage change from their initial starting values.

As well as reducing the values of RSS and RSS_B each procedure must also produce a final estimate of mean shape $\hat{\mu}_{SL}$ as close as possible to that of the base configuration from which they were simulated. Therefore, the full Procrustes distance $d_F(\hat{\mu}_{SL}, B)$ will also be examined, along with the change from the initial value $d_F(\hat{\mu}_0, B)$, obtained with all landmarks fixed. However, because of the way the configurations were simulated (using zero mean, symmetrical distributions), the initial estimates of mean shape, $\hat{\mu}_0$, should very nearly have exactly the same shape as the base configuration from which they were simulated, with $d_F(\hat{\mu}_0, B) \approx 0$, providing the number of simulations is sufficiently large. To allow for instances where $d_F(\hat{\mu}_0, B) > 0$, the change in shape between values of $d_F(\hat{\mu}_{SL}, \hat{\mu}_0)$ will also be noted.

In order to establish what size of change in the shape of the base configuration is still likely to be unimportant in practice, a similar investigation to that in 5.5.6 was performed. Firstly each base configuration was scaled up to have radius=250 pixels or side length=500 pixels, roughly the size each would be when located in the centre of a 1000×1000 pixel digital image. For each configuration, one landmark was then selected at random and after applying a random disturbance of either (-1,0), (1,0), (0,1) or (0,-1) to its coordinates, the full Procrustes distance between the new and original version of each configuration calculated. After repeating the process over a hundred times, the largest change in shape produced by a 1-pixel change in any one landmark coordinate was found to be no more than to 0.001 in both cases. The same process was then

repeated by randomly disturbing all 8 landmarks of each scaled base configuration by one of the 4 coordinate vectors above. The largest change in shape produced by doing so was found to be 0.004 units for the regular octagon and 0.003 units for the 8-landmark square. Consequently any values of $d_F(\hat{\mu}_{SL}, \hat{\mu}_0)$ or $d_F(\hat{\mu}_{SL}, B)$ of these magnitudes, produced by the semi-landmark methods, are unlikely to be important.

10.3 Results

For each of the semi-landmark methods and datasets, a convergence criterion of $\Delta RSS < 0.001$ was used, as in the two preceding chapters. An examination of the iteration histories (omitted here) revealed that this allowed use of the same criterion with each method on each of the 8 combinations of w and σ, allowing direct comparison of results, while stopping each process before entering its "slow manifold". This is also in agreement with the discussion of 5.5.6, which suggested an acceptable convergence criterion for ΔRSS with 100 configurations (here also of eight landmarks), in the region of $n \times 2 \times 10^{-5} = 0.002$.

10.3.1 Simulations from a circle

Fig. 10.9 displays the scatters of final Procrustes registered configurations, $X_i^{P_{SL}}$, i=1,...,100, about their estimated mean shape $\hat{\mu}_{SL}$, after running each of the full Procrustes (FP), nearest point (NP) and original bending energy (BE) semi-landmark methods, on each of the eight simulated datasets. Values of the original and final RSS values (corresponding to the variation in fig. 10.9) obtained using each technique can be seen in table 10.1(left). In table 10.1(right), are the values of RSS_B, defined in section 10.2. The figures in brackets indicate the percentage reduction in RSS and RSS_B, from their original values in the grey cell above. Fig. 10.10 presents the values of RSS_B and percentage change in RSS_B as interaction plots. The first and third rows show the effect of changes in simulated angular range ($-w$,$+w$) on separate plots for each level of noise σ, for each of the semi-landmark methods and the second and fourth rows show the effect of changes in σ at different levels of w.

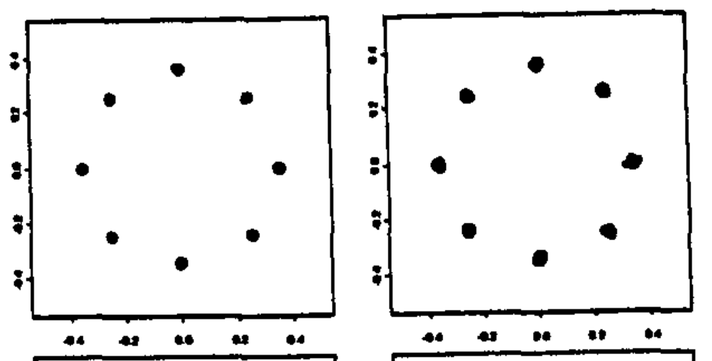
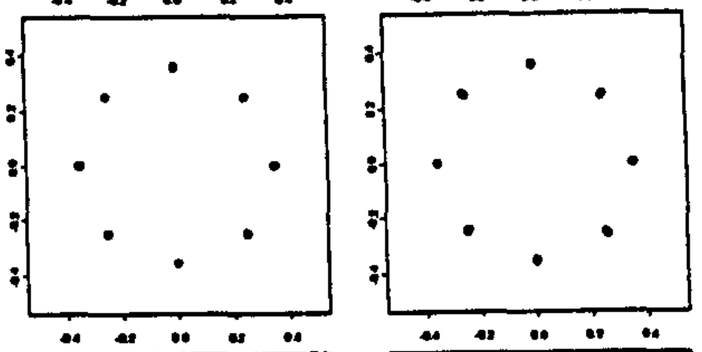
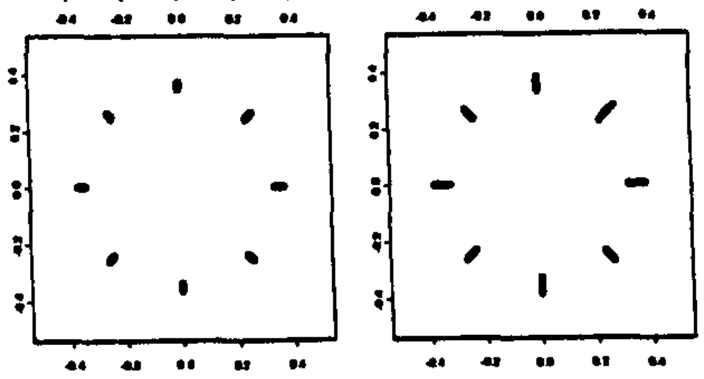
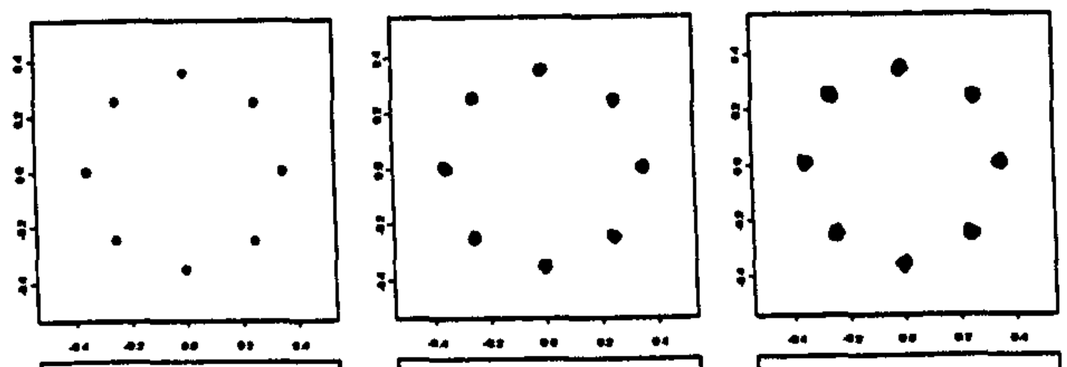
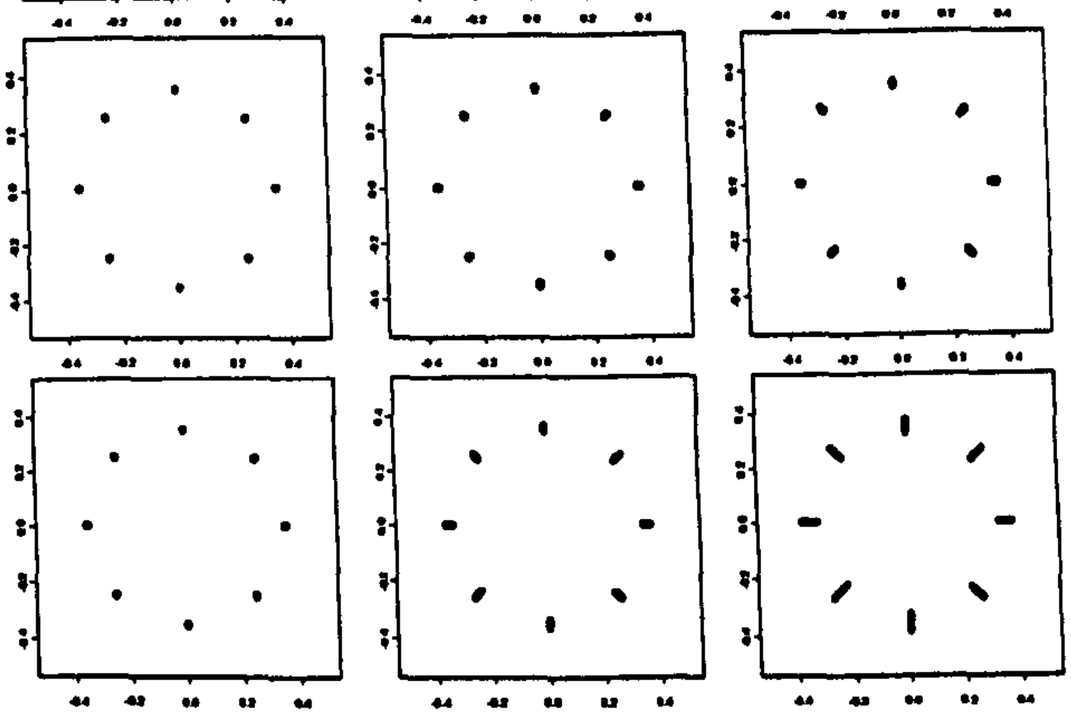
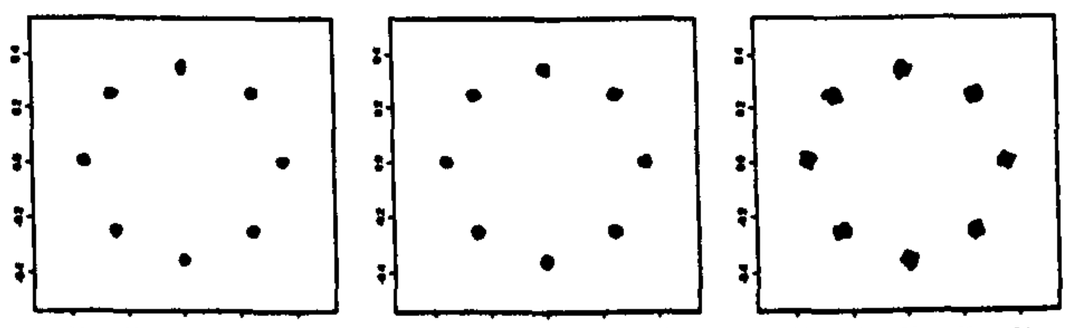
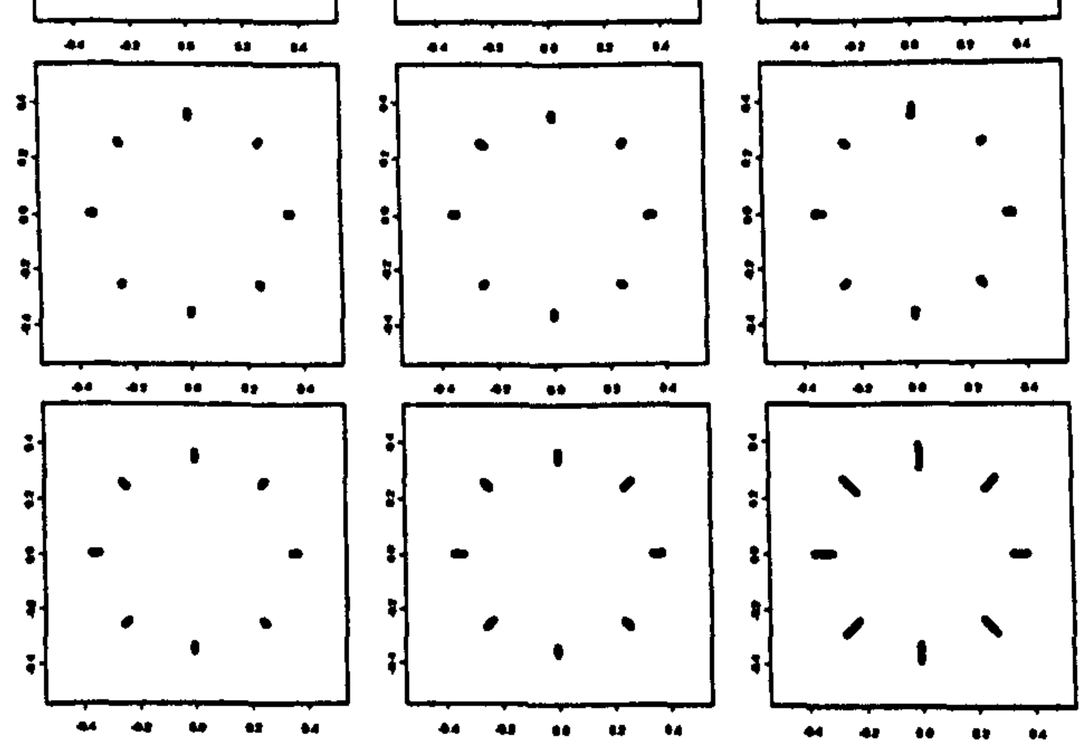
		σ=0.00	σ=0.02	σ=0.04
w=0°	Bending energy			
	Full Procrustes			
	Nearest point			
w=7.5°	Bending energy			
	Full Procrustes			
	Nearest point			
w=15°	Bending energy			
	Full Procrustes			
	Nearest point			

Fig 10.9: Final Procrustes fits after running each of the semi-landmark methods on each set of circle configurations, simulated with various combinations of angular variation w and noise σ.

	Semi-landmark criterion	RSS (%Δ RSS from original in grey)			RSS_B (%Δ RSS_B from original in grey)		
		σ=0.00	σ=0.02	σ=0.04	σ=0.02	σ=0.02	σ=0.04
w=0	None (All fixed)		0.057	0.233		0.058	0.235
	Bending energy		0.021 (62.6%)	0.079 (66.3%)		0.032 (44.8%)	0.119 (49.4%)
	Full Procrustes		0.001 (98.6%)	0.003 (98.8%)		0.003 (94.8%)	0.010 (95.7%)
	Nearest point		0.024 (58.7%)	0.094 (59.7%)		0.025 (56.9%)	0.096 (59.1%)
w =7.5	None (All fixed)	0.407	0.473	0.639	0.409	0.477	0.648
	Bending energy	0.001 (99.7%)	0.035 (92.6%)	0.095 (85.1%)	0.004 (99.0%)	0.043 (91.0%)	0.131 (79.8%)
	Full Procrustes	0.001 (99.8%)	0.005 (99.0%)	0.013 (97.9%)	0.003 (99.3%)	0.011 (97.7%)	0.039 (94.0%)
	Nearest point	0.001 (99.8%)	0.028 (94.0%)	0.095 (85.1%)	0.004 (99.0%)	0.032 (93.3%)	0.104 (84.0%)
w =15	None (All fixed)	1.720	1.808	1.973	1.733	1.819	1.989
	Bending energy	0.023 (98.7%)	0.043 (97.6%)	0.133 (93.2%)	0.036 (97.9%)	0.056 (96.9%)	0.174 (91.3%)
	Full Procrustes	0.006 (99.7%)	0.008 (99.6%)	0.016 (99.2%)	0.020 (98.8%)	0.023 (98.7%)	0.050 (97.5%)
	Nearest point	0.013 (99.2%)	0.034 (98.1%)	0.116 (94.1%)	0.025 (98.6%)	0.044 (97.6%)	0.132 (93.4%)

Table 10.1: Final values of RSS and RSS_B for configurations simulated from a circle using various *w* and σ.

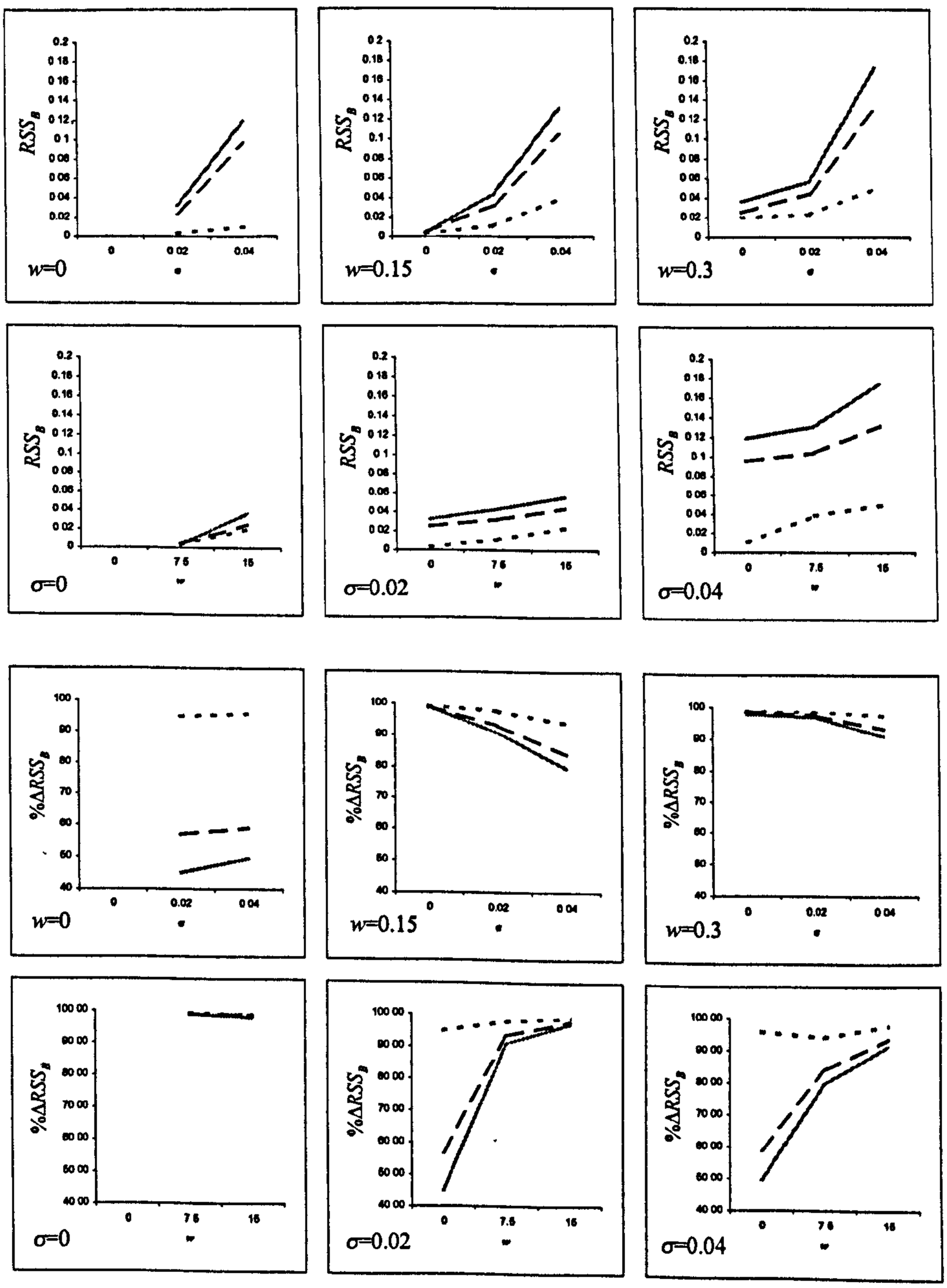


Fig 10.10 (top two rows): Final values of RSS_B for configurations simulated from a circle using various *w* and σ. (Bottom two rows): Percentage change in RSS_B from original value (all landmarks fixed).

10.3.1.1 Main observations

As was found in all previous applications, an examination of the final Procrustes fits (fig. 10.9), after running each of the semi-landmark methods, shows that when the FP or NP criteria are used, all remaining variation at the semi-landmark positions is perpendicular to their final chord directions. The BE method was also successful in reducing the variation along these directions, but again that which remains appears more isotropic. In contrast to the results of previous chapters however, here the variance in the final semi-landmark positions can be seen to be always notably larger for the NP method, compared to the FP method. This is most evident for the datasets simulated with greater noise σ (and to a lesser extent is also true for larger *w*). Even when a lower level of convergence (0.0001) was used, the visual differences in the scatters of points and in values in table 10.1 remained and so this result is not due to differences in the actual level of convergence reached. In addition, the NP method took far fewer iterations to converge compared to previous applications (see table 10.2 below).

	Semi-landmark criterion	Number of iterations		
		σ=0.00	σ=0.02	σ=0.04
w=0	Bending energy		6	6
	Full Procrustes		4	5
	Nearest point		3	4
w=7.5	Bending energy	2	5	7
	Full Procrustes	2	7	6
	Nearest point	4	4	5
w=15	Bending energy	3	5	7
	Full Procrustes	4	5	6
	Nearest point	5	5	6

Table 10.2: Number of iterations required for convergence at ΔRSS<0.001.

In table 10.1, values of RSS (and RSS_B) are always of the same magnitude, and follow similar patterns with w and σ for each method. Differences between RSS and RSS_B do however, increase slightly with w and σ, suggesting an increasing difference in shape between $\hat{\mu}_{SL}$ and base configuration B with each method, which we examine in section 10.3.1.4. For each combination of w and σ, the BE criterion always performed worst (in terms of values of RSS or RSS_B), whereas the FP method always performed best, confirming the observations in fig 10.9. Note also that the difference in values between the BE and FP methods also increases slightly with σ. We consider the results for different combinations of w and σ in more detail in section 10.3.1.3.

10.3.1.2 Differences between semi-landmark methods.

Fig. 10.11 shows what happens on the first one or two iterations when using each of the semi-landmark methods, for one of the configurations simulated with maximum amount of noise (σ=0.04), but zero angular disturbances (w=0) at each of its landmark positions. Fig. 10.12 shows the first iteration of each of the methods for one of configurations simulated with maximum angular variation (w=15) but no noise (σ=0). In the usual notation, let $X^P_{(r-1)}$ denote the nominal positions of a configuration at the start of iteration r following GPA registration to estimated mean shape $\hat{\mu}_{(r-1)}$, $vec^{-1}(vec(X^P_{(r-1)})-U_{(r)}\lambda_{(r)})=X^{new}$ the new positions of the landmarks of $X^P_{(r-1)}$, following movement of the semi-landmarks, by distances given by vector $\lambda_{(r)}$ along unit chord directions defined by matrix $U_{(r)}$. $X^P_{(r)}$ is then the Procrustes registered fit of $X^{new}_{(r)}$ to the new estimate of mean shape $\hat{\mu}_{(r)}$, obtained at the end of the rth iteration.

For configurations simulated with σ=0.04, w=0, GPA registration to the initial estimate of mean shape $\hat{\mu}_0$ (obtained with all landmarks fixed), results in isotropic variation at the residuals of each landmark as was seen in fig. 10.3(right) top right. Isotropic noise at the original landmark locations, results in isotropic errors in the Procrustes coordinates (Dryden & Mardia, 1998). For σ=0, w=15 however, GPA registration to $\hat{\mu}_0$ results in this variation being distributed more around the outline of the circle, rather than only isotropically, as in fig. 10.3(right) bottom left. For large σ, there is also greater variation in the first set of chords approximating tangent directions to a circle at each landmark. See for example, the configuration in fig 10.11, 2nd panels of part a, b and c. This will also be true (but to a lesser extent) for configurations simulated with σ=0, w=15, as can be seen in the example in fig. 10.11, 2nd panels of parts a, b and c.

For the FP method, the sum of squared distances between the corresponding landmarks of $vec^{-1}(vec(X^P_{(r-1)})-U_{(r)}\lambda_{(r)})$ and $\hat{\mu}_{(r-1)}$ is minimised, by allowing the semi-landmarks to move along their chords while $vec^{-1}(vec(X^P_{(r-1)})-U_{(r)}\lambda_{(r)})$ is superimposed to $\hat{\mu}_{(r-1)}$ over similarity transformations. Since here the chord directions are always approximately around the outline of a circle, this mainly allows an improved rotational fit to the landmarks of $\hat{\mu}_{(r-1)}$, most notably on the first iteration. In addition, some of the variation perpendicular to the circle outline is also able to be removed by varying

the scaling factor, which acts in all directions from the centre the configuration being superimposed, as the semi-landmarks move along their chords. At the end of the iteration, GPA registration to an essentially unchanged mean, with $\hat{\mu}_{(r)} \approx \hat{\mu}_{(r-1)}$ will then give a very similar result to the OPA superimposition optimised when determining the new positions of the semi-landmarks. The results are very similar, regardless of w and σ, as can be seen in figs. 10.11(b) and 10.12(b), although slightly larger values of RSS (and RSS_B), with increasing σ and w result, as the approximation by the chord directions, to the outline of a circle becomes worse (see section 10.3.1.3).

In fig 10.11, the fourth panel of the top row of (b) illustrates how use of the FP criterion on such configurations often produces large movements in the semi-landmarks along their chords in order to optimise the OLS superimposition to $\hat{\mu}$, shown in the panel below. Despite moving large distances (and further away from the landmarks of $\hat{\mu}_{(1)}$), the combined movement of the landmarks hardly affects the actual shape of the configuration, with the fit to $\hat{\mu}_{(1)}$ being only marginally improved.

Using the NP criterion, the distances between corresponding landmarks are minimised on each movement step without allowing the registration of the configuration to the mean to change. An improved rotational fit results since the chords approximate the outline of a circle; on the first iteration, the semi-landmarks move to positions along their chords at equally spaced angles around the outline (as in $\hat{\mu}_{(r-1)}$), to points at nearest Euclidean distance to their corresponding positions in $\hat{\mu}_{(r-1)}$. However, since the new positions have not been determined to optimise the superimposition of $vec^{-1}(vec(X^P_{(r-1)}) - U_{(r)}\lambda_{(r)})$ to $\hat{\mu}_{(r-1)}$ over scale and location as well, more variation will remain in directions perpendicular to the outline of $\hat{\mu}_{(r-1)}$. This is particularly true on the first iteration when the chord directions are most subject to the effects of noise, as shown in fig 10.11(c) as some chords will then not pass as near to the corresponding landmarks in the mean shape as they would if σ were smaller. To a lesser extent, this is also true for larger w (see for example, fig 10.12(c)). Here however, the chords approximate the outline of a circle more closely than when there is noise and pass nearer to the corresponding landmarks in $\hat{\mu}_{(r-1)}$ following the initial GPA registration. Consequently, so for larger w, less variation will remain normal to the circle outline, compared to with larger σ.

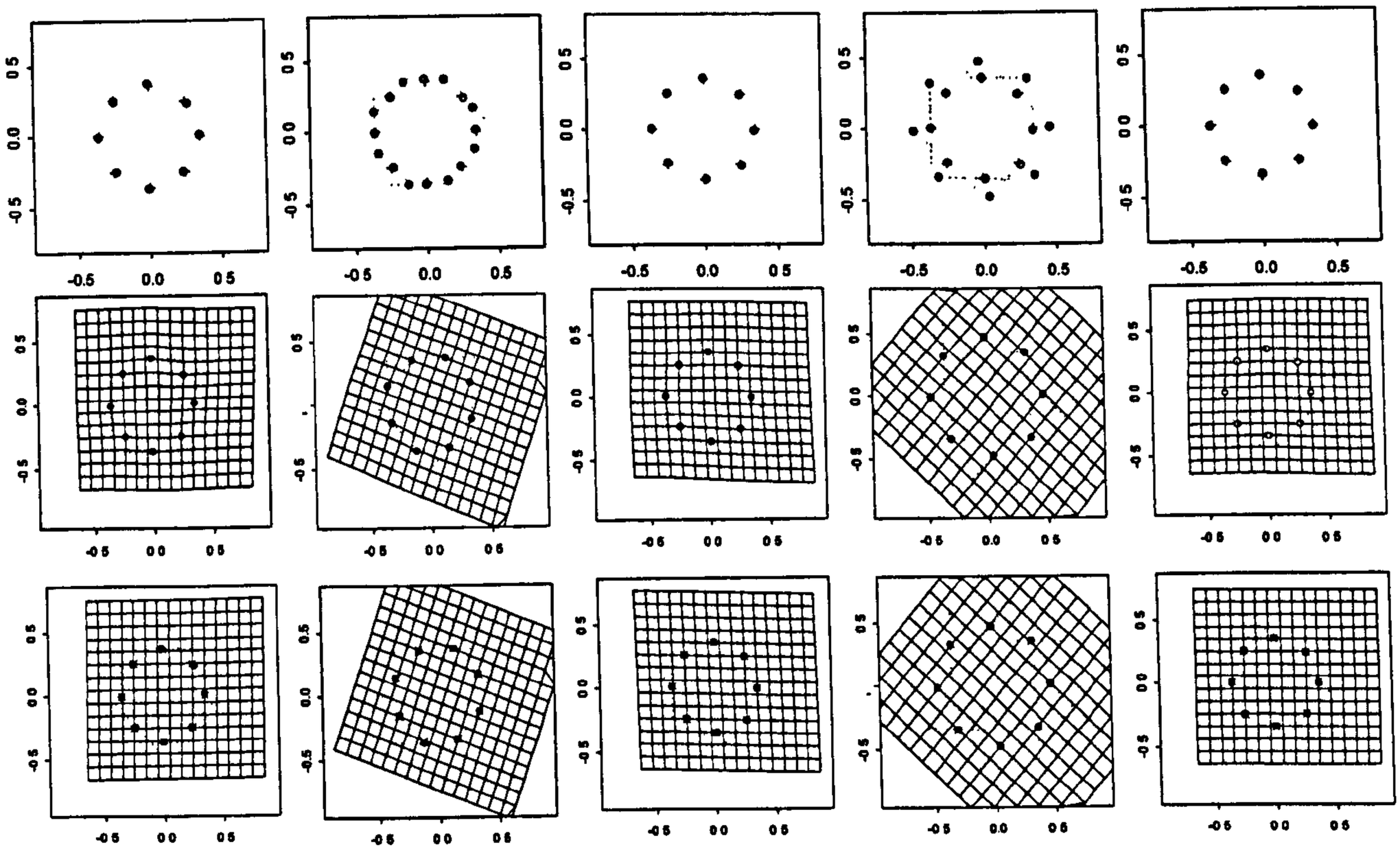
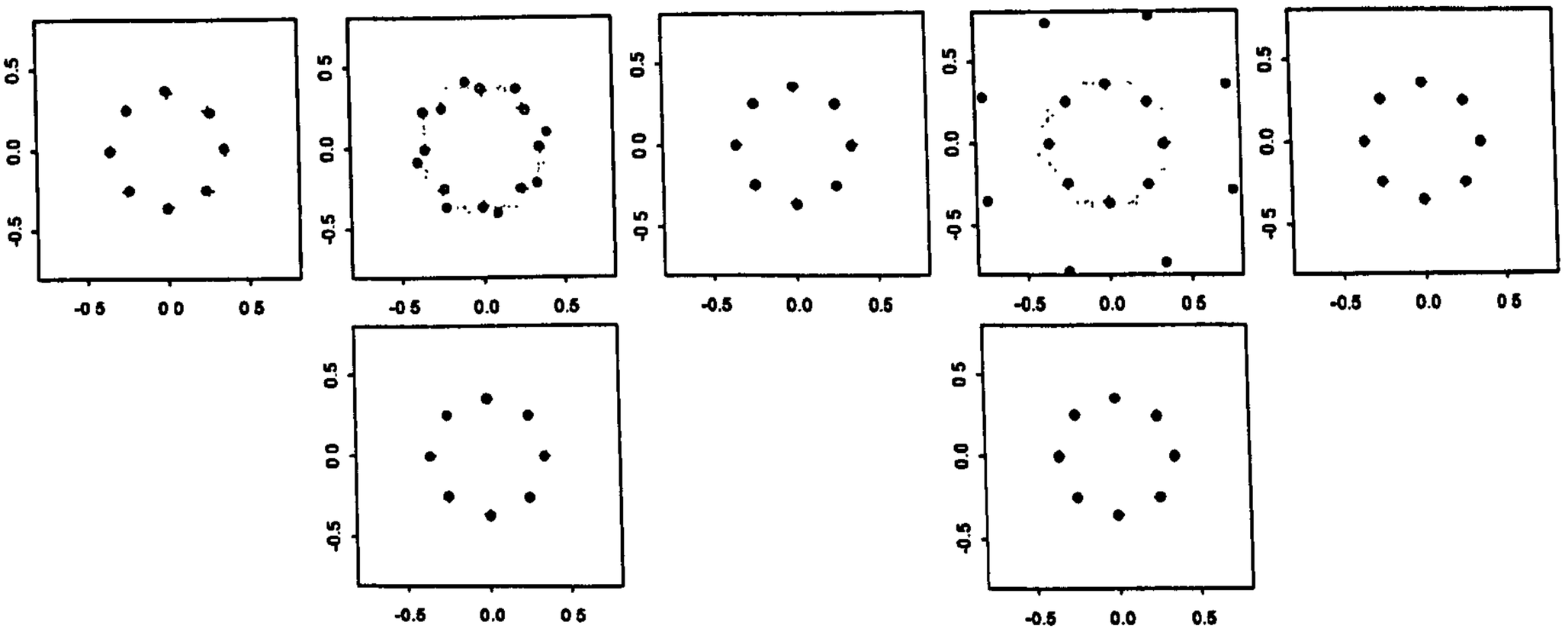
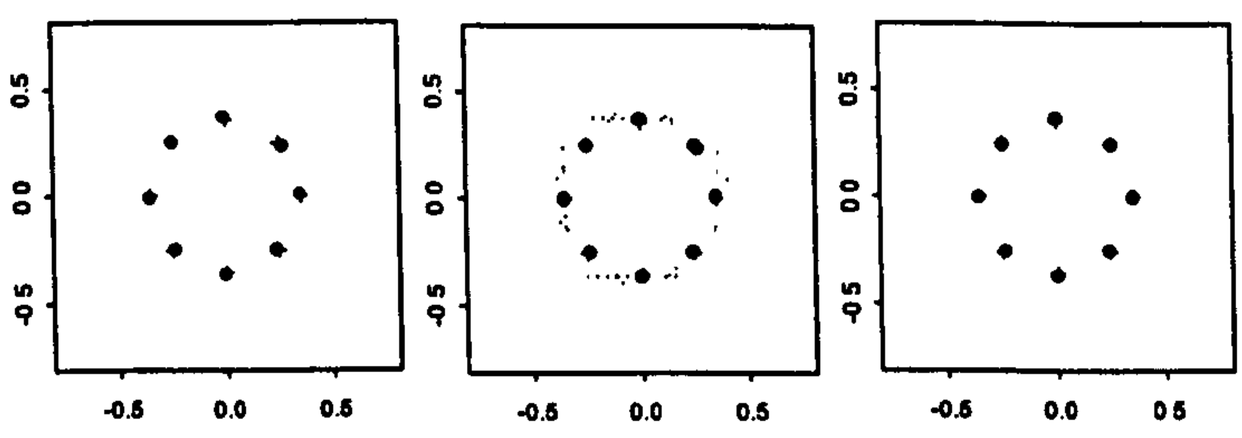
First row of parts (a-c):	Initial GPA registration of config. to $\hat{\mu}_{(0)}$ to give $X^{P}_{(0)}$	1st iteration (r=1) Movement of semi-landmarks to give $X^{new}_{(1)}$	GPA registration to new mean $\hat{\mu}_{(1)}$ to give $X^{P}_{(1)}$	2nd iteration (r=2): Movement of semi-landmarks to give $X^{new}_{(2)}$	GPA registration to new mean $\hat{\mu}_{(2)}$ to give $X^{P}_{(2)}$
(a) BE					
(b) FP					
(c) NP					

Fig 10.11: First one or two iterations of semi-landmark methods for a single circle configuration generated with w=0, σ=0.04, using (a) Bending energy (BE), (b) Full Procrustes (FP), (c) Nearest point (NP) criteria. Key for all panels: (Plusses) current mean $\hat{\mu}_{(r)}$, (filled circles) $X^{P}_{(r)}$ (odd cols) and $X^{new}_{(r)}$ (even cols), (empty circles) nominal semi-landmark positions $X^{P}_{(r-1)}$. First row of (a-c): As described at top of page. Additional rows: (a) (2nd row) Deformation grid for PTPS mapping from $\hat{\mu}_{(r)}$ to $X^{P}_{(r-1)}$ (odd cols) or $X^{new}_{(r)}$ (even cols). (a) (3rd row) Deformation grid for affine component of PTPS mapping in panel above, (diag. crosses=fitted affine coordinates).

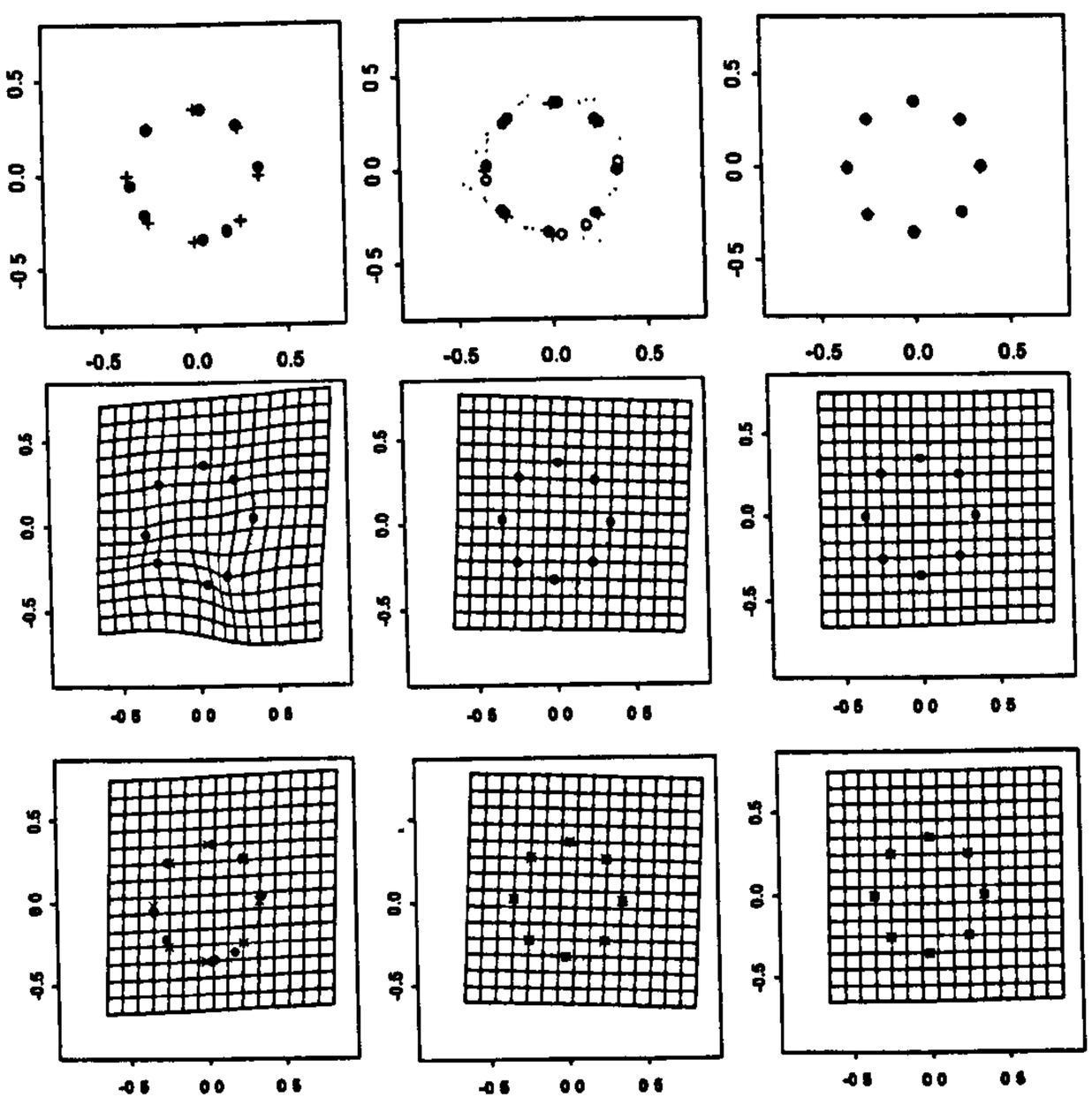
First row of parts (a-c):

Initial GPA registration to $\hat{\mu}_{(0)}$ to give $X^P_{(0)}$

1st iteration (r=1): Movement of semi-landmarks to give $X^{new}_{(1)}$

GPA registration to new mean $\hat{\mu}_{(1)}$ to give $X^P_{(1)}$

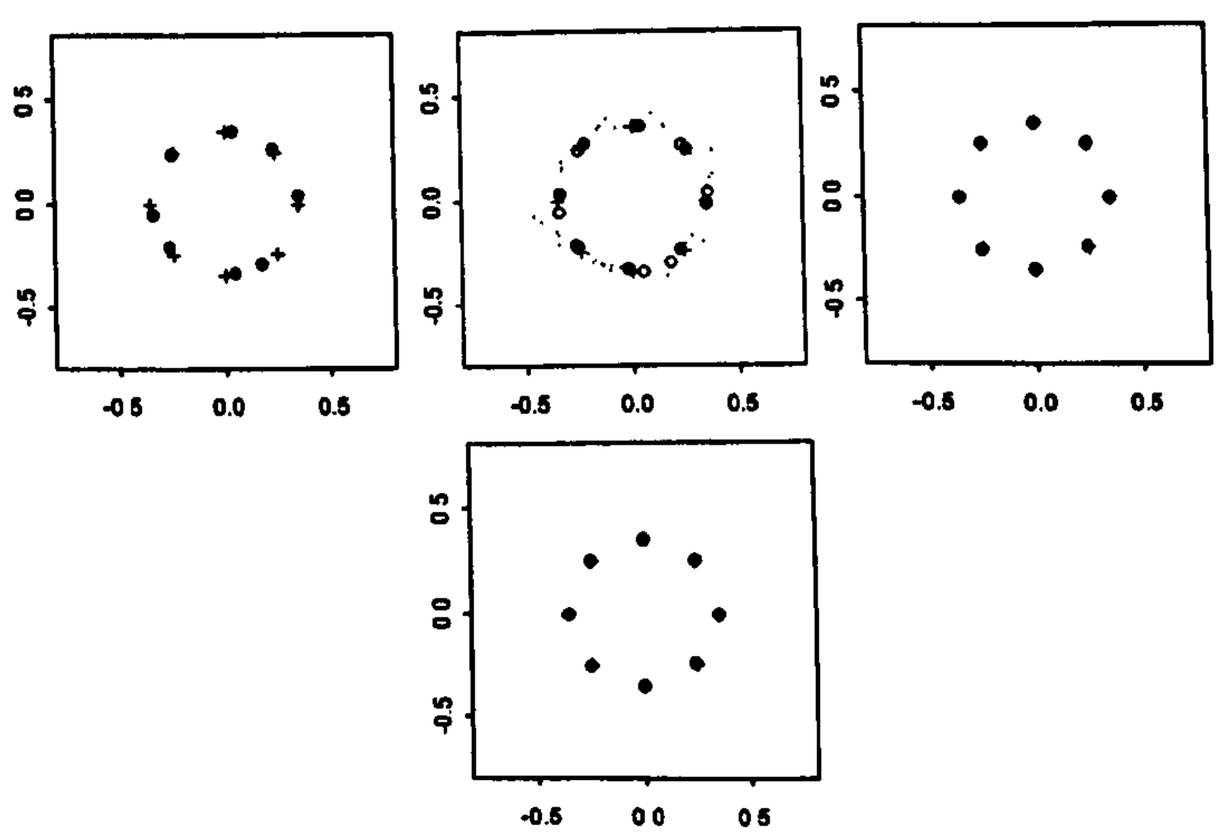
(a) BE



Deformation grid for PTPS mapping from $\hat{\mu}_{(r)}$ to $X^P_{(r)}$ (odd cols) and $X^{new}_{(1)}$ (even cols).

Deformation grid for affine component of PTPS mapping in panel above, (diag. crosses=fitted affine coordinates).

(b) FP



Minimisation of FP criterion shown as superimposition of $X^{new}_{(1)}$ to $\hat{\mu}_{(0)}$.

(c) NP

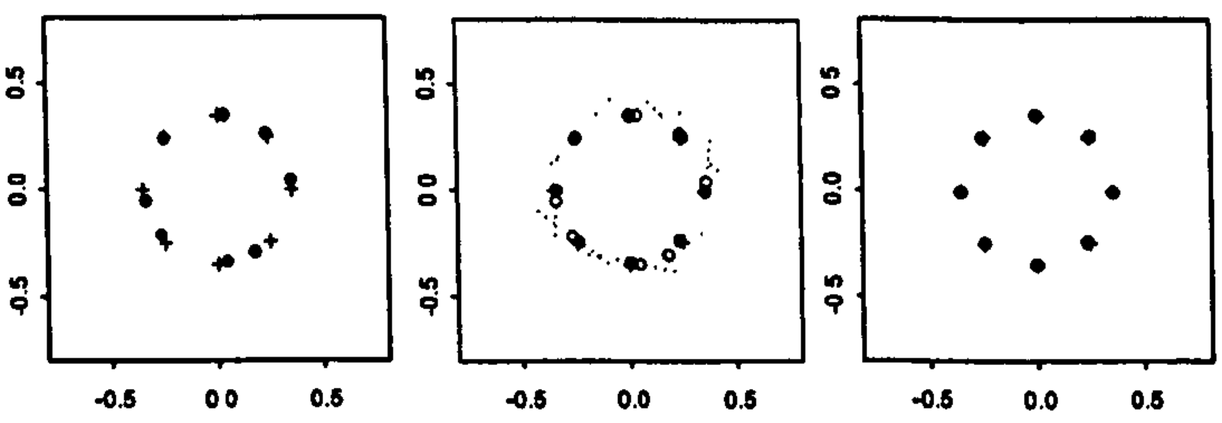


Fig 10.12: First iteration of semi-landmark methods for a single circle configuration generated with w=15, σ=0, using (a) Bending energy (BE), (b) Full Procrustes (FP), (c) Nearest point (NP) criteria. Key for all panels: (Plusses) current mean $\hat{\mu}_{(r)}$, (filled circles) $X^P_{(r)}$, (odd cols) and $X^{new}_{(1)}$ (even cols), (empty circles) nominal semi-landmark positions $X^P_{(0)}$.

When the configuration is registered to a new (although practically unchanged) mean at the end of the first iteration, the fit in the rotational directions of the chords has essentially been optimised already, but variation will remain perpendicular to the circle outline from the previous movement step, as the new GPA scaling is are only able to remove some of this (as noted above). From the second movement step onwards however, the semi-landmarks then hardly move. Despite any changes in the optimal scale of the superimposition, for the next movement step the semi-landmarks are almost already at their " nearest points" to those in $\hat{\mu}_{(1)}$ because of the rotational fit achieved by moving to points at equally spaced angles around the outline, on the previous iteration. Therefore, unlike in other applications, here it is difficult to remove variation normal to the chord directions by iterating between the movement and GPA steps. Usually the choice of semi-landmarks and chord directions involved and/or presence of fixed landmarks, results in the GPA fit, following each movement step always being improved, as the reduction in variance achieved at the by moving the semi-landmarks is distributed around the rest of the configuration. Here however, most of the variance perpendicular to the circle outline from the end of the first iteration, remains.

For the BE method no large changes in shape occur because the chord directions are non-parallel, lying roughly around the outline of a circle, which as explained in 9.3.3.2, restricts the possibility of any large changes in the mapping of $\hat{\mu}_{(r-1)}$ to $vec^{-1}(vec(X^{P}_{(r-1)})-U_{(r)}\lambda_{(r)})$ being produced (other than in terms of size and/or a rotation). Even though the choice of semi-landmarks and chord directions constrains the affine component of the PTPS transformation to be nearly a similarity transformation, variation still remains in the directions of the semi-landmarks' chords, rather than entirely perpendicular to them, as with the FP method. As described in 9.3.3.2, the new semi-landmark positions are determined by a sum of generalised, rather than ordinary, least squares between the fitted values of the affine/similarity transformation of $\hat{\mu}_{(r-1)}$ and $vec^{-1}(vec(X^{P}_{(r-1)})-U_{(r)}\lambda_{(r)})$. Therefore, in contrast to the FP method, the semi-landmarks do not move to positions which minimise Euclidean distances between the chords in $vec^{-1}(vec(X^{P}_{(r-1)})-U_{(r)}\lambda_{(r)})$ and the corresponding landmarks in $SG(\hat{\mu}_{(r-1)})$ (or equivalently between $SG(vec^{-1}(vec(X^{P}_{(r-1)})-U_{(r)}\lambda_{(r)}))$ and $\hat{\mu}_{(r-1)}$), leaving some variation in their positions along the chords, as well as normal to

them. Figs 10.11(a) and 10.12(a) illustrate how on the first iteration an affine transformation of $\hat{\mu}_{(r-1)}$ comprising a rotation (and increase in scale on the second iteration in fig. 10.11(a)), enables the BE or GLS residuals between the affine mapping and $vec^{-1}(vec(X^P_{(r-1)}) - U_{(r)}\lambda_{(r)})$ to be minimised. As with the FP method, slightly larger values of *RSS* (and RSS_B), result with increasing σ and *w*, as the approximation by the chords, to the outline of a circle becomes worse (see section 10.3.1.3 below).

10.3.1.3 Differences with σ and *w*

With each method, the remaining variation in shape in fig 10.9 clearly appears to be greater when the standard deviation of simulated noise σ is larger (at each level of *w*). Although less apparent, a larger scatter of variation in the Procrustes fits also appears to result when the range (-*w*,*w*) of angular displacements sampled from a Uniform distribution is larger (at each level of σ). Both these observations are confirmed by examining the values of *RSS* and RSS_B in table 10.1.

Since the chosen chord directions (are intended to) approximate the directions along which the angular variation was simulated, each of the methods is more able to remove this variation than noise (which was simulated isotropically and so occurs perpendicular to the outline, as well as along it). As the range of angular variation (-*w*,+*w*) is increased, this has only a small effect on the different methods' ability to remove this variation as it is still roughly along the chord directions. Increasing *w* does result in a poorer approximation to the circumference of a circle by the chord directions, which explains the slightly larger *RSS* (remaining variation) with larger *w*, in fig. 10.9. As σ is increased, we then have the problem described in 10.3.1.2 above, where any variation perpendicular to the chord directions is more difficult to remove, particularly for the NP method, although the FP and (to a lesser extent) BE methods are able to have some success, since we allow the superimposition of $\hat{\mu}$ to $vec^{-1}(vec(X^P) - U\lambda)$ change as we minimise the OLS (or GLS) distances between landmarks.

Differences between how the methods perform at different levels of *w* and σ can be seen in fig 10.10, which plots values of RSS_B (top two rows) and the percentage change

in RSS_B from the initial value obtained with all landmarks fixed (bottom two rows). Aside from the effects of increasing w and σ already mentioned above, it can be seen that as σ is increased, the difference between the values of RSS_B obtained by the three different methods is increased, regardless of the level of w (fig 10.10 top row). In other words, with increased noise the BE method is affected most, then the NP and FP methods. Reasons why the NP method is affected more than the FP with increasing noise were explained above. The difference between the remaining variance of the BE and other two methods increases as σ is increased is because more variation is simulated in directions which cannot be removed by the PTPS mapping of $\hat{\mu}$ to $vec^{-1}(vec(X^P)-U\lambda)$. The FP and NP method are able to remove the increased variation along the directions of the chords, whereas the BE method always leaves some proportion of the simulated variation along the chord directions, due to its use of a GLS criterion, as described in 10.3.1.1. As the noise is increased in all directions, the amount of variance not removed by the BE method along the chords will also increase. A similar effect on the value of RSS_B obtained by each method, also occurs as w is increased at each level of σ although this is less noticeable since less variation in the directions of the chords is produced, compared to when σ is increased.

The percentage reduction in RSS (or RSS_B) in table 10.1 allows us to account for the fact that the initial values of RSS (or RSS_B) will be higher with larger w and σ. In this respect, the FP criterion once again clearly performs best. At all levels of w and σ, the percentage reduction exceeds 98.5%, coping equally well with both increased noise and angular variation. Of note is how well the FP criterion performs when there is only noise added to the base positions compared to the NP and BE criteria (see table 10.1 or fig 10.9 for w=0). For w>0, the patterns in the percentage reduction in RSS_B are quite similar for the BE and NP methods, decreasing (performing worse) as σ is increased as we would expect from the discussion above (see fig 10.10 3rd row), but increasing (performing better) as w is increased (see fig 10.10 bottom row). The latter observation is explained by the fact that angular variation is much more easily removed by these methods than noise. After simulating a wider range of angular variation, producing larger initial values of RSS_B or RSS, the percentage reduction in RSS_B or RSS is greater since the additional angular variation is better approximated by the semi-

landmarks chords, along which variation is removed. Conversely, increasing σ increases the variation at each landmark in every direction, which is more difficult to remove (with the FP method having most success).

10.3.1.4 Changes in estimated mean shapes

As well as a reduction in variance between configurations, we also hope that the mean of the final fits ($\hat{\mu}_{SL}$) is the same shape as the base configuration from which the datasets were generated. Table 10.3 (left) details the full Procrustes distances between $\hat{\mu}_{SL}$ and B and between the initial estimated mean $\hat{\mu}_0$ and B. As we would expect, values of $d_F(\hat{\mu}_0, B)$, indicated by shaded boxes, increase slightly with both w and σ as the effects of the larger variation in shape influence the estimate of $\hat{\mu}$. Therefore in brackets we also quote the change in $d_F(\hat{\mu}_{SL}, B)$ from $d_F(\hat{\mu}_0, B)$. Table 10.3 (right) also details the change in mean shape from the initial estimated mean, i.e. $d_F(\hat{\mu}_{SL}, \hat{\mu}_0)$.

The smallest values of $d_F(\hat{\mu}_{SL}, B)$ at each combination of w and σ were always found for the NP method, where there is a slight increase in the dissimilarity of $\hat{\mu}_{SL}$ from B as σ is increased and a more notable increase for larger w. However, each of these values represent no change from the initial dissimilarity measure $d_F(\hat{\mu}_0, B)$, obtained for each combination of w and σ, since $\hat{\mu}_{SL}$ is always practically unchanged from the initial estimated mean shape ($d_F(\hat{\mu}_{SL}, \hat{\mu}_0) < 0.001$ always). The largest values of $d_F(\hat{\mu}_{SL}, B)$ occur when using the BE criterion, with table 10.3 suggesting that the dissimilarity between $\hat{\mu}_{SL}$ and B generally increases with σ, but hardly changes with w. In addition, the change from $d_F(\hat{\mu}_0, B)$ also tends to be greater for larger σ, reflecting the fact that the change in estimated mean, $d_F(\hat{\mu}_{SL}, \hat{\mu}_0)$ also increases with σ. For the FP method $d_F(\hat{\mu}_{SL}, B)$ increases with both w and σ and as with the BE method, the change from $d_F(\hat{\mu}_0, B)$ increases slightly with σ, since $d_F(\hat{\mu}_{SL}, \hat{\mu}_0)$, also increases with σ.

Overall the figures suggest that in contrast to the NP technique, the FP and (more notably) BE methods, produce final estimates of mean shape that change slightly from

their original estimates when there is increased noise. This is also true (although to a lesser extent) when $\hat{\mu}_{SL}$ is compared to base shape B. However, even when σ =0.04, the difference between $\hat{\mu}_{SL}$ and B was never more than 0.007 (FP method) or 0.016 (BE method) greater than the original difference in shape between $\hat{\mu}_0$ and B. Recall from section 10.2.3 that a change in shape of order 0.004 could result from just a one pixel change at each of the landmarks of B (when scaled to have radius=250 pixels). In view of this, the sizes of changes observed are of no real practical concern for both the new semi-landmark methods and (by a slightly weaker argument) original technique.

	Semi-landmark criterion	$d_F(\hat{\mu}_{SL}, B)$ (change from $d_F(\hat{\mu}_0, B)$)			$d_F(\hat{\mu}_{SL}, \hat{\mu}_0)$		
		σ=0.00	σ=0.02	σ=0.04	σ=0.00	σ=0.02	σ=0.04
w=0	None (All fixed)	0.000	0.003	0.004			
	Bending energy	0.000 (0.000)	0.010 (0.007)	0.020 (0.016)	0.000	0.009	0.017
	Full Procrustes	0.000 (0.000)	0.004 (0.001)	0.008 (0.004)	0.000	0.004	0.007
	Nearest point	0.000 (0.000)	0.003 (0.000)	0.004 (0.000)	0.000	0.000	0.000
w =7.5	None (All fixed)	0.005	0.006	0.009			
	Bending energy	0.005 (0.000)	0.009 (0.003)	0.019 (0.010)	0.000	0.005	0.013
	Full Procrustes	0.005 (0.000)	0.008 (0.002)	0.016 (0.007)	0.000	0.009	0.013
	Nearest point	0.005 (0.000)	0.006 (0.000)	0.009 (0.000)	0.000	0.000	0.000
w =15	None (All fixed)	0.011	0.010	0.013			
	Bending energy	0.011 (0.000)	0.012 (0.002)	0.020 (0.007)	0.002	0.005	0.015
	Full Procrustes	0.012 (0.001)	0.012 (0.002)	0.018 (0.005)	0.004	0.007	0.016
	Nearest point	0.011 (0.000)	0.010 (0.000)	0.013 (0.000)	0.001	0.001	0.001

Table 10.3: Full Procrustes distances of final mean shapes to base and initial mean configuration estimates from circle simulations.

10.3.2 Simulations from a square

In contrast to the samples generated from a circle, here only the landmarks along the sides of the configurations are allowed to move along escribed chords. Again, we aim to reduce the variation in shape between the simulations and hope that their mean shape is the same shape as the base configuration from which they were generated.

Fig. 10.13 displays the scatters of final Procrustes registered configurations, $X_i^{P_{SL}}$, i=1,...,100, about their estimated mean shape $\hat{\mu}_{SL}$, after running each of the semi-landmark methods, on each of the eight simulated datasets. Values of the original and final *RSS* values (corresponding to the variation in fig. 10.13) obtained using each technique can be seen in table 10.4(left). In table 10.4 (right), are the values of RSS_B. As in table 10.1, the figures in brackets indicate the percentage reduction in *RSS*

and RSS_B, from their original values in the grey cell above. Fig. 10.14 presents the values of RSS_B and percentage change in RSS_B as interaction plots. For each of the semi-landmark methods, the first and third rows show the effect of changes in the simulated linear displacement range (-w,+w), on separate plots for each level of noise σ, and the second and fourth rows show the effect of changes in σ at different levels of w.

10.3.2.1 Main observations

An examination of the remaining variation in shape after running each of the semi-landmark methods (fig. 10.13) shows that all three methods give very similar results. For the FP and NP methods, any remaining variation in the semi-landmark positions is again perpendicular to the final directions of the chords (where here the chord directions do not change between iterations). However, in contrast to the circle simulations the scatters appear identical with no difference in the extent of variation in these directions between these two methods, as confirmed by the values of RSS and RSS_B in table 10.4. For the BE method, a slight amount of variation remains along these directions again, which we discuss further below. With all methods, a reduction in variance also occurs at the fixed landmarks where the reduction in variance achieved at the semi-landmarks, allows the GPA fits of the configurations to be improved at all other positions. Table 10.5 shows that the BE and FP methods took the same number of iterations for each combination of w and σ, with the NP method always taking an increasing number of iterations as w is increased.

	Semi-landmark criterion	Number of iterations		
		σ=0.000	σ=0.0125	σ=0.0250
w=0	Bending energy		2	2
	Full Procrustes		2	2
	Nearest point		2	3
w=0.15	Bending energy	2	2	2
	Full Procrustes	2	2	2
	Nearest point	4	4	4
w=0.30	Bending energy	2	2	2
	Full Procrustes	2	2	2
	Nearest point	5	5	6

Table 10.5: Number of iterations required for convergence at ΔRSS<0.001.

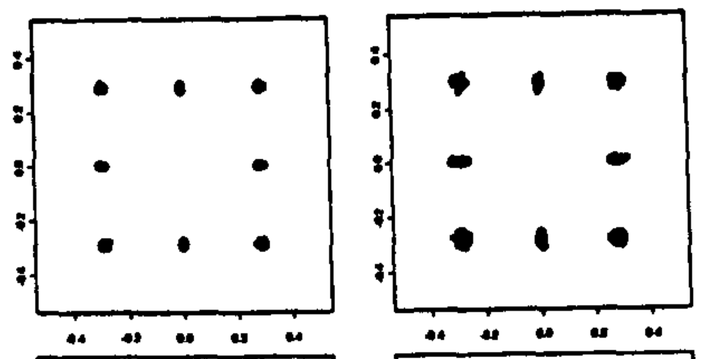
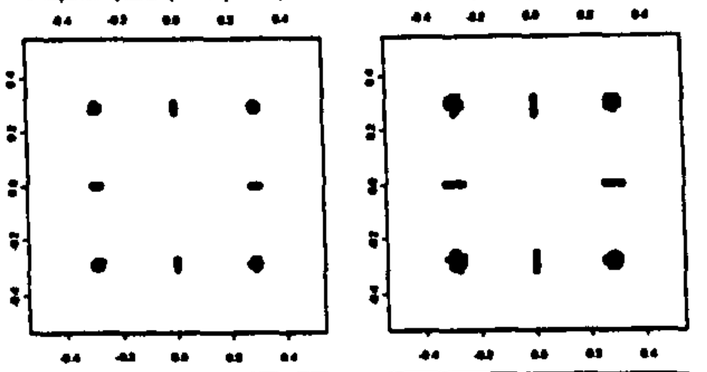
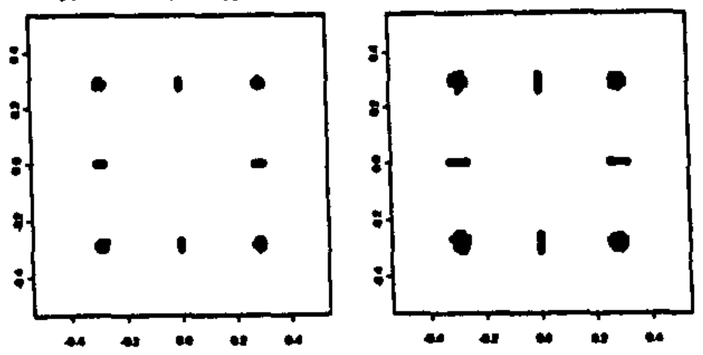
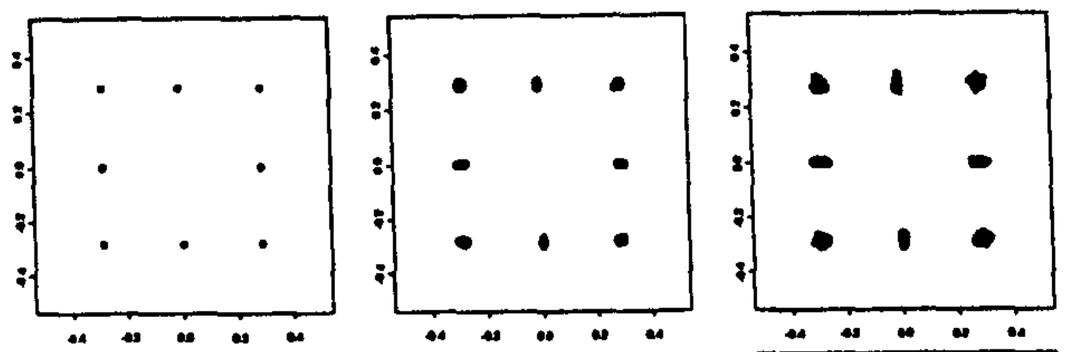
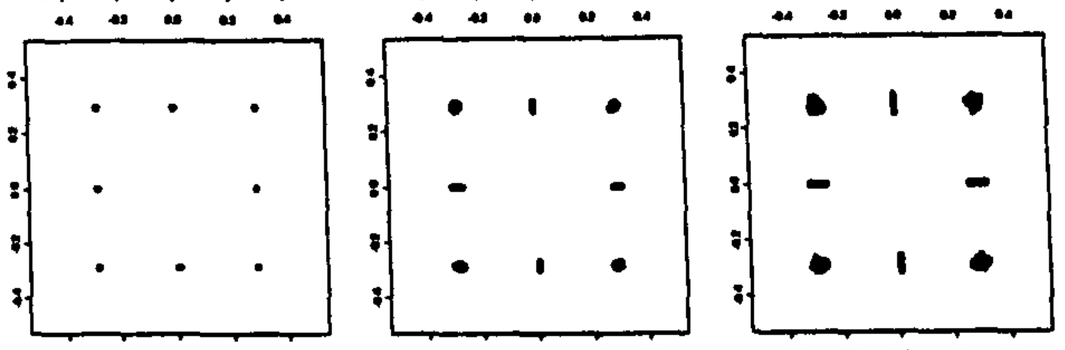
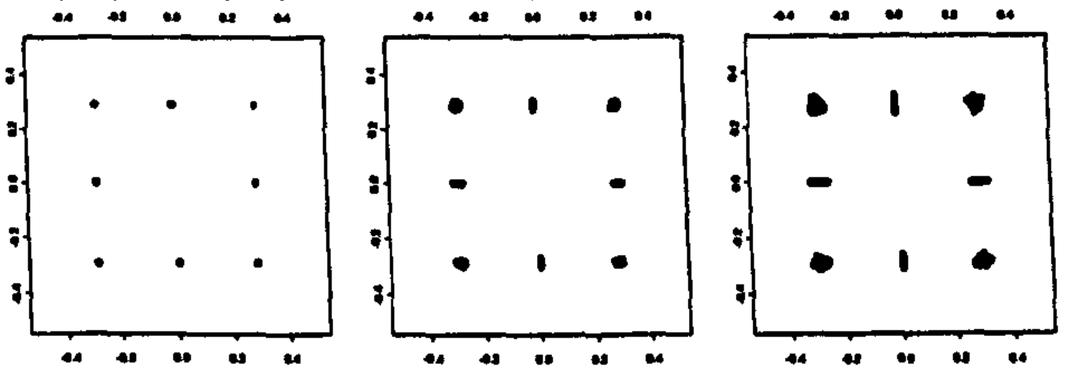
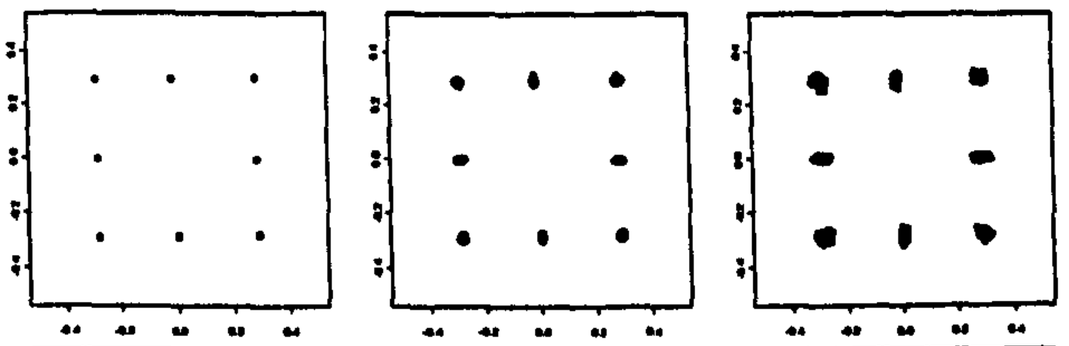
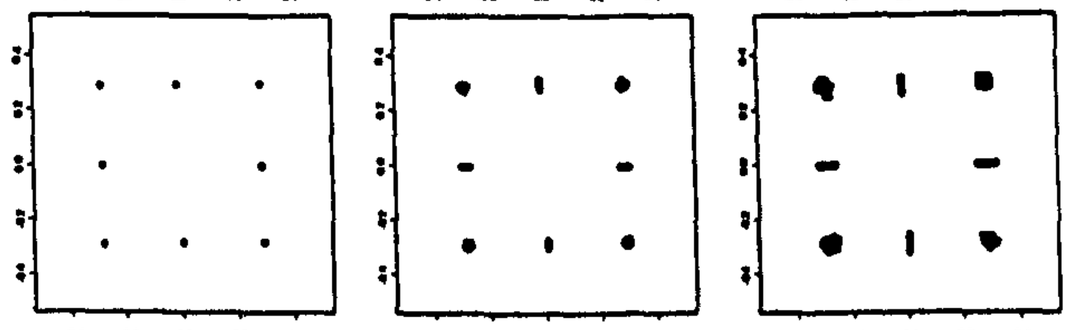
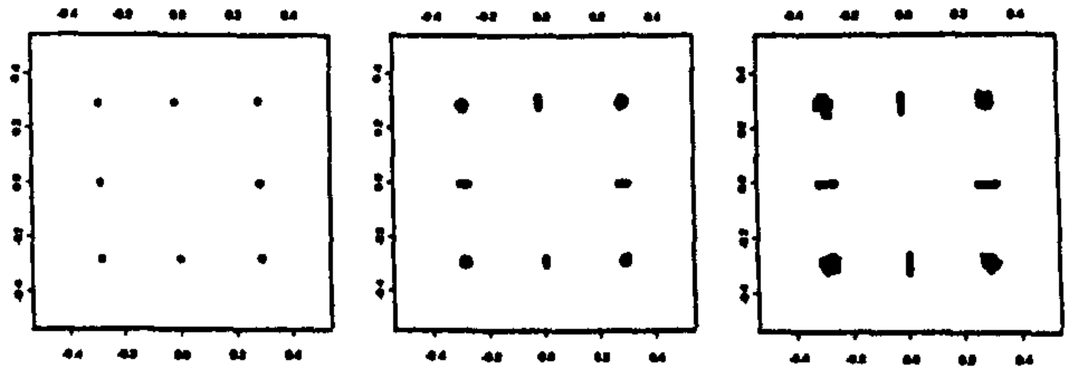
		σ=0.0000	σ=0.0125	σ=0.0250
w=0.0	Bending energy			
	Full Procrustes			
	Nearest point			
w=0.15	Bending energy			
	Full Procrustes			
	Nearest point			
w=0.3	Bending energy			
	Full Procrustes			
	Nearest point			

Fig 10.13: Final Procrustes fits after running each of the semi-landmark methods on each set of square configurations, simulated with various combinations linear variation *w* and noise σ.

	Semi-landmark criterion	RSS (%Δ RSS from original in grey)			RSS_B (%Δ RSS_B from original in grey)		
		σ=0.0000	σ=0.0125	σ=0.0250	σ=0.0000	σ=0.0125	σ=0.0250
w=0	None (All fixed)		0.061	0.259		0.062	0.261
	Bending energy		0.044 (27.7%)	0.180 (30.6%)		0.045 (27.4%)	0.181 (30.7%)
	Full Procrustes		0.041 (33.1%)	0.168 (35.3%)		0.041 (33.9%)	0.169 (35.2%)
	Nearest point		0.041 (32.9%)	0.168 (35.3%)		0.042 (32.3%)	0.169 (35.2%)
w=0.15	None (All fixed)	0.741	0.781	0.971	0.749	0.788	0.978
	Bending energy	0.000 (100%)	0.046 (94.1%)	0.185 (80.9%)	0.008 (98.9%)	0.054 (93.1%)	0.192 (80.4%)
	Full Procrustes	0.000 (100%)	0.044 (94.4%)	0.175 (82.0%)	0.008 (98.9%)	0.051 (93.5%)	0.182 (81.4%)
	Nearest point	0.000 (100%)	0.044 (94.4%)	0.175 (82.0%)	0.008 (98.9%)	0.051 (93.5%)	0.182 (81.4%)
w =0.3	None (All fixed)	2.852	2.892	3.052	2.878	2.917	3.076
	Bending energy	0.000 (100%)	0.044 (98.5%)	0.198 (93.5%)	0.026 (99.1%)	0.069 (97.6%)	0.225 (92.7%)
	Full Procrustes	0.000 (100%)	0.042 (98.6%)	0.182 (94.0%)	0.025 (99.1%)	0.066 (97.7%)	0.208 (93.2%)
	Nearest point	0.000 (100%)	0.042 (98.6%)	0.182 (94.0%)	0.025 (99.1%)	0.066 (97.7%)	0.208 (93.2%)

Table 10.4: Final values of RSS and RSS_B for configurations simulated from a square using various *w* and σ.

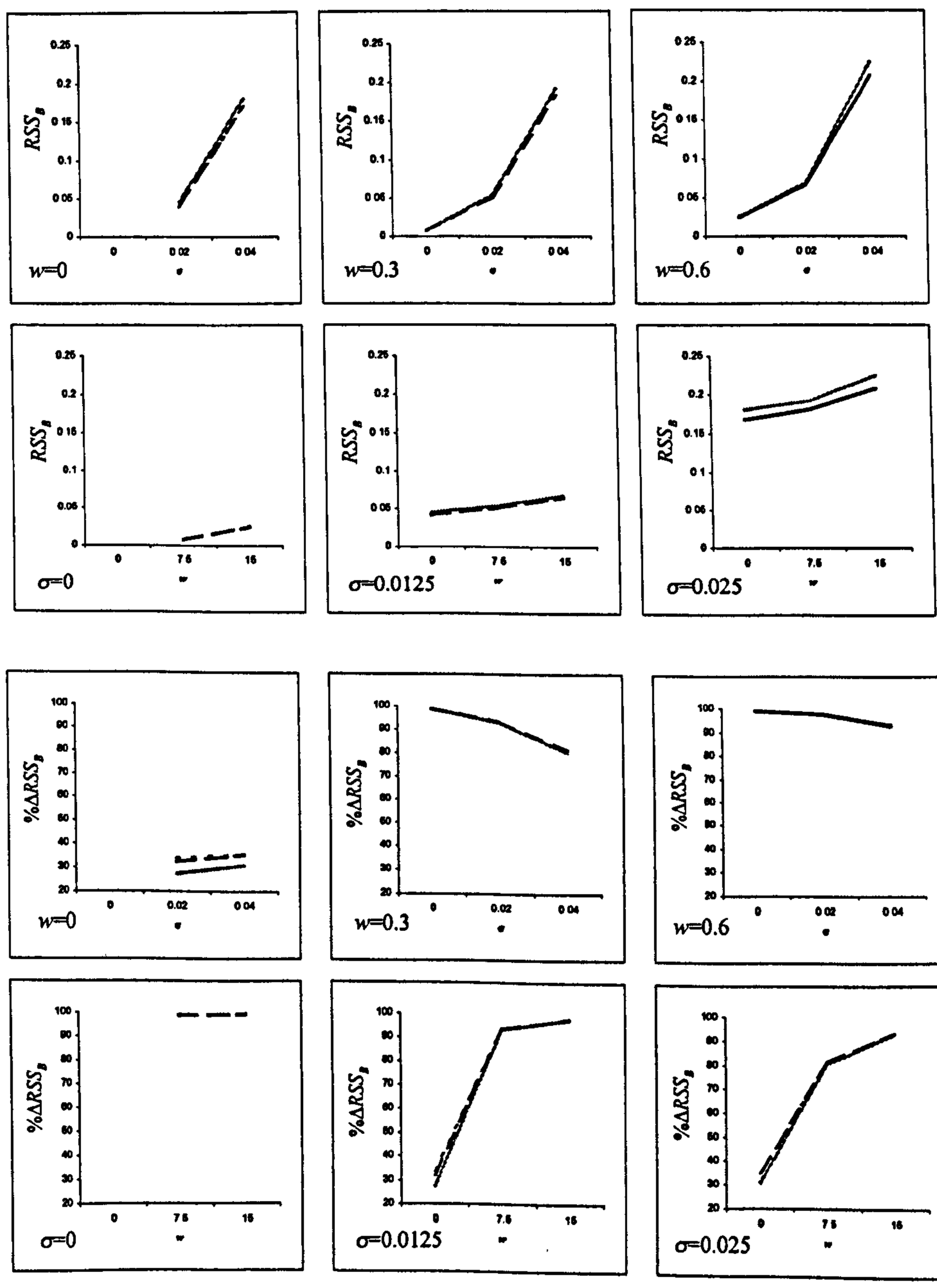


Fig 10.14 (top two rows): Final values of RSS_B for configurations simulated from a square using various *w* and σ. (Bottom two rows): Percentage change in RSS_B from original value (all landmarks fixed).

In table 10.4, values of RSS and RSS_B are always of the same magnitude, and follow similar patterns with w and σ for each method. Differences between RSS and RSS_B do however, increase slightly with w for each method (but not σ), suggesting an increasing difference in shape between $\hat{\mu}_{SL}$ and base configuration B (with w) which we examine in section 10.3.2.4. For each combination of w and σ, the BE criterion always performed slightly worse (in terms of RSS or RSS_B) than the FP and NP methods, confirming the observations in fig 10.13. Notice also that the difference in values between the BE and other methods also increases slightly with σ.

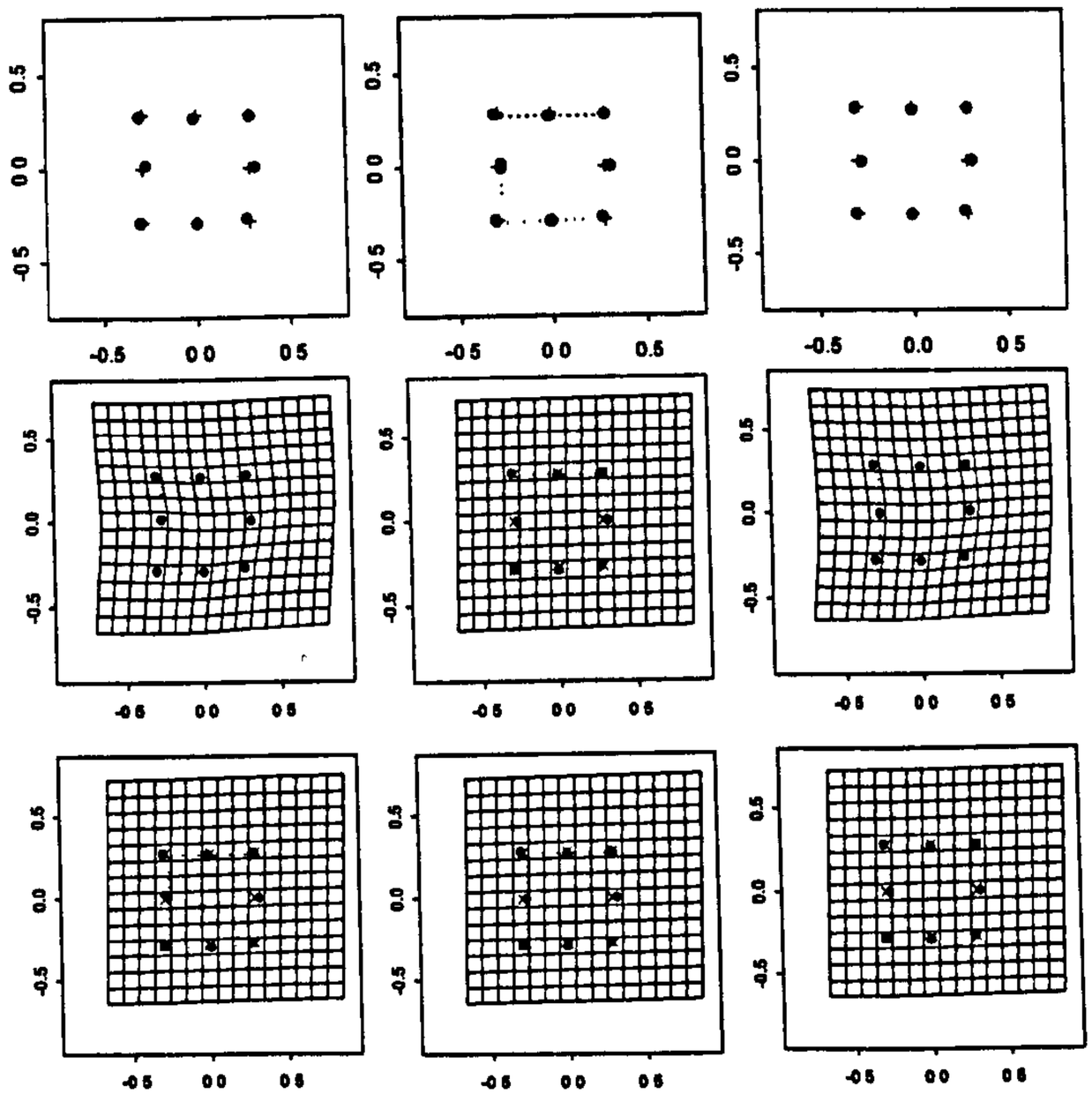
10.3.2.2 Differences between semi-landmark methods

Fig. 10.15 shows the results of the first iteration, using each of the BE, FP and NP methods, for one of configurations simulated with only maximum noise (σ=0.025, w=0) at each of its landmark positions. Fig. 10.16 shows the first one or two iterations of each of the methods for one of the configurations simulated with maximum variation along its sides, but no noise (w=0.3, σ=0).

As before, for configurations simulated with isotropic noise at each of the landmark positions (σ=0.025, w=0), GPA registration to the initial estimate of mean shape $\hat{\mu}_0$ (obtained with all landmarks treated as fixed), results in isotropic variation at the residuals of each landmark as seen in fig 10.7(right) top right. For larger σ, this again leads to greater variation in how well the first set of chords approximate the sides of the square along which the four semi-landmarks may move, as for example in the 2nd panels of parts (a), (b) and (c) in fig. 10.15. Here however, on subsequent iterations the chord directions do not change. For configurations simulated with only variation along the sides of the base shape, there is no variation in the chord directions as the four corner landmarks (which determine the directions of the semi-landmarks chords), are always in identical positions on each configuration. For w=0.3, σ=0, GPA registration to $\hat{\mu}_0$ results in isotropic residuals at each of the fixed corner landmarks, as can be seen in the 2nd panels of parts (a), (b) and (c) of fig. 10.16, but greater variation in those between them, along the sides of the square (as seen in fig.10.7(right) bottom left).

First row of parts (a-c):	Initial GPA registration to $\hat{\mu}_{(0)}$ to give $X^P_{(0)}$	1st iteration (r=1): Movement of semi-landmarks to give $X^{new}_{(1)}$	GPA registration to new mean $\hat{\mu}_{(1)}$ to give $X^P_{(1)}$

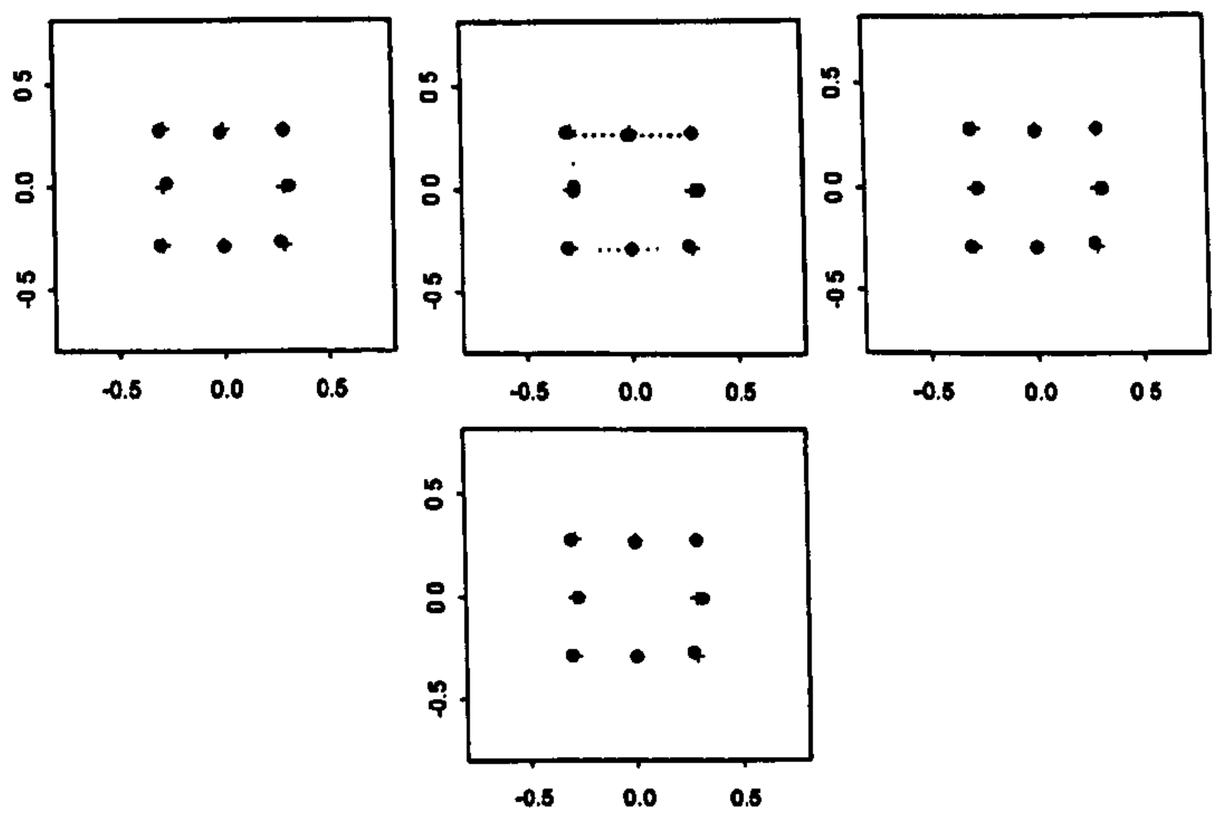
(a) BE



Deformation grid for PTPS mapping from $\hat{\mu}_{(r)}$ to $X^P_{(r)}$ (odd cols) and $X^{new}_{(1)}$ (even cols).

Deformation grid for affine component of PTPS mapping in panel above, (diag. crosses=fitted affine coordinates).

(b) FP



Minimisation of FP criterion shown as superimposition of $X^{new}_{(1)}$ to $\hat{\mu}_{(0)}$.

(c) NP

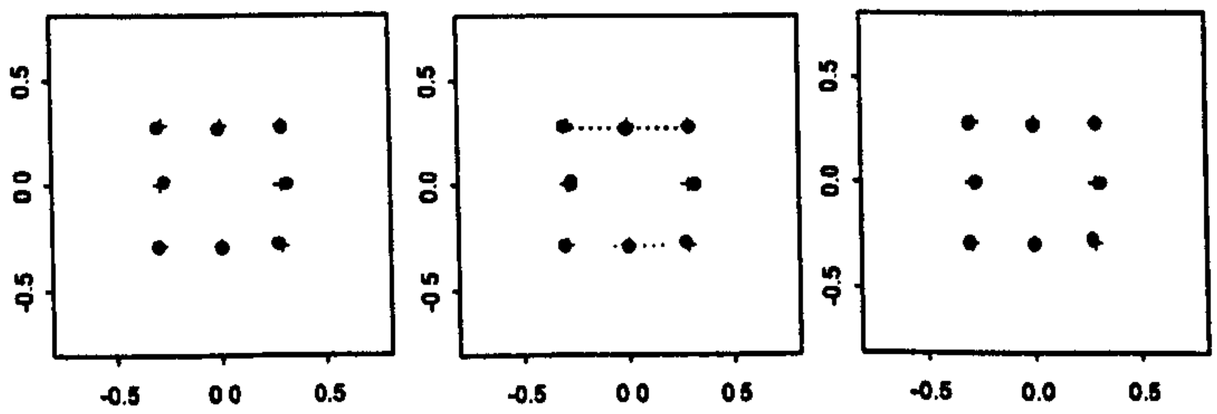


Fig 10.15: First iteration of semi-landmark methods for a single square configuration generated with w=0, σ=0.025, using (a) Bending energy (BE), (b) Full Procrustes (FP), (c) Nearest point (NP) criteria.
Key for all panels: (Plusses) current mean $\hat{\mu}_{(r)}$, (filled circles) $X^P_{(r)}$, (odd cols) and $X^{new}_{(1)}$ (even cols), (empty circles) nominal semi-landmark positions $X^P_{(0)}$.

First row of parts (a-c):

Initial GPA registration of config. to $\hat{\mu}_{(0)}$ to give $X^{P}_{(0)}$

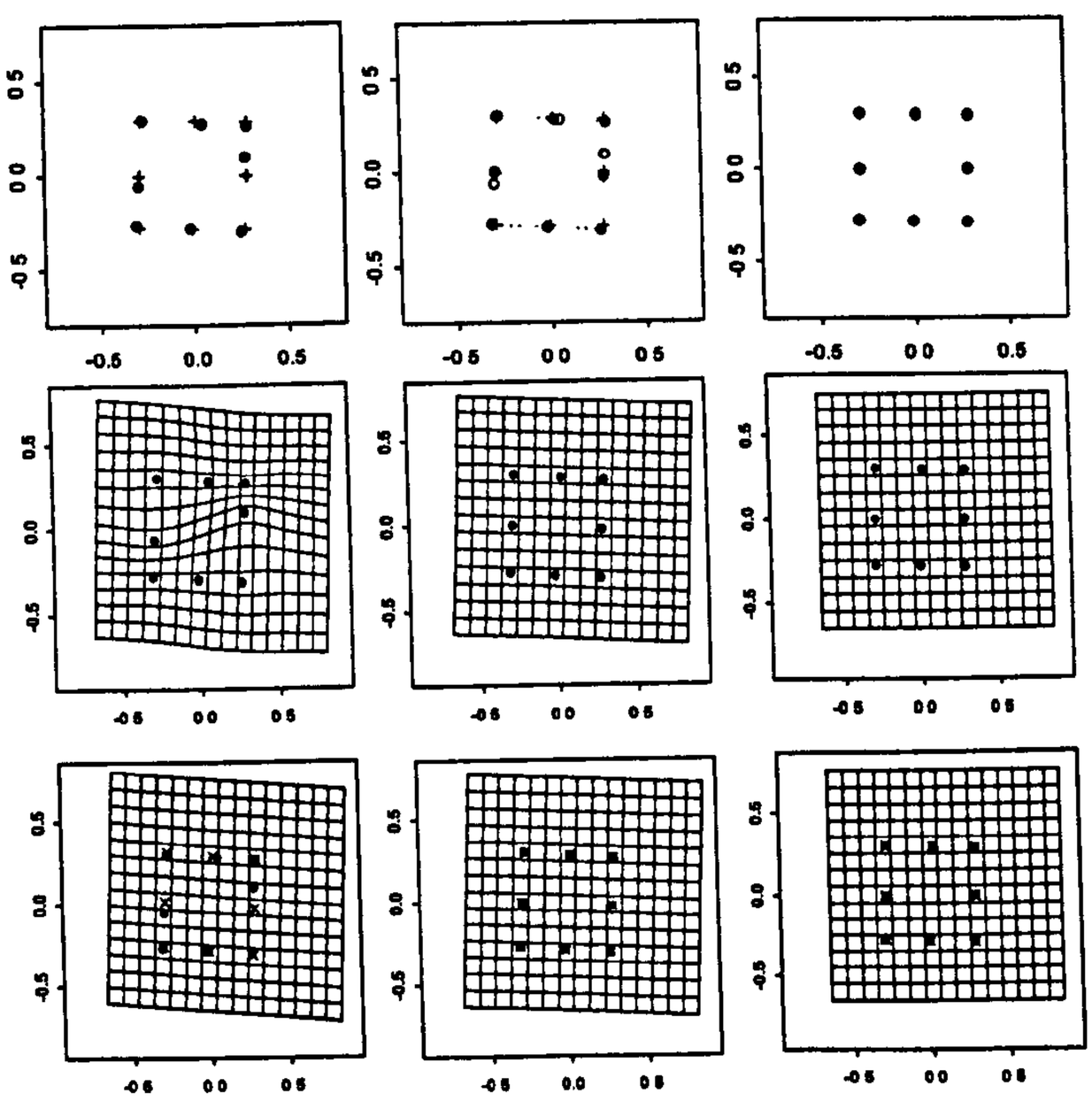
1st iteration (r=1) Movement of semi-landmarks to give $X^{new}_{(1)}$

GPA registration to new mean $\hat{\mu}_{(1)}$ to give $X^{P}_{(1)}$

2nd iteration (r=2): Movement of semi-landmarks to give $X^{new}_{(2)}$

GPA registration to new mean $\hat{\mu}_{(2)}$ to give $X^{P}_{(2)}$

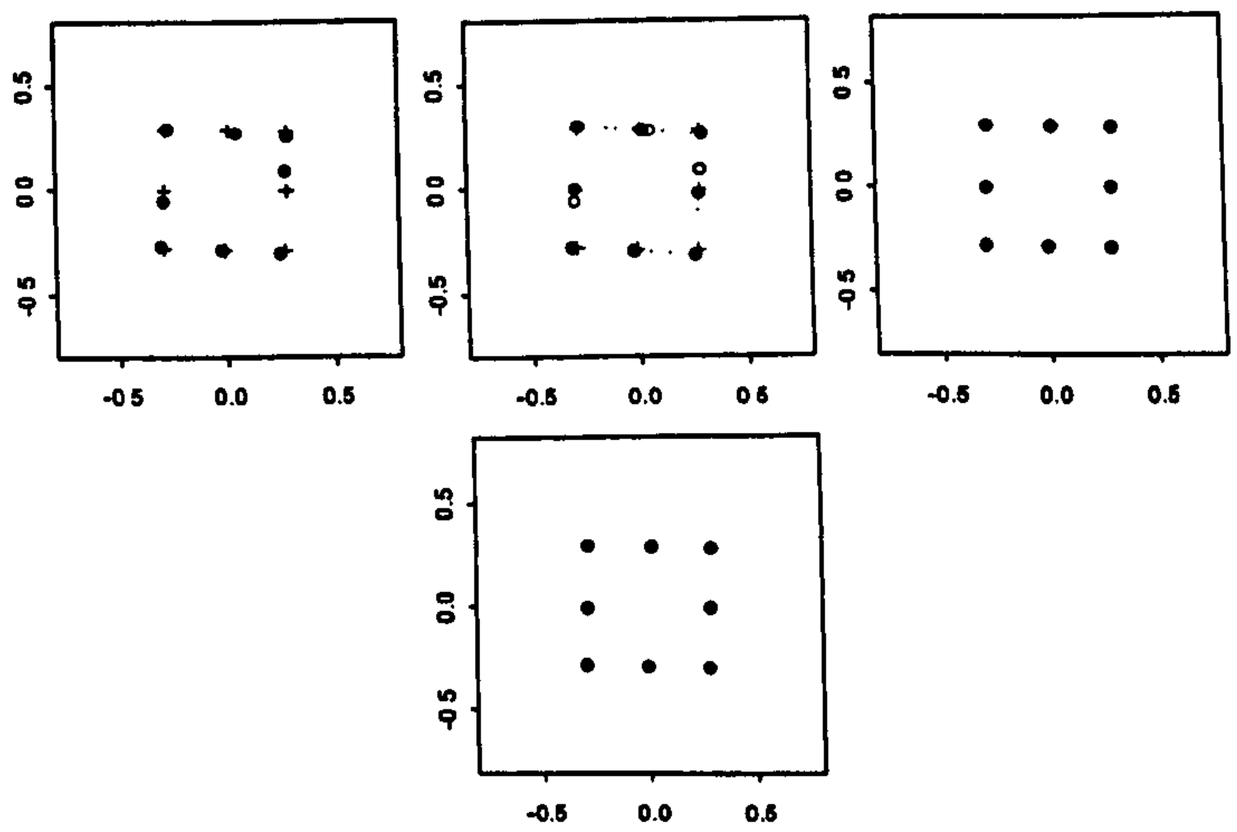
(a) BE



Deformation grid for PTPS mapping from $\hat{\mu}_{(r)}$ to $X^{P}_{(r)}$ (odd cols) and $X^{new}_{(1)}$ (even cols).

Deformation grid for affine component of PTPS mapping in panel above, (diag. crosses=fitted affine coordinates).

(b) FP



Minimisation of FP criterion shown as superimposition of $X^{new}_{(1)}$ to $\hat{\mu}_{(0)}$.

(c) NP

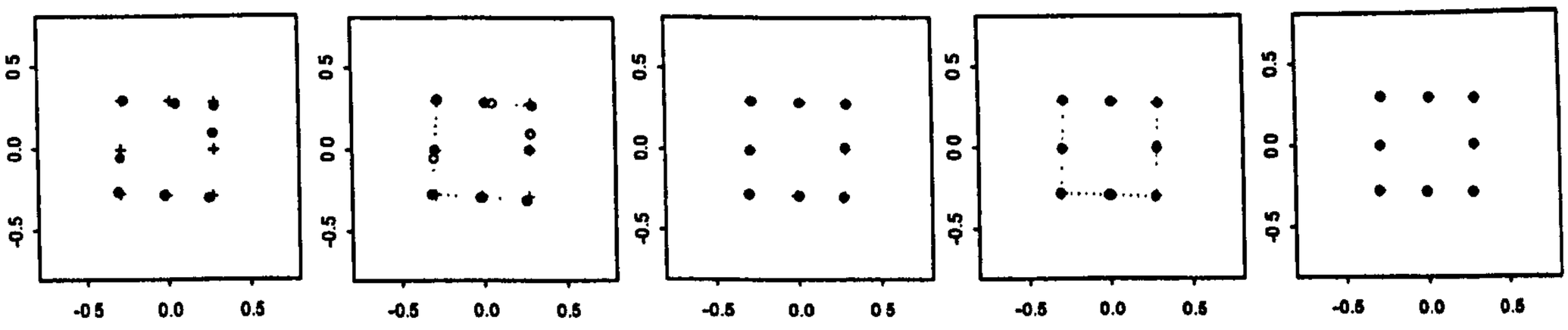


Fig 10.16: First one or two iterations of semi-landmark methods for a single square configuration generated with w=0.3, σ=0, using (a) Bending energy (BE), (b) Full Procrustes (FP), (c) Nearest point (NP) criteria. Key for all panels: (Plusses) current mean $\hat{\mu}_{(r)}$, (filled circles) $X^{P}_{(r)}$ (odd cols) and $X^{new}_{(r)}$ (even cols), (empty circles) nominal semi-landmark positions $X^{P}_{(r-1)}$.

The similarity of results on the first iteration, when using different criteria can be clearly seen in figs. 10.15 and 10.16. Recall from sections 9.3.3.2.1 and 10.3.1.2 how on any iteration, the BE method produces more similar results to those obtained using the FP method when the affine component of the PTPS transformation is close to a similarity transformation. As can be seen in fig. 10.15(a), even with large σ, due to the mapping required to the fixed corner landmarks, the affine transformation of $\hat{\mu}_{(0)}$ to $vec^{-1}(vec(X^P_{(0)}) - U_{(1)}\lambda_{(1)})$ is almost an identity, with the semi-landmarks moving along their chords to minimise the sum of squared non-affine components between the fitted affine values and $vec^{-1}(vec(X^P_{(0)}) - U_{(1)}\lambda_{(1)})$. The slight shearing is required because of the simulated noise at the landmarks of the target configuration and its effect on the semi-landmarks' chord directions (see fig 10.15(a), third row). Consequently the new semi-landmark positions are almost the same as those determined by the FP criterion, allowing only similarity transformations of $\hat{\mu}_{(0)}$ (or $vec^{-1}(vec(X^P_{(0)}) - U_{(1)}\lambda_{(1)})$, with the slight difference in new semi-landmark positions being due to an ordinary, rather than generalised least squares criterion being optimised (see 2nd panels in the first rows of fig. 10.15(a) and (b)). However, when there is no added noise and only variation in the four landmarks along the sides of the square, the BE and FP methods give identical results, as can be seen in fig 10.16 (parts a and b). For the BE criterion, the optimal affine component in the PTPS mapping is the full Procrustes superimposition of $\hat{\mu}_{(0)}$ to $vec^{-1}(vec(X^P_{(0)}) - U_{(1)}\lambda_{(1)})$. Since the configuration of the four fixed corner landmarks is always identical in $\hat{\mu}_{(0)}$ and $vec^{-1}(vec(X^P_{(0)}) - U_{(1)}\lambda_{(1)})$ and any difference in shape occurs only along lines joining them, along which the semi-landmark may move, both the affine mapping and FP superimposition is able to map the landmarks of $\hat{\mu}_{(0)}$ onto the corner landmarks and semi-landmarks chords exactly. No non-affine (bending) components are required since the optimal positions for the semi-landmarks to adopt along their chords are at positions corresponding precisely to where the affine (similarity) transformation of these landmarks in $\hat{\mu}_{(0)}$ maps them to; half way between each pair of corner landmarks, along each chord.

For the same two configurations, figs. 10.15(c) and 10.16(c) show the corresponding results using the NP method. For w=0.3, σ =0 and w=0, σ =0.025, the end results are practically identical to the FP method, although more iterations are required with larger w (see table 10.5). In contrast to the circle simulations, here it is possible to remove the

same variation as the FP method by iterating between the movement and GPA steps. Recall that for the circle configurations, the choice of semi-landmarks and chord directions meant that following GPA at the end of the first iteration, the semi-landmarks were then already at their 'nearest points' relative to $\hat{\mu}_{(1)}$ for the second movement stage and so hardly move again, leaving any variation in directions normal to the outline unable to be removed on the first iteration. Here however, when variation at the semi-landmarks is removed, the presence of fixed landmarks means that the improved GPA fit of each configuration, at all landmarks, allows the updated chord directions to pass closer to the corresponding landmarks in $\hat{\mu}_{(r)}$. This means that over the course of the iterations, the procedure is able to be more successful in removing variation perpendicular to, as well as along the chords, as with the FP method.

10.3.2.3 Differences with σ and w

Fig 10.13 shows that when no noise is added (σ=0) to the corner or side landmarks, each method is successful in removing all variation in shape along the sides of each square, regardless of the value of w used, since the chords are defined precisely in the directions along which the variation was simulated. This is confirmed by the first column of *RSS* values in table 10.4, which are all zero. As σ is increased however, the final remaining variation in shape is clearly larger at each level of w for all of the methods. Variation in the corner landmarks defining the chord directions means that the chords are no longer in consistent directions on the configurations. In addition, noise at the simulated positions of the side landmarks also means that the relative positions of the chords are different and, for the NP method, following GPA registration, the chords will not always pass through identical positions in the vicinity of these landmarks in the mean shape $\hat{\mu}_{(r-1)}$. As a result, the optimal OLS or GLS criteria between positions on these chords and the corresponding landmarks in $\hat{\mu}_{(r-1)}$, $SG(\hat{\mu}_{(r-1)})$ or the affine fit of $\hat{\mu}_{(r-1)}$ to $vec^{-1}(vec(X^P_{(r-1)}) - U_{(r)}\lambda_{(r)})$, will also be affected and thereby the ability of the different methods to reduce the variation in shape in directions perpendicular to the chords. Fig. 10.13 and table 10.4 show that once again, at each combination of w and σ, the uniform variation simulated along the outline (here sides) of the configurations is more easily removed than the simulated noise. For any

method and given amount of noise σ, there appears to be little difference between the scatters of final fits at different levels of w. As w is increased, this has only a very small or (for σ=0) no effect, on the ability of the different methods to remove this variation, as it is still in the directions in which the semi-landmarks are (approximately) allowed to move. Despite this lack of a visual difference however, table 10.4 shows that with larger values of σ, the values of RSS (or RSS_B) differ more as w is increased, as the effects of noise on the initial directions and locations of the chords reduces the ability of each of the methods to remove variation normal to the sides of the squares.

Differences between how the methods perform at different levels of w and σ can also be seen in fig. 10.14 which plots values of RSS_B (top two rows) and the percentage change in RSS_B from the initial value with all landmarks fixed (bottom two rows). The more or less consistent effect of increasing σ at each level of w described above can be seen in the first row. The second row shows the slightly less noticeable effect of increasing w at each level of σ and the resulting increase in difference between the NP/FP methods and BE method. Using the BE criterion, a proportion of variation remains along the semi-landmarks' chords which the NP and FP methods are always able to remove. When the noise at all landmarks is increased in all directions, then, as was found with the circle simulations, the amount of variance not able to be removed by the BE method along the chords will also increase (although here only slightly).

In terms of the percentage reduction in RSS (or RSS_B) in table 10.4, the difference between the three methods is less apparent, except at w=0 where the simulated variation in shape is smaller and so differences in the final values of RSS_B between methods, result in a greater percentage differences in the values of $\%\Delta RSS_B$. However, in general there is no superior performance by one of the methods in terms of these measures, as was found when using the FP criterion in 10.3.1. For w>0, the third row of fig 10.14 shows that for all three techniques the percentage reduction in RSS_B decreases when there is increased noise, although the decrease with σ is smaller with larger w and (in the fourth row) $\%\Delta RSS_B$ increases when w is increased. Again this indicates how the increased variation simulated along the directions represented by the semi-landmarks' chords is removed more easily than noise. For larger w, a greater initial

value of RSS_B (or RSS) is produced and so when variation in the positions of the semi-landmarks along their chord directions is reduced, a greater proportion of the variation originally simulated is removed, resulting in a greater percentage reduction in RSS_B.

10.3.2.4 Changes in estimated mean shapes

Table 10.6(left) details the full Procrustes distances between final estimates of mean shape $\hat{\mu}_{SL}$ and base configuration B from which the samples of configurations were generated. In grey are the full Procrustes distances between each original estimated mean $\hat{\mu}_0$ and B. As before we would hope that the $\hat{\mu}_{SL}$ have the same shape as the base configuration and that there has been no increase in the difference in shape (between $\hat{\mu}_{SL}$ and B), compared to the initial value of $d_F(\hat{\mu}_0, B)$. As in table 10.4, the values of $d_F(\hat{\mu}_0, B)$, indicated by shaded boxes, increase slightly with w, as the effects of larger variation in shape influences the estimate of $\hat{\mu}_0$, but here the increase with σ is less noticeable. Table 10.6(right) details the change in mean shape from the initial estimate, i.e. $d_F(\hat{\mu}_{SL}, \hat{\mu}_0)$.

Note that at each combination of w and σ, the values of $d_F(\hat{\mu}_{SL}, B)$ and $d_F(\hat{\mu}_{SL}, \hat{\mu}_0)$ for each method are identical to 3 decimal places, indicating that the change in shape produced by each technique is always of the same magnitude. Table 10.6(left), shows that $d_F(\hat{\mu}_{SL}, B)$ increases with w and only very slightly with σ (in fact for the largest values of w, differences in $d_F(\hat{\mu}_{SL}, B)$ across σ are zero to 3 decimal places). This also ties in with the slight differences observed between RSS and RSS_B with increasing w, in table 10.4. However, for each method and combination of w and σ, the values of $d_F(\hat{\mu}_{SL}, B)$ represent no change from the initial dissimilarity measures $d_F(\hat{\mu}_0, B)$, obtained with all landmarks fixed, with $d_F(\hat{\mu}_{SL}, B) \approx d_F(\hat{\mu}_0, B)$ to 3 decimal places for all w and σ. i.e. the final estimates of mean shape $\hat{\mu}_{SL}$ are no more different to B than each initial estimate, $\hat{\mu}_0$ was originally. Table 10 (right) indicates that $d_F(\hat{\mu}_{SL}, \hat{\mu}_0)$ increases slightly with w but not σ, although $d_F(\hat{\mu}_{SL}, \hat{\mu}_0) < 0.004$ always. Recall from the discussion of 10.2.3 that a change in shape of order 0.003 could result from just a 1-pixel change at each of the landmarks of the base shape (when scaled to have side

length=500 pixels). Therefore, these figures produced by each of the methods, are again of no real practical concern.

	Semi-landmark criterion	$d_F(\hat{\mu}_{SL}, B)$ (change from $d_F(\hat{\mu}_0, B)$)			$d_F(\hat{\mu}_{SL}, \hat{\mu}_0)$		
		σ=0.0000	σ=0.0125	σ=0.0250	σ=0.0000	σ=0.0125	σ=0.0250
w=0	None (All fixed)		0.002	0.004			
	Bending energy		0.002 (0.000)	0.004 (0.000)	0.000	0.000	0.000
	Full Procrustes		0.002 (0.000)	0.004 (0.000)	0.000	0.000	0.000
	Nearest point		0.002 (0.000)	0.004 (0.000)	0.000	0.000	0.000
w=0.15	None (All fixed)	0.009	0.009	0.008			
	Bending energy	0.009 (0.000)	0.009 (0.000)	0.008 (0.000)	0.001	0.001	0.001
	Full Procrustes	0.009 (0.000)	0.009 (0.000)	0.009 (0.000)	0.001	0.001	0.001
	Nearest point	0.009 (0.000)	0.009 (0.000)	0.009 (0.000)	0.001	0.001	0.001
w =0.3	None (All fixed)	0.016	0.016	0.016			
	Bending energy	0.016 (0.000)	0.016 (0.000)	0.016 (0.000)	0.004	0.004	0.004
	Full Procrustes	0.016 (0.000)	0.016 (0.000)	0.016 (0.000)	0.004	0.004	0.004
	Nearest point	0.016 (0.000)	0.016 (0.000)	0.016 (0.000)	0.004	0.004	0.004

Table 10.6: Full Procrustes distances of final mean shapes to base and initial mean configuration estimates from square simulations.

10.4 Summary and discussion

In this chapter we have investigated use of the new and original semi-landmark methods with sets of configurations generated from two simple geometric shapes. In the first investigation, configurations were produced from eight equally spaced landmarks, located around the circumference of a unit radius circle. Samples of 100 configurations were generated by displacing each landmark around the outline by an angular value, sampled from a uniform distribution and then adding isotropic noise, from a bivariate normal distribution. In the second investigation, a unit sided square was used with landmarks at each of the corners and four others placed exactly half way along each of the sides. Sets of 100 configurations were generated by producing perturbations of the four landmarks along each side, using displacements sampled from a uniform distribution and then adding isotropic noise to all landmarks, again from a bivariate normal distribution. Eight different sets of configurations were produced from each of the base shapes, by varying the range w and common standard deviation σ of the uniform and bivariate normal distributions respectively, allowing us to investigate the effect on the success of the different semi-landmark methods. Each set of simulated configurations were regarded as 'distortions' or representations of the same 'true' shapes, which we would hope the various methods would be able to recognise and

recover. For the simulations from a circle, all landmarks were set as semi-landmarks and allowed to move iteratively along escribed chords approximating the outline of a circle. For the simulations from a square, only the landmarks along the sides of each configuration were allowed to move, here along chords between the two adjacent corner landmarks, so as to approximate the sides of a square. Since the corner landmarks were not allowed to move, these directions were unable to change between iterations. A successful implementation of each method on any simulated dataset, would result in the variance in shape (*RSS*) being reduced to as small a value as possible, with the final estimated mean being the same shape as that from which the configurations were generated. The percentage reduction in *RSS*, compared to the value obtained for each dataset following an initial GPA with all landmarks fixed, was also considered since the initial value will be greater if larger simulation ranges were used.

In each investigation, differences in the final estimated means from the base configuration used to generate each dataset (and from the initial GPA estimates of mean shape), were found to be negligible, indicating no bias resulting from any of the methods. Since $d_F(\hat{\mu}_0, B) \approx d_F(\hat{\mu}_{SL}, B)$, this meant that similar conclusions would also have been reached had we examined the variance of the final fits about their base configuration (RSS_B), rather than values of *RSS*.

For each method, the final values of *RSS* were seen to be larger with increased noise (σ) compared to increased angular/linear variation (w). This was because the latter is much easier to remove than noise, having been simulated along directions which the chords are then defined to represent, rather than in all directions. (On the squares when there was no noise, for any w the final *RSS* value was zero, since the semi-landmarks were allowed to move precisely in the directions in which the linear variation was simulated.) With increased noise, the amount of remaining variation perpendicular to the chord directions (which is more difficult to remove), was greater, as the effect in both applications was to produce more variation in how well the first set of chords (which did not change, in the case of the squares), approximate the circumference of a circle or sides of a square. Noise simulated at the positions determining the directions of the chords resulted in inconsistent directions on different configurations, whereas at the positions of the semi-landmarks, this affects the locations of the chords in relation to the

rest of the configuration. Conversely, with larger w, similar effects were produced (circle simulations only) but to a much lesser extent, since the increased variation at the locations of the semi-landmarks and those determining their chord directions, is in directions which the chords were then defined to represent.

In both sets of investigations, all three methods were found to be successful in reducing the variation in Procrustes fits along the specified chord directions. In each application, the FP method produced the best results (in terms of smallest remaining variation in shape) with each combination of w and σ, although practically identical results were produced using the NP method when used on the square simulations, as in the applications of previous chapters. Despite producing slightly larger values of *RSS*, the BE method again gave comparable results to the FP method, as with several of the examples of chapter 9. As previously noted, if there are fixed landmarks in the vicinity of the semi-landmarks or if the semi-landmarks' chords are non-parallel, this restricts the possibility the their movement being in accordance with a large change in the affine component of the PTPS mapping from $\hat{\mu}_{(r-1)}$ (since $vec^{-1}(vec(X^P_{(r-1)}) - U_{(r)}\lambda_{(r)})$ must be matched to by the affine fit from $\hat{\mu}_{(r-1)}$ with minimum sum of squared non-affine components). If the optimal affine transformation is close to similarity transformation (as in both sets of applications here), the new positions of the semi-landmarks are very similar to those determined by the FP method, with any difference in their final positions being due to a GLS rather than ordinary sum of squares being optimised. This typically means that some variation remains along the final chord directions (as well as normal to them) with the BE method (but not with the FP technique), leading to higher final *RSS* values with the former. This difference in final *RSS* values between the two methods was seen to be larger with increased noise, as this generates more variance in directions which cannot be removed by a GLS criterion, but can by minimising an ordinary sum of squared Euclidean distances, over similarity transformations.

With the exception of the NP method, when used on the circles, each technique also had some success in removing some of the variation perpendicular to the final chord directions and at the other fixed landmarks (in the case of the squares). The FP method again produced the best results (with similar results being produced by the NP method when used on the square, but not circle simulations). For the NP method, we have seen

how GPA distributes the reduction in variance, achieved at the movement stage along the semi-landmarks' chords, around the rest of each configuration, when registering to an updated mean $\hat{\mu}_{(r)}$, resulting in an improved GPA fit for each configuration. When using the FP method, the OLS fit to $\hat{\mu}_{(r)}$ is optimised directly as the semi-landmarks move along their chords and so if $\hat{\mu}_{(r)} \approx \hat{\mu}_{(r-1)}$, the subsequent GPA registration of the new configuration is nearly always improved. Similar results were found using the BE method, when the affine component of the optimal PTPS transformation from $\hat{\mu}_{(r-1)}$ to each $vec^{-1}(vec(X^{P}_{(r-1)}) - U_{(r)}\lambda_{(r)})$ was close to a similarity transformation (although less variation is actually able to be removed due to the GLS criterion, as explained above).

As was found in the applications of the two previous chapters, with the FP and NP methods, any remaining variation at each of the semi-landmark positions in the final Procrustes fits was entirely perpendicular to the final chord directions on each configuration. For the square configurations, the patterns of remaining variation in the fits were visually identical for both new methods. Despite the NP technique taking a greater number of iterations to converge, it was again notably faster in terms of processing time, as seen in all previous applications. However, with the circle configurations (and for the first time in any of our investigations), the variance in the final semi-landmark positions (here all landmarks) was seen to be always greater for NP method, compared to the FP. Usually, the NP method is able to give very similar results by iterating between the movement and GPA steps. Providing the semi-landmarks move, the subsequent GPA registration to $\hat{\mu}_{(r)}$ is always improved, which typically allows the updated chord directions to pass successively closer to the landmarks in $\hat{\mu}_{(r)}$ and thereby more variation to be removed. Here however, on the first iteration, the landmarks move to position along their chords at equally spaced angles around the outline since this gives the points at nearest Euclidean distance to their corresponding positions in $\hat{\mu}_{(r-1)}$. Following GPA at the end of the iteration, if $\hat{\mu}_{(r)} \approx \hat{\mu}_{(r-1)}$, the semi-landmarks of the new fits are then almost already in their optimal (NP) positions in the rotational directions of the chords, for the next movement step and so hardly move again, leaving any variation perpendicular to the outline, unable to be removed on the first iteration. With increased noise, more perpendicular variation was found to remain, as the initial approximation to the outline by the chords became worse.

Chapter 11

Summary and future work

11.1 Introduction

In this, the final chapter, we bring together the main points and findings from this study and consider the benefits and implications of what has been achieved.

In section 11.2 we summarise the content of each chapter, recalling the key purpose of each main section and the major issues which needed to be addressed. Section 11.3 then reviews the results of our investigations of each of the new semi-landmark methods, comparing their main properties with those of the original bending energy technique. In the conclusions of 11.4, we return to review the aims set out at the start of this thesis and offer recommendations and advice as to when use of the new methods will be more beneficial than employing Bookstein's (1996a,d,e) approach to assigning landmark homologies, in order to remove unwanted patterns of variation in shape; not only for studies of tooth shape, but also in more general situations. Finally in section 11.5 we consider suggestions and ideas for further exploration and development of this work. We recall some of the alternative approaches, discussed in earlier sections, for addressing difficulties associated with the analysis of tooth shape and discuss new ideas and proposals for use in wider applications, where lack of landmark correspondence is to be accounted for or removed in some way.

11.2 Summary of thesis by chapter

Chapter 1 gave background details to the project, highlighting the need for this study. The overall aim is to improve the way the shape of teeth is analysed, and in terms of methodological development, to allow lack of exact correspondence between features on different objects to be addressed in a more comprehensive manner than is possible

using existing techniques. Details of the equipment and methods of data collection available were given, as well as a thesis outline, which briefly described the content of each chapter.

Here we first introduced the reader to the field of shape analysis, describing how methods for the study of the shape require that location, rotation and scale are removed or accounted for in some way, with the aim being to estimate average shapes and the structure of shape variability and to carry out inferences on population quantities. Having provided preliminary definitions of the different tooth types, crown surfaces and features, standard orthodontic measurements and formal definitions of terms such as 'landmarks', we noted how previous studies of tooth shape had typically been limited to taking traditional orthodontic measurements and performing standard univariate statistical analysis on quantities or 'indices' derived from them. For example, Peck & Peck's (1972) crown shape index. However, in recent years, improved techniques for the analysis of shape had emerged in the fields of mathematics and statistics, which, in contrast to the 'traditional' methods, capture the geometry of the objects involved and preserve this information throughout the analysis. Since we require that methods are suitable for use on un-extracted teeth, a key problem identified was that any such analysis must be based only on information from the visible edges and surfaces of a tooth. Methods must be able to recognise that teeth may have the same shape but differ in how they are represented for analysis, due to differences between patients in the position of their gingival tissue.

Chapter 2 reviewed and evaluated the most popular methods for the analysis of shape, in terms of their suitability as a starting point for our investigations. The techniques were separated by their associated data source into 'landmark' and 'outline' approaches. The former were further categorised as either 'distance', based on inter-landmark measurements' or 'geometric' techniques, where information on the relative locations of landmarks is always retained. For this study we had the option of using either type of data. Details of discussion forums, websites, mailing lists and an overview of the currently available software for performing the various methods were also described.

The advantages and disadvantages of the different techniques were considered, noting difficulties likely to arise when used for the study of tooth shape and identifying which

approaches offered the best scope for possible development and/or modification, in order to address issues such as unwanted differences in shape resulting from variation in the position of patients' gingival tissue. The most logical starting point for our investigations was seen to be the Generalised Procrustes Analysis (GPA) technique. Along with the other tools of the 'morphometric synthesis', it was noted how, in recent years, this has become the standard protocol for investigations of shape when objects may be represented as configurations of landmark data, having advantages over both other landmark and outline methods, in terms of developed theory, statistical properties and the variety of descriptive and inferential techniques which may be utilised. A key clinical benefit over other methods noted, was that it allows interpretation of results as pictures in the original space of the objects being considered, since the geometry of the objects is retained throughout the analysis.

Chapter 3 provided a comprehensive description and demonstration of Procrustes analysis, introducing these ideas for the first time to the study of tooth shape. Using two sets of configurations of landmark coordinates, obtained from buccal images of central incisors from patients with hypodontia and a corresponding control group, we illustrated how, after optimally matching configurations to account for the unwanted effects of location, scale and rotation, we can then define ways of measuring dissimilarity (or distance) between shapes, estimate mean shapes, and examine shape variation. Several different methods for estimating the Procrustes mean shape and corresponding registered fits have been proposed and so a comparison of the different algorithms and solutions was presented, detailing differences between the scaling options used and the resulting summary measures. Details were also given of how landmark configurations can be regarded as single points in Kendall's shape space (Bookstein, 1986, Kendall, 1984), with metric defined by the 'Procrustes' distance between shapes and how calculation of an appropriate tangent space, a linearised version of shape space in the vicinity of the estimated average, forms the fundamental geometric setting for further analytical investigations, allowing most conventional descriptive and inferential techniques to be adapted and applied successfully.

Despite being able to demonstrate a difference in mean buccal surface shape between our two groups, issues and caveats relating to the general suitability of Procrustes methods were noted. GPA (in its standard form) assumes that the landmarks are

reproducible and have good correspondence between cases, but in study of tooth shape there are several difficulties in meeting these requirements that need to be addressed.

The effect of operator inconsistencies in the positioning of landmarks, on the Procrustes analysis of shape was considered in **chapter 4**, where we investigated the importance of this problem with data obtained from the buccal and occlusal surfaces of various tooth types. Using different operators' images and landmark representations from a variety of patients, the consequences of location inconsistencies were evaluated by calculating its effect on the recorded variation in Procrustes fits obtained from each set of multiple representations. Two new methods were presented for establishing the proportion of variation in shape due to operator inconsistencies in obtaining landmarks relative to actual variation between patients, i.e. the inter-operator reliability. The first extended the work of Arnqvist & Martensson (1998), summing estimated components of variance from the principal component scores of each set of Procrustes fits. The second used analogous calculations based on a 2-way ANOVA partition of the sum of squared Euclidean distances (approximate sum of squared Procrustes distances) between the Procrustes registered configurations, eliminating the need to calculate principal components. For the latter, details of how to obtain a similar one-way (intra-operator) reliability measure, which may be of use for future shape reliability studies, were also given. Overall, it was found that operator inconsistencies in the location of landmarks are a particular problem if teeth are to be represented and analysed as landmark configurations. For the occlusal surfaces especially, reliability was found to be poor.

An examination of the 'within case (patient)' covariance structures in the Procrustes fits by principal components analysis and feedback from the operators helped to identify which particular patterns and directions of variation in shape had resulted from inconsistent representation of the same patient's tooth surface. For the buccal surfaces, positions around the outline, such as those at the ends of the mesio-distal width, the papilla endpoints and corners of the incisal edge were found to be more difficult to identify consistently than (say) the landmarks at the ends of the LACC or cusp tips. For the occlusal surfaces, a potential source of error was found to be in the subjective orientation of the tooth surface in the bucco-lingual direction prior to imaging. Such findings have implications in GPA where each landmark carries equal 'importance'. In addition to recording inconsistencies, it was also noted how 'nuisance' variation,

resulting from differences between patients in the position of their gingival margin and inter-dental papilla, also leads to variations in shape (between patients) which are of no interest, regardless of the reproducibility of landmarks in these regions. However, the reliability figures have no way of taking this into account. In terms of the aims of this thesis, methods of analysis will need to accommodate these problems, if landmark data are to be used to describe variations in tooth shape. Teeth which are identical in shape but have differences in Procrustes coordinates due to unreliable positioning of landmarks and differences in the position of landmarks due to a patient's gingival tissue need to be regarded as, or identified as, the same shape.

At the end of chapter 4, it was noted that the papilla and gingival landmarks are still useful, because they provide the best indication we have of the cemento-enamel junction and the mesio-distal dimensions of the teeth in these areas. Similarly, points which do not match exactly, but lie shifted (say) around the outline of a tooth (which would be of no interest), still describe differences in shape perpendicular to these directions, where we expect most of the between patient variation to be. In light of this, recent work on 'semi-landmarks' (Bookstein, 1996a,d,e) appeared to offer one way in which lack of precise landmark correspondence may be overcome and so the usefulness of this technique was considered in **chapter 5**. The method recognises that certain landmarks may be known to lie along particular lines or curves, but are difficult to locate precisely. Here we described how an additional standardisation step is introduced into the GPA procedure, allowing these landmarks to move along pre-specified directions, typically along 'escribed chords' (a line through the nominal position, parallel to the line joining the two neighbouring points). For each configuration, new locations along these chords are determined as those which minimise the bending energy (BE) of the pair of thin plate splines (PTPS) transformation from the current Procrustes mean shape. A new mean shape and Procrustes fits are then computed, along with new chord directions, and the process repeated until the semi-landmarks stop moving. The final resulting set of Procrustes fits may then be analysed in the same way as the results of an ordinary GPA.

Interpolating spline transformations and their use in shape analysis (using a pair of thin-plate splines) were described at the start of chapter 5. Details of how the PTPS formulation is then modified to allow certain landmarks to move along chords to assign

point-to-point correspondences was then described, as well as analogies of the PTPS and semi-landmark formulations with generalised least squares (GLS) affine fitting. Use of the semi-landmark method for the analysis of tooth shape was then considered, using information gained from the reliability datasets to specify appropriate directions in which the landmarks might be allowed to move. At the start of these investigations (in 2000), no readily available software existed to perform the computations and so specific S-plus routines had to be written. In addition, documentation of the practical use of the technique was limited; only papers with Bookstein as an author had used the method. One unexpected issue that did arise concerned the level of convergence that can be achieved when using the method. However, it was argued that in practical terms, with most datasets this matter will typically be unimportant. For addressing the issue of unwanted variation in landmarks due to the position of a patient's gum, the semi-landmark method was seen to produce some undesirable results. Use of the BE criterion to determine the new positions of gingival landmarks, with chords representing directions of between patient variation, was found to produce large, unrealistic changes in their positions.

In **chapter 6**, we therefore considered ways in which the semi-landmark or Procrustes method could be modified in order to eliminate this problem and address the issue of unwanted variation in the positions of gingival landmarks. We examined other possible mappings and objective criteria that may be optimised (with reference to the current mean shape) when determining the new positions of the semi-landmarks along their chords and the use of existing variants of Procrustes and spline mappings, which aim to allow for lack of landmark correspondence during registration.

Use of pairs of smoothing splines to prohibit large movements of the semi-landmarks away from their original positions was first considered. We showed how the original semi-landmark method can be considered as a special case of an anisotropic smoothing spline and how its associated objective function may be used to restrict large movements of semi-landmarks by adjusting the relative weights used to penalise the BE and distances moved by the semi-landmarks. However, it proved difficult to find a common weighting parameter which gave a suitable balance between the gingival landmarks moving too far or not at all. We also considered the use of more general, higher order splines and described how along with PTPS and (pairs of) smoothing

splines, these are special cases of multiple kriging predictors. We showed how the formulae may be modified to include semi-landmarks, and how this could then allow a more general class of polynomial functions to be used in the 'global' part of the mapping and more generalised penalty functions to be investigated, again with the relative weights of the BE and movement components pre-specified. However, use of polynomials higher than first order was found to produce several unwanted results, including folding.

Use of affine mappings (polynomials of first order) had been noted as being central to the issue of gingival semi-landmarks moving large distances along their chords, and so ways of restricting or penalising the shape change produced by the affine component of a PTPS mapping, when determining the new positions of semi-landmarks, were considered. Various measures of affine shape change were investigated and penalised relative to the BE of any mapping. However, again it was found difficult to specify the relative weights of the two penalties in order to work equally well on all configurations.

We then turned our attention to transformations and optimisation criteria not based on spline mappings, for the determination of new semi-landmark positions. This included the minimisation of the full Procrustes (FP) distance between the mean and individual configurations or simply moving the semi-landmarks to the nearest point (NP) along their chords to the corresponding landmarks in the mean shape (following prior GPA registration). Of the new methods proposed and investigated, these appeared to offer the most promising results for removing unwanted variations in the gingival landmark positions and so were taken forward for further investigation.

For completeness, the final part of chapter 6 reviewed some of the other variants of Procrustes and spline mapping techniques that have been proposed for use when there is lack of precise correspondence between landmarks along particular directions or where one would wish account for certain unwanted differences in shape in some way. The ideas of weighted Procrustes analysis were presented, as well as Green's (1996) extension of the semi-landmark technique to include curves and outlines.

The remainder of this thesis comprised a detailed investigation and evaluation of the FP and NP techniques, both for the study of tooth shape and when used on configurations

generated from simple geometric shapes. Firstly in **chapter 7**, we considered some of the implementation issues arising with the use of the new methods, as we did for Bookstein's original BE method in chapter 5, such as convergence assessment and, for the NP criterion, the importance of the GPA method used. For future reference, details were also given of the specifically written S-plus function used for performing the investigations, including details of the various input and output arguments that may be specified and the algorithm used to optimise the FP criterion, since this does not have a straightforward analytical solution.

In **chapter 8** we investigated how well the new semi-landmark methods performed, compared with the original BE technique, in addressing the gingival variation problem, using configurations representing the buccal surface of upper central incisors. As before, the key aim was that any variation in shape due to differences in the positions of the patient's gingival landmarks be removed in the same way as we filter out differences due to size, location and orientation, by allowing these landmarks to move iteratively along chords representing the directions of unwanted variation. In addition, it was also required that each method be able to identify that representations of the same tooth, but with the gingival margin and inter-dental papillae in different positions, are still the same shape. Since we had no way of investigating this using study casts (as the gum is always in one position), images of patients' teeth where the entire actual crown is visible were used and different possible landmark representations of the clinical crown (which would be visible in practice) generated by simulating different possible positions of a patient's gum. Use of probability models was made with parameters chosen to reflect what is known about gingival variation from both the dental literature and experience of periodontologists.

After giving careful consideration to the choice of summary measures to be used to compare the results of the different techniques, it was found that when using the NP and FP criteria, both methods gave very similar results and were successful in eliminating variation in the gingival landmark positions in the directions of the specified chords, both within and between cases, and also at the fixed (non-gingival) landmarks (see section 11.3.1). For all but 2 out of the 32 sets of simulations, the final positions of the semi-landmarks were within the ranges of 'feasible' variation, identified a priori. In contrast, despite eliminating variation within cases, use of the BE method was found to

produce an increase in variation between cases. For many of the sets of simulations, different representations of the same tooth all converged to the same unrealistic shapes. After categorising the outcome for each case into one of three groups, discriminant analysis was used to identify the shapes of the actual crowns they were simulated from that best distinguishes between results. This confirmed that the outcome of using the BE method with these particular configurations, semi-landmarks and chord directions, depends on the proximity of the fixed landmarks to each other (in the direction of the semi-landmarks chords), relative to their corresponding positions in the mean shape (as previously suggested in chapter 5).

In **chapter 9** we compared how well the original BE and new FP and NP methods performed in other tooth shape applications, specifically in addressing some of the main reliability issues identified in chapter 4. Use was made of the same datasets obtained by multiple operators from the buccal and occlusal surfaces of a variety of tooth types, from which reliability figures had previously been obtained. From the buccal surfaces, operator inconsistencies in the locations of landmarks along the sides of the teeth and at the corners of their incisal edges were found to be a common problem and for the occlusal surfaces unwanted landmark variation along the incisal edge and cusps tips of teeth appeared to result from orientation differences in the bucco-lingual (BL) direction at the imaging stage. Suitable chord directions having been defined to represent the directions along which lack of landmark homology occurs, the effectiveness of each method was judged by how much of an improvement in the reliability figures was able to be achieved. For the occlusal surface applications, we also introduced the idea of constraining a subset of the semi-landmarks to always move by inter-dependent proportions along their chords, representing the effect of tilting each three dimensional surface in the BL direction. In each application, all three methods were found to be successful in reducing the variation at the semi-landmark positions along the chosen chords directions, when compared to the original GPA registration of the data. A marked improvement in the reliability figures indicated that most of the variation removed had been 'within-cases' (due to operator inconsistencies) rather than between. As was found in chapter 8, the results for the FP and NP methods were almost identical in each application, with all remaining variation in shape at the semi-landmarks normal to the final directions of each configuration's chords. However, in contrast to chapter 8, here use of the BE criterion was found to produce more comparable results to the NP

and FP methods, although with some variation always remaining along the directions of the final escribed chords, as well as normal to them (see section 11.3.2).

We then also investigated how each of the methods perform when we also include semi-landmarks and chord directions representing variation due to differences in the position of patients' gingival tissue, again on the buccal surfaces of upper central incisors and now also on the corresponding occlusal surface with the landmark at the lingual endpoint of the BL width free to move along the BL direction. Here the BE method again produced unrealistic shapes, but both the FP and NP methods were successful in reducing the variation at each of the semi-landmarks, again leaving only that perpendicular to their final chord directions. An example was given of how the means of two populations of upper central incisors may be compared, having used the new semi-landmark methods to remove unwanted variation due to both operator inconsistencies and gingival variation, and a significant difference between the mean shapes of hypodontia and control patients was demonstrated.

In **chapter 10**, we compared how the semi-landmark methods performed in removing different patterns of variation from configurations generated from simple geometric 'base' shapes. In the first investigation, a circle with eight equally spaced landmarks around the circumference was used as the base shape and in the second, a square with landmarks at each of the corners and four others placed exactly half way along each of the sides. Samples of configurations were generated as perturbations of all (in the case of the circle) or some (in the case of the square) of the base shape landmarks around the circle/square perimeter, by sampling displacements from a uniform distribution and then adding isotropic (bivariate normal) noise to all positions, so that they no longer lie on the outline. Different parameter values were used to produce eight different sets of configurations from each base shape with various ranges and patterns of variation in order to investigate the effect on the success of the different semi-landmark methods. Each set of generated configurations were regarded as distortions of the same 'true' shape (their base shape), which we would hope the semi-landmark methods would be able to recognise and recover, with the variance in shape being reduced to as small a value as possible and the final estimated mean of each sample being the same shape as that from which the configurations were generated. For the simulations from a circle, all landmarks were set as semi-landmarks and allowed to move iteratively along

escribed chords approximating its circumference. For those from a square, only the landmarks along the sides of each configuration were allowed to move, along escribed chords between each two adjacent corner landmarks, so as to approximate the sides of a square. Since the corner landmarks were fixed, these chord directions were unable to change between iterations.

As was found in previous applications, for the FP and NP methods, any remaining variation at each of the semi-landmark positions in the final Procrustes fits was entirely perpendicular to the final chord directions on each configuration. For the square configurations, the patterns of remaining variation in the final Procrustes fits were visually identical for the FP and NP methods, but for the circles, this was seen to be always larger for NP method (see section 11.3.2). For all three methods, the remaining variation in shape was found to be larger when there is increased noise compared to increased angular/linear variation, as the former is much easier to remove, having been simulated along directions which the chords are meant to represent. The effect of increased noise in both applications was to produce greater variation in the directions and locations of the first set of chords on different configurations and therefore in how well they approximate the outline of a circle or square. Despite leaving some variation in the direction of the final chords, which became larger with increased noise (see section 11.3.2), the BE method again gave comparable results to the two new methods.

11.3 New semi-landmark methods: Overview and summary of key points

Following registration by GPA, we have seen how the new positions of semi-landmarks may be calculated as the nearest point (NP) along each chord to the corresponding landmark in the mean shape or as those which minimise the full Procrustes (FP) distance (or, equivalently, the ordinary sums of squares over similarity transformations) between all corresponding landmarks of the individual and mean configurations. The key findings from our investigations into the behaviour of these new methods are summarised below, noting differences and similarities between their properties, and when compared to the results of using the original BE method. The section ends with a look at recent developments of the 'TPSrelw' program by F.J. Rohlf, available on the

SUNY Stonybrook website. Following development of the new ideas presented here and the writing of S-plus routines to implement these and the original technique, a feature was added in 2003 for implementing Bookstein's original semi-landmark method (see section 5.5.4), and in summer 2004, a number of modifications were made.

11.3.1 Registration, biasness and processing

- Although ordinary least squares (OLS) and FP criteria do not produce the same minimum for their objective functions, the resulting configurations have the same shape (see 6.2.3). Note also that with the FP method, the resulting configuration is not the Procrustes fit to the mean; the fit to a new mean is obtained at the end of the iteration.

- With the FP (and BE) method, there is no dependency of where the semi-landmarks move to on the prior alignment of the mean and individual configuration, whereas with the NP method the new positions will be influenced by their registration. However, use of different GPA methods was found to have very little effect on the end results.

- With each new method the change in mean shape from the original estimate was found to be negligible in each application (with a full Procrustes distance of no more than 0.004 units between the initial and final means), indicating a lack of systematic bias. For the BE method however, there were several applications where large movements of semi-landmarks occurred, leading to more substantial changes (see 11.3.2).

- For the FP method, the final Procrustes fits are determined on more or less the first iteration, since rather than just minimising Euclidean distances between points on the chords and landmarks in mean shape on the movement step, the configuration may also be rotated, scaled or translated, with all new semi-landmark positions determined simultaneously. If the updated mean is practically unchanged, the semi-landmarks are almost already in their optimal positions for the next movement step. With the NP method, more iterations are required since on any movement step, the new positions of the semi-landmarks are determined without allowing the configuration to change its registration to the mean. This is done at the end of the iteration, with respect to a new (usually practically unchanged) mean. The GPA fit is always improved (see 11.3.3), which typically then allows the updated chord directions to pass successively closer to

the landmarks in the mean and thereby more variation to be removed. However, even with the mean hardly changing, several more iterations are required before similar results to the FP method are produced. By comparison, the BE method was found to usually take around the same number of iterations as the FP method.

- In terms of actual processing time, the FP method always took considerably longer than the NP (or BE) methods, despite the NP method requiring more iterations to reach the same level of convergence. This is because the NP and BE criteria have explicit analytical solutions for the determination of the semi-landmark positions, whereas the FP criterion requires the use of an optimisation algorithm.

11.3.2 Final results and potential for unrealistic shapes

- The FP and NP methods generally yield almost identical end results (see sections 7.3 and 8.3.3.1). Indeed they are usually visually identical. In all but one of our applications, the final values of $RSS(\hat{\mu})$ (the residual sums of squares of the landmarks of the Procrustes fits about those of the estimated mean, which is equal to the residual sums of squares about the arithmetic mean of the Procrustes fits (RSS) if the iterative, rather than explicit, GPA method is used) were always practically identical.

The exception was when the NP method was used on the simulations from circles, where the variance in final semi-landmark positions (here all landmarks) was seen to be larger. On the first movement step, the semi-landmarks moved to positions along their chords at equally spaced angles around the outline (to give points at least Euclidean distance to those in the mean). However, since the mean is practically unchanged, following GPA the semi-landmarks are then almost already at their NP positions for the next movement step, as the fit in the rotational directions of the chords has already been optimised. Consequently, the semi-landmarks hardly move again, leaving most of the variation perpendicular to the outline that remained from the first iteration.

- Both new methods were always successful in eliminating the variance along the specified chord directions, leaving any remaining variation in shape in the final Procrustes fits at these positions entirely perpendicular to their final chord directions.

- Despite often also being successful in this regard (see next point), on several occasions the BE method was seen to produce large unrealistic movements of the semi-landmarks of some configurations, resulting in an increase, rather than decrease in $RSS(\hat{\mu})$.

  In order to produce the minimum bending energy PTPS transformation of the mean, it was noted how the semi-landmarks of a configuration move to positions which result in the mapping which gives the closest possible 'affine fit' of the mean (in terms of the sum of squared 'non-affine' components) to the configuration's fixed landmarks and any positions along the semi-landmark's chords. That is, the PTPS mapping tries to represent the transformation as 'globally' as possible using the affine component, so that there is the smallest possible variation in the corresponding non-affine or 'local' components added to ensure that points are mapped exactly to their targets (the fixed landmarks and chord positions of the target configuration). When all the semi-landmarks are free to move in roughly the same direction and there are no nearby fixed landmarks, or other semi-landmarks whose movement is restricted to a different direction, large movements of the semi-landmarks will occur when a notable change in the optimal affine component of the PTPS transformation from the mean shape is able to occur in the direction of the chords, determined by the configuration of landmarks effectively fixed in these directions, relative to those in the mean shape. Unrealistic shapes are produced as there is nothing in the BE criterion to stop the semi-landmarks moving unrealistically large distances in accordance with this new mapping, as this reduces the size of the non-affine components (BE). However, when there are fixed landmarks near to the semi-landmarks or differences in the chord directions result in certain semi-landmarks being effectively constrained in the direction of other semi-landmarks chords, the affine component of the PTPS mapping is unable to chance considerably, restricting the possibility of large semi-landmark movements occurring.

For example, in section 5.5.6.2, it was seen that when the BE method was used with semi-landmarks and chords representing gingival landmark variation in the buccal view, the optimal positions for them to adopt were in accordance with the PTPS transformation of the fixed landmarks around the lower (incisal) outline in the direction of the roughly parallel chords (and in particular it's affine component), which resulted in the semi-landmarks moving large distances. If (say) the fixed landmarks were relatively closer together in the directions of the chords in the target configuration,

compared to the mean, this resulted in the semi-landmarks moving downwards in order to minimise the GLS residuals of the affine fit to the fixed landmarks and chord positions. There were no fixed landmarks or semi-landmarks with different chord directions nearby to the gum landmarks to stop large changes in the affine component occurring. In Bookstein's 1996a,d,e original application, large movements did not occur because the choice of chord directions, around a closed outline, constrained the semi-landmarks to stay around the outline, with the shape of the affine component not really able to change.

However, with the new methods the semi-landmarks are unable to move large distances since, when using the FP criterion, it is required that the fit of the entire configuration to the mean shape is optimised and with the NP method, the prior GPA registration ensures that the semi-landmarks need not move too far in order to find positions along their chords at nearest Euclidean distance to corresponding landmarks in the mean.

- When the BE method is successful in removing variation along the specified chord directions, producing more comparable results to those obtained by the FP and NP methods (see chapters 9 and 10), what variation is left appears more isotropic, with some still remaining along the direction of the final chords, as well as perpendicular to them (and hence producing a larger value of $RSS(\hat{\mu})$).

As noted above, for the BE method to produce similar results, this requires that there are fixed landmarks in the vicinity of the semi-landmarks, or that the chords are in non-parallel directions (which must be matched to exactly by the PTPS mapping from the mean), so that large changes in shape cannot occur. For the FP criterion, the new semi-landmark positions at each movement stage are determined by minimising an ordinary sums of squares between all landmarks of the target shape and mean over similarity transformations of either configuration (both approaches give the same answer, but for consistency with the BE and NP methods, we can think of this in terms of the former). When using the NP method, there is no transformation of the mean or individual configuration involved, just a minimisation of Euclidean distances (or equivalently, ordinary sums of squares) between the chords and corresponding landmarks in the mean. However, iterative re-alignment by GPA, ensures that when using that same convergence criterion, the FP and NP methods usually yield practically identical results.

With the BE method, new semi-landmarks positions are found by minimising a generalised sums of squares over affine superimpositions of the mean to the individual configuration as the semi-landmarks vary along their chords. In addition to similarity transformations, the affine mapping allows scaling of the mean in either the x or y direction and shearing, when superimposing to the target configuration, whereas with the FP method, transformation of the mean is restricted to rigid transformations of location, rotation and scale only. Consequently, the new positions of the BE, FP and, after several iterations, NP methods will be similar when the affine component of the PTPS mapping is close to being a similarity transformation. When this is true, at any movement step the remaining difference in the new positions, between the FP and BE methods, will be due to an ordinary, rather than generalised, least squares criterion being optimised. Variation in the semi-landmarks will still remain in directions along their final chord positions, rather than only perpendicular to them, as when using the new methods, since at each movement step, the semi-landmarks move to positions which minimise a weighted sum of correlated distances between the affine transformed version of the mean and the individual configuration, rather than ordinary Euclidean distances between corresponding landmarks, leading to larger contribution to $RSS(\hat{\mu})$.

11.3.3 Effects of noise and variation at fixed landmarks

- With increased noise at all initial landmark positions, the amount of variation remaining perpendicular to the final chords, unable to be removed, increases. For the new methods, that which is along the chord directions is removed entirely by the movement of the semi-landmarks as described above, whereas perpendicular variation is only able to be removed at the GPA registration stage and, for the FP method, when allowing the superimposition of the mean and individual configuration to change during the movement step. Additionally, noise produces greater variation in the directions and locations of the first set of chords on different configurations, which also contributes to the variation unable to be removed.

As was seen in chapter 10, the difference in final $RSS(\hat{\mu})$ between the FP and BE methods becomes larger with increased simulated noise, since more variance is generated in directions which cannot be removed by the landmarks moving in

accordance with a PTPS mapping of the mean and a GLS criterion, but can by minimising an OLS of Euclidean distances over similarity transformations.

- As well as reducing the variation at the positions of the semi-landmarks, it was also found that after running each new method (and between successive iterations), changes also occurred in the patterns of variance at the other, fixed landmarks, with a decrease in the contribution to $RSS(\hat{\mu})$ always resulting for each configuration. By removing variance at the semi-landmarks, this is then no longer able to influence the variation at the other landmarks by being distributed around the configuration by GPA and so by using these methods, the true variation in shape at these positions is able to be represented more accurately.

Sections 7.4, 8.3.3.3 and 9.3.3 described how on each iteration, with the FP method, the fixed and semi-landmarks are matched simultaneously as the semi-landmarks move along their chords, resulting in an improved overall fit to the current mean (and new mean if this hardly changes). Similar results will also occur with BE method, if the affine component resembles a similarity transformation. With the NP method, the GPA registration step, gives equal weight to all pairs of corresponding landmarks in the new mean and each configuration and distributes the reduction in Euclidean distances achieved by moving the semi-landmarks, around the rest of the configuration. Until convergence, this always results in a decrease in the contribution to $RSS(\hat{\mu})$ from each configuration. Typically therefore, each method has some success in removing some of the variation perpendicular to the final chord directions (with the exception of the NP method used on the circles). However, how the variance at each fixed landmark changes will depend on the choice of semi-landmarks, chord directions and configurations in question. For example, in chapter 8, a reduction in variance was noted at all fixed landmarks after using both new methods whereas in chapter 9, a reduction in variation at some landmarks was seen to be at the expense of an increase at others.

11.3.4 Convergence

For each of the new methods and with the original BE method, there is always a level of tolerance below which the change in $RSS(\hat{\mu})$ ($\Delta RSS(\hat{\mu})$) between successive iterations

will not converge, with $\Delta RSS(\hat{\mu})$ remaining around the same magnitude once this level has been reached, regardless of the GPA method used. Once the process reaches a particular level of convergence, the semi-landmarks start to move in the same direction on each iteration, by very small but consistent amounts. This results in a systematic change in the mean shape, which requires that semi-landmarks move in a similar way on the following iteration and so on, resulting in a small, but consistent increase or decrease in $\Delta RSS(\hat{\mu})$ between iterations. These systematic changes in the mean are very small, but large in comparison to the continual changes in shape of some of the configurations relative to their mean, thereby offsetting any decreases in $RSS(\hat{\mu})$ produced, slowing convergence and stopping $\Delta RSS(\hat{\mu})$ falling below a certain level.

In practice however, this issue does not matter. For all of the datasets considered and with each of the semi-landmark methods, the changes in shape concerned were very small, with $\Delta RSS(\hat{\mu})$ no greater than $\varepsilon = 0.001 \times n$ where n is the number of configurations involved. This guideline was suggested in 5.6.6, by examining how large a change in $\Delta RSS(\hat{\mu})$ is possible from a 1 pixel change at all 8 landmarks of 100 similar configurations and gives a level of convergence at least as stringent as the recommendations of Gower (1975) and Rohlf & Slice (1990), that a change in $RSS(\hat{\mu})$ of 0.001 or 0.01 is adequate for convergence of the iterative GPA method. A general approach, as we have used throughout our investigations, would be to identify the iteration where $\Delta RSS(\hat{\mu})$ slows (perhaps by plotting values against iteration number) and then to check whether or not an acceptable level of convergence for the number of configurations involved has been achieved. The semi-landmark process can then be re-run for this pre-specified level of convergence, without the danger of the configurations becoming unrealistic in shape, from running the routine too long.

11.3.5 Recent updates to TPSrelw

A very recent (2004) modification to the semi-landmark facility of Rohlf's TPSrelw program has been to include an option to update only the mean, rather than the set of configurations at the end of each iteration. Presumably, at the final iteration the semi-landmarks actually move to the new positions which result in no change in the mean shape, otherwise this does nothing to eliminate the variance in semi-landmark positions

along their chords. In the accompanying 'help' file, Rohlf argues that iteratively updating the mean "may produce more stable computations, although this may depend on the data in question". We tried this by modifying the '**semi.it.new**' function, but found that this still does not stop the problem of there always being some low level of tolerance, below which each process fails to converge.

Until summer 2004, Rolf's program used the BE criterion for the determination of new semi-landmark positions, but recently this has been updated to include an option to 'slide' the landmarks along their chords to minimise the squared Euclidean distances between corresponding landmarks (in the same way as the NP method introduced here). However, there is still some restriction on the direction of chords which may be used and ambiguity over the level of tolerance and assessment of convergence used, although results were comparable with those of our own when using the NP method.

11.4 Conclusions

This thesis presents new statistical methodology for both the analysis of tooth shape and for use in more general applications where objects are to be represented and analysed as homologous configurations of landmarks and one would wish to account for lack of precise correspondence between certain points in some way.

In contrast to previous studies, which have typically been based on linear measurements or quantities derived from them, we have shown how use of modern statistical techniques can enable dental researchers to work with the full geometry of the teeth, which has otherwise been lost or ignored. A key benefit is that this allows graphical representation of results in the original space of the objects and has many advantages over other, distance-based, methods in terms of various statistical properties such as power (Adams *et al.*, 2004). In addition, the new semi-landmark methods introduced here offer ways which investigators can allow for unwanted differences in shape, resulting from differences between patients in the position of their gingival tissue, an important difficulty that has previously been neglected. Moreover, poor reliability in the recorded locations of landmarks along particular directions may be taken into account, which also enables more meaningful comparisons of tooth shape to be made.

In chapter 6 it was noted that Bookstein had recently observed that there had been very little documentation on the practical use of the semi-landmark technique and stated that he had hoped that someone would explore other possible ways to generate new semi-landmark positions. In this and the subsequent chapters of this thesis, we have proposed several methods for doing this and identified two particular choices of criteria which appear to offer improvements and advantages over the original approach. Although the bending energy (BE) technique will still be useful in situations where it is desirable to standardise configurations with respect to their non-affine dissimilarities from the mean shape, the modifications presented here appear to be much more useful for determining the new semi-landmark positions in more general applications.

Often there will be situations where the different methods produce similar results, typically if there are fixed landmarks in the vicinity of the semi-landmarks or differences in the chord directions of semi-landmarks, and so the PTPS transformation from the mean is unable to change substantially. However, as we have seen in some of the investigations in this thesis, when this is not the case (as will be likely in many situations), use of the BE criterion may lead to increased variation in shape, with the resulting configurations for some objects becoming very dissimilar from the shape they were intended to represent. For removing variation in shape due to differences in the positions of landmarks along particular directions, the BE method clearly fails in such situations, whereas use of the new FP or NP criteria was always been found to produce successful results. With these methods the movements of the semi-landmarks are unable to produce large changes in shape, since the fit of the entire configuration to the mean is being optimised (FP method) or because prior GPA registration ensures that the semi-landmarks need not move too far along their chords in order to find points at nearest Euclidean distance to their corresponding positions in the mean (NP method).

Over the course of the iterations the new methods will typically yield almost identical results, with all remaining variation in the final Procrustes fits, at the semi-landmarks, being entirely perpendicular to their final chord directions. However, one drawback found with larger samples, was that convergence of the FP method can be very slow, always taking considerably longer than the NP method, despite the latter requiring more iterations to reach the same level of convergence.

11.5 Future work

In this final section we consider ideas for further exploration and development. We recall some of the suggestions made in earlier sections of this thesis for modifying the original semi-landmark technique and discuss ways of extending the new methodology, for both the study of tooth shape and use in more general applications. In the usual notation, let $T=\hat{\mu}$ denote the current estimate of mean shape and $vec^{-1}(vec(Y^0)-U\lambda)$ the new version of a $k\times 2$ configuration Y^0, where U is a ($2L\times 2$ matrix of zeroes and pre-specified unit chord directions for the L semi-landmarks as defined in section 5.4.1, and λ is a vector of L scalars to be determined.

11.5.1 Alternative optimisation criteria

11.5.1.1 Combinations of penalties

In chapter 6, we saw how a weighted combination of more than one objective function may be used to determine the new locations of semi-landmarks along their chords. New positions were found as those which minimise the BE of the PTPS mapping from the mean to the fixed landmarks and chords positions of the target configuration, relative to either a weighted sum of squared distances moved by the semi-landmarks (with the resulting penalty function taking the form of an anistropic smoothing spline), or various measures of the shape change produced by the affine component of the mapping. We could therefore also consider other combinations of objective criteria, including ones involving the new FP or NP penalties for determining new locations of semi-landmarks. For example, if positions are still to be determined by a PTPS mapping from the current mean shape, we could use functions such as:

$$((vec(Y^0)-U\lambda)^T(I_2\otimes\Gamma^{11})(vec(Y^0)-U\lambda)^T)+\alpha d_F^2(vec^{-1}(vec(Y^0)-U\lambda),\hat{\mu})$$

$$\text{or}\quad ((vec(Y^0)-U\lambda)^T(I_2\otimes\Gamma^{11})(vec(Y^0)-U\lambda)^T)+\alpha d_F^2(vec^{-1}(vec(Y^0)-U\lambda),Y^0)$$

in order to ensure that the shape of a configuration does not change too much from that of the mean or from itself, when using the original BE criterion. A cross-validation method could also be developed to determine α in some automatic way.

11.5.1.2 Incorporating information on the accuracy of semi-landmark locations

A further idea considered in chapter 6, was how the original BE technique may be considered as particular case of an anisotropic smoothing spline, where landmarks in the reference configuration are mapped to positions defined as bivariate error ellipses in the target configuration. Each semi-landmark's chord may be considered as an ellipse with infinite variance in one direction and zero in the other. In section 6.2.1.4 we showed how penalties could be placed on the extent of movement allowed for each semi-landmark away from its nominal position, by specifying the variance of its associated ellipse in the direction of its chord. The objective function then penalises the BE of the mapping relative to the weighted sum of squared distances moved by the semi-landmarks. For the gingival variation problem, it was found difficult to identify weights that would work equally well on all configurations, however, there could well be other applications where this idea may be useful. One approach could be to specify variances along chords to represent the operator reliability or degree of uncertainty associated with each semi-landmark position.

In fact, the component of the objective function that penalises the movement of semi-landmarks, away from their nominal positions, could instead be used with the FP or NP criterion. For example, if $\sigma_1^2,...,\sigma_{SL}^2$ are a series of weights, we could consider minimising:

$$(U\lambda)^T diag(\sigma_1^2,...,\sigma_{SL}^2)^{-1}(U\lambda)+\alpha\left\|SG(vec^{-1}(vec(Y^0)-U\lambda)-\hat{\mu}\right\|^2$$

to determined the new positions of semi-landmarks. Such a method could be used when we wish to allow the shape of a configuration to change, but not if this involves the more reliable semi-landmarks moving large distances along their chords. Again, the weights $\sigma_1^2,...,\sigma_{SL}^2$ could be pre-specified to reflect the relative certainty with which semi-landmarks are located (in the direction of their chords).

11.5.1.3 Weighted Procrustes criterion

Another way of representing error variance, here at both the fixed and semi-landmark locations, when determining the new positions of the semi-landmarks along their

chords, could be to use a weighted, rather than OLS, Procrustes criterion as described in section 6.4.1.1. For example, an objective function of the form:

$$vec(SG(vec^{-1}(vec(Y^0)-U\lambda)-\hat{\mu})^T\Sigma_S^{-1}vec(SG(vec^{-1}(vec(Y^0)-U\lambda)-\hat{\mu})$$

or $$vec(vec^{-1}(vec(Y^0)-U\lambda)-SG(\hat{\mu}))^T\Sigma_S^{-1}vec(vec^{-1}(vec(Y^0)-U\lambda)-SG(\hat{\mu}))$$

with error covariance matrix Σ_S^{-1}, specified to reflect the certainty with which the fixed and semi-landmarks are identified, and in which directions, could be used to ensure that the superimposition matched the more reliable landmarks preferentially, when determining new locations for the semi-landmarks. For example, one option could be to determine the full Procrustes superimposition to the mean on only the fixed landmarks, since these are those about whose locations we are most certain. The new positions of the semi-landmarks would then effectively be determined by applying the same transformation to semi-landmarks chords and identifying positions which minimize the OLS or GLS from the corresponding landmarks in the mean.

11.5.1.4 Shape features and use of covariates

Several other ideas were also suggested in chapter 6 for other criteria which may be optimised when determining the new positions of semi-landmarks. For example, in some studies it may be useful to move the semi-landmarks to match some feature of the mean (or some reference shape), such as a ratio of inter-landmark distances or angle between a group of landmarks. Another option could be to determine new semi-landmark positions through values of other covariates, known to be related to shape.

11.5.2 Improvements to the semi-landmark methods

11.5.2.1 Better approximation of outlines and curves

In chapter 8, it was noted how the resulting positions of semi-landmarks sometimes did not lie on the actual outline of the object they were meant to represent. This will generally occur when the approximation to a curved set of possible positions by the straight lines of the chords is poor. One way to improve this would be to use the actual strings of pixel coordinates denoting the curve or outline of the object, which could

easily be extracted from a digital image using most modern image analysis packages. Use of the NP, FP or BE criterion would then involve a search along the strings of possible positions, to determine the new locations of the semi-landmarks. An alternative, for when such a detailed representation is not possible, but when plenty of landmarks are available, could be to represent the possible new semi-landmark positions by a polynomial curve, based on the nearest (say) 1, 2 or 3 landmarks either side of its nominal semi-landmark position. However, in this study, the use of such methods to represent possible new positions of gingival semi-landmarks, was not be possible, since complete outline information is nearly always obscured in one direction by gingival tissue. Perhaps one idea could be to fit a polynomial curve, extending beneath the gum, based on the visible part of the crown outline, as we did when using chord directions.

11.5.2.2 Speeding up the FP method

For the FP method, convergence was seen to be much slower than when using the BE or NP criteria, since there is no straightforward analytical solution for determining the new positions of semi-landmarks on each movement step. The investigations here used the 'nlmin' function of S-plus, but one could consider other algorithms or further ways in which this process could be speeded up.

11.5.2.3 Extension of new methods to three dimensions

A logical extension of the semi-landmark methods is to three dimensional (3D) surfaces. One deficiency of most morphometric analyses is that they are nearly always 2D representations of 3D objects, but with laser scanners becoming ever more affordable it is becoming increasingly easier to acquire 3D or pseudo-3D data. GPA algorithms are easily able to cope with data in *m* dimensions, as described in section 3.2 and the FP and NP criteria should be easily be transferable to three-dimensional data problems. However, as Adams *et al.* (2004) note when considering extensions of Bookstein's original method to 3D applications, this requires that the semi-landmarks be able to move in directions approximating surfaces, rather than curves, which is a complex problem. Recent work by Bookstein and co-workers (see Mitterocker & Gunz, 2002), has involved an approach which starts by densely sampling a surface of three dimensional points on each individual object and triangulating these into a

'surface mesh'. On a reference object (e.g. the mean shape), a low resolution mesh is produced to reduce the number of vertices, while maintaining the surface information. Data from each object are then mapped by splines onto the reference and nearest surface points taken as homologous to those of the reduced mesh. These points are then allowed to slide in a plane tangential to the local curvature of the surface when determining their new minimum BE positions and as before the adjustments are repeated until the process converges.

In terms of using the NP method with this approach, determination of new semi-landmark positions would simply involve using the penultimate step described above (following prior GPA registration to the reference object). i.e. finding surface positions or points in a tangent plane, which minimised Euclidean distances to points on the reduced mesh. For the FP criterion, the more difficult task would be modifying the Procrustes technique to allow superimposition of landmarks and points which varied along surfaces rather than along chords. A computer intensive search could be used to identify surface positions which simultaneously minimised the sum of squared distances to points on the reduced mesh of the reference object and between pairs of corresponding landmarks, as the entire object varied over similarity transformations.

11.5.3 An alternative, Bayesian approach to semi-landmark problems

An alternative to using the techniques for inference described in section 3.2.3 or 3.2.4 would be to consider a Bayesian approach. Recall from section 6.4.1.2 that a $k \times m$ configuration X is assumed to be a perturbation from a model with mean μ and covariance matrix Σ, which has then been translated, rotated and re-scaled, by vector $1_k \gamma'^T$, matrix Γ' and scalar β' and observed as $X = \beta' (\mu + E) \Gamma' + 1_k \gamma'^T$, where $vec(E)$ follows a distribution with zero mean and covariance matrix Σ. Assuming that a sample of configurations, $X_1,...,X_n$ is observed without error, each observed X is then a function of μ, Σ, β', Γ' and γ' with some joint probability density and likelihood function $L(X_1,...,X_n | \mu, \Sigma, \gamma', \beta', \Gamma')$. If $\pi(\mu, \Sigma, \gamma', \beta', \Gamma')$ defines a joint prior distribution for μ, Σ, β', Γ' and γ', then by Bayes' rule, a joint posterior density for μ, Σ, β', Γ' and γ' is given by:

$$p(\mu,\Sigma,\gamma',\beta',\Gamma'|X_1,...,X_n) \propto \pi(\mu,\Sigma,\gamma',\beta',\Gamma') \times L(X_1,...,X_n|\mu,\Sigma,\gamma',\beta',\Gamma').$$

For analysis, interest would then lie in the marginal posterior distributions for Σ and μ, the mean and covariance of the $(\mu + E)$, or equivalently of the $\beta X\Gamma + 1_k\gamma^T$ where β, Γ and $1_k\gamma^T$ are the reverse similarity transformations of β', Γ' and γ', as described in section 6.4.1.2. Dryden & Mardia (1998) consider a Bayesian approach to inference in terms of distributions for points in shape or pre-shape space and discuss different choices of prior distribution for the mean and/or covariance parameters.

In order to include semi-landmarks, we would also need to define joint priors for the scalars $\lambda_{(l)}$, l=1,...,L, defining the distance moved by semi-landmarks along each unit chord direction, or alternatively for the error covariance Σ_e associated with the location of each configuration's semi-landmarks. For the latter, this could be done in either the same space in which the biological or actual covariance Σ was defined or in the space of the observed configurations. The former would be difficult unless one was to work in terms of shape and size. For example, the $X_1,...,X_n$ could be assumed to be observed from a Binormal (X_{true}, Σ_e) distribution where X_{true} is a function of μ, Σ, β', Γ' and γ', as defined above and Σ_e determines the chord directions of the semi-landmarks. If $\pi(\mu,\Sigma,\gamma',\beta',\Gamma',\Sigma_e)$ defines a joint prior distribution for μ, Σ, β', Γ', γ' and Σ_e and $L(X_1,...,X_n|\mu,\Sigma,\gamma',\beta',\Gamma',\Sigma_e)$ is the new likelihood of the observed data, the posterior distribution for μ, Σ, β', Γ', γ' and Σ_e would be given by:

$$p(\mu,\Sigma,\gamma',\beta',\Gamma',\Sigma_e|X_1,...,X_n) \propto \pi(\mu,\Sigma,\gamma',\beta',\Gamma',\Sigma_e) \times L(X_1,...,X_n|\mu,\Sigma,\gamma',\beta',\Gamma',\Sigma_e).$$

Equivalently, a prior $\pi(\mu,\Sigma,\gamma',\beta',\Gamma',\lambda)$ for μ, Σ, β', Γ', γ' and $\lambda=(\lambda_{(1)},...,\lambda_{(L)})^T$, could be defined in terms of size and shape or in the space of the observed configurations, along with a likelihood $L(X_1,...,X_n|\mu,\Sigma,\gamma',\beta',\Gamma',\lambda)$ for the observed X and a posterior distribution $p(\mu,\Sigma,\gamma',\beta',\Gamma',\lambda|X_1,...,X_n)$ determined by Bayes rule. Again, in practice, we would only be interested in the marginal distributions of Σ and μ, rather than those of the $\lambda_{(l)}$ (or Σ_e) and other nuisance parameters β', Γ' and γ'.

A key benefit of such an approach would be its potential to formally capture the expert opinion of dentists and periodontal researchers, in a less ad-hoc way than was used in the simulation study of Chapter 8, by means of prior probability distributions. The main difficulty, however, would be the likely problems with computation. Even with the introduction of artificial constraints on the problem, such as the use of convenient parametric forms for prior distributions, the form of the posterior distribution would be intractable and need to be integrated numerically rather than analytically in order to obtain posterior probabilities. This could be done, for example, by using MCMC methods, but would be computationally intensive.

11.5.4 Multivariate reliability measure

In chapter 4, new methods were presented for obtaining an overall reliability measure of different operators' landmark representations of the same set of objects, here teeth of a certain type, from different patients. The consequences of location inconsistencies were evaluated by calculating their effect on the recorded variation in Procrustes fits, quantifying the proportion of variance in shape attributable to systematic operator differences and random errors, relative to the actual variation between cases (patients). In section 4.4.2, we showed how the same figure may be obtained by summing components of variance from the principal component (PC) scores of the Procrustes fits, estimated in the same way as Arqvist & Martensson (1998), using mean squares from 2-way ANOVA partitions, or by the same calculations based on a single 2-way partition of the sum of squared Euclidean distances (approximate sum of squared Procrustes distances) between the fits.

In section 4.4.3, we noted how the variance assumptions of both methods, defined on each (independent) set of PC scores as $\sigma^2_{C_r} + \sigma^2_{O_r} + \sigma^2_{E_r}$, correspond to a covariance matrix Σ for the residuals of the Procrustes fits of the form:

$$\sum_{r=1}^{2k-4} (\sigma^2_{C_r} + \sigma^2_{O_r} + \sigma^2_{E_r}) \psi_r \psi_r^T$$

where $\sigma^2_{C_r}, \sigma^2_{O_r}$ and $\sigma^2_{E_r}$ are the variances in PC scores between cases, operators and due to random errors, along each of the r =1,...,2k-4 principal components, ψ_r of Σ

and k is the number of landmarks on each configuration. (The ψ_r are estimated by the PC's of the sample covariance matrix of the Procrustes fits and $\sigma^2_{C_r}$, $\sigma^2_{O_r}$ and $\sigma^2_{E_r}$ from the sample component scores). The appropriateness of the resulting form for Σ is clearly one area for future exploration. In addition, it may be possible to produce other reliability measures for the Procrustes fits, based on multivariate random effects models, rather than univariate models along each PC.

11.5.5 Other applications and accessibility

Of course, there will be many other situations other than for the study of tooth shape where these new techniques may be beneficial, both in other areas Dentistry and in applications within a wider variety of different fields and disciplines. Some of the possible uses that have been suggested and discussed with other researchers have included; comparing facial profiles before and after surgery, where for example, landmarks around the mandibular jaw-line can be difficult to locate precisely, investigation of edge profiles of forensic bite marks from upper and lower anterior teeth and applications to fossil bone remains, which have become broken or altered in some way, making identification of homologous landmarks difficult.

As Adams et al. (2004) note, for landmark data, the standard protocol based on Procrustes methods is now widely accepted and, by using Bookstein's (1996a,d,e) semi-landmark method, outlines or curves may also be analysed using these methods. What the new techniques presented here offer is an improved, more robust methodology for more general situations where use of the original BE approach may produce undesirable results and where assigning points of correspondence by Euclidean distances seems more appropriate than minimising the BE to a consensus shape. Future work would be to make the 'semi.it.new' S-plus routine, detailed in section 7.4 (and its associated subroutines) more accessible and user friendly. For those familiar with scrip files in S-plus, one option would be to make these and details of user supplied arguments available on-line, in a manner similar to Dryden's 'shapes' package of routines. Further work could be to reproduce the routines in a windows based interface, similar Rohlf's TPS series of programs, with drop-down menus for choosing the different options and the ability to define chord directions interactively, rather than as matrices.

References

Adams, D. C., Rohlf, F. J., Slice, D. E. (2004). Geometric morphometrics: Ten years of progress following the 'revolution'. *Italian Journal of Zoology*. 71:5-16.

Alt, K. W. and Turp, J. C. (1998). Hereditary Dental Anomalies. In Alt, K.W., Rosing FW, and Teschler-Nicola, M., editors, *Dental Anthropology: Fundamentals, Limits and Prospect,* pages 95-128. Springer, New York.

Alvesalo, L. (1971). The influence of sex chromosome genes on tooth size in man. *Suomen Hammaslaakariseunan Toimituksia*, 67: 3-54.

Alvesalo, L. and Tigerstedt, P. M. (1974). Heritabilities of human tooth dimensions. *Hereditas*, 77(2):311-318.

Andresen, P. R., Bookstein, F. L., Conradsen, K., Ersboll, B., Marsh, J. and Kreiborg, S. (2000). Surface-bounded growth modelling applied to human mandibles. *IEEE Transactions on Medical Imaging*. 19:1053-1063.

Arnqvist, G. and Martensson, T. (1998). Measurement error in geometric morphometrics: Empirical strategies to assess and reduce its impact on measures of shape. In Klingenberg, C. P. and Bookstein, F. L., editors, *Acta Zoologica Academiae Scientiarum Hungaricae*, Vol. 44(1/2), pages 73-96. Hungarian Natural History Museum, Budapest.

Bahti, S. N. and Harrison, V. E. (1987). Operational performance of the travelling microscope in the measurement of dental casts. *British Journal of Orthodontics*, 14(3): 147-153.

Bailey, R. C. and Byrnes, J. (1990). A new, old method for assessing measurement error in both univariate and multivariate morphometric studies. *Systematic Zoology*, 39:124-130.

Ballard, M.L. (1944). Asymmetry in tooth size: a factor in the aetiology, diagnosis and treatment of malocclusion. *Angle Orthodontist*, 14: 67-71.

Barrett, M. J., Brown, T. and Macdonald, M. R. (1963). Dental observations on Australian aborigines: mesio-distal crown diameters of permanent teeth. *Australian Dental Journal*, 8: 150-155

Bialit, H. L. (1975). Dental variation among populations: An anthropologic view. *Dental Clinics of North America*, 19: 125-139.

Biggerstaff, R. H. (1969). Electronic methods for the analysis of the human post-canine dentition. *American Journal of Physical Anthropology*, 31(2):235-242.

Besl, P. J. and McKay, N. D. (1992). A method for registration of 3D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14:239-256.

Bland, J. M. and Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet*, 1:307-310.

Boas, F. (1905). The horizontal plane of the skull and the general problem of the comparision of variable forms. *Science*, 21:862-863.

Bodner, L., Bar-Ziv, J., and Becker, A. (2001). Image accuracy of plain film radiography and computerized tomography in assessing morphological abnormality of impacted teeth. *American Journal of Dentofacial Orthopedics*, 120(6): 623-628.

Bookstein, F. L. (1978). *The measurement of biological shape and shape change*. Springer-Verlag, Berlin.

Bookstein, F. L. (1984). A statistical method for biological shape comparisons. *Journal of Theoretical Biology*, 107:475-520.

Bookstein, F. L. (1986). Size and shape spaces for landmark data in two dimensions (with discussion). *Statistical Science*, 1:181-242.

Bookstein, F. L. (1989). Principal warps: thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11:567-585.

Bookstein, F. L. (1991). *Morphometric Tools for Landmark Data: Geometry and Biology*. Cambridge University Press.

Bookstein, F. L. (1996a). Applying landmark methods to biological outline data. In Mardia, V K. V., Gill, C. A., and Dryden, I. L., editors, *Proceedings in Image Fusion and Shape Variability Techniques*, pages 59-70. University of Leeds Press.

Bookstein, F. L. (1996b). Biometrics, biomathematics and the morphometric synthesis. *Bulletin of Mathematical Biology*, 58:313-365.

Booksetein, F. L. (1996c). Combining the tools of Geometric Morphometrics. In Marcus, L. F., Corti, M., Loy, A., Naylor, G. J. P. and Slice, D. E., editors, *Advances in Morphometrics*, pages 103-112. NATO ASI Series A: Life Sciences, 284. Plenum: New York.

Bookstein, F. L. (1996d). Landmark methods for forms without landmarks: morphometrics of group differences in outline shape. *Medical Image Analysis*, 1:225-243.

Bookstein, F. L. (1996e). Visualizing group differences in outline shape: Methods from Biometrics of landmark points. *Lecture Notes in Computer Science*, 1131:405-410.

Bookstein, F. L. (1997). Shape and the information in medical images: a decade of the morphometric synthesis. *Computer Vision and Image Understanding*, 66:97-118.

Bookstein, F. L. (1998). A hundred years of morphometrics. In Klingenberg, C. P. and Bookstein, F. L., editors, *Acta Zoologica Academiae Scientiarum Hungaricae*, Vol. 44(1/2), pages 73-96. Hungarian Natural History Museum, Budapest.

Bookstein, F. L. (1999). Linear methods for nonlinear maps: Procrustes fits, thin-plate splines and the biometric analysis of shape variability. In Toga, A. W., editor, *Brain Warping*, Academic Press.

Bookstein, F. L., Chernoff, B., Elder, R., Humphries, J., Smith, G., and Strauss, R. (1985). *Morphometrics in Evolutionary Biology*. Special Publication No. 15, Academy of Natural Sciences: Philadelphia.

Bookstein, F. L. and Sampson, P. D. (1990). Statistical models for geometric components of shape change. *Communications in Statistics, Theory and Methods*,19:1939-1972.

Bookstein, F. L., Sampson, P. D., Streissguth, A. P., and Connor, P. D. (2001). Geometric morphometrics of corpus callosum and subcortical structures in the fetal-alcohol-affected brain. *Teratology*, 64:4-32.

Bookstein, F. L., Schaefer, K., Possinger, H., Seidler, H., Fieder, M., Stringer, C., Weber, G., Arsuga, J., Slice, D,. Rohlf, F. J., Recheis, W., Mariam, A., and Marcus, L. (1999) Comparing front cranial profiles in archaic and modern Homo by morphometric analysis. *Anatomical Record - The New Anatomist*, 257:217-224.

Bookstein, F. L., Streissguth, A. P., Sampson, P. D., Connor, P. D., and Barr, H. (2000a). Corpus callosum shape and neuropsychological deficits in adult males with heavy fetal alcohol exposure. *Neuroimage*, 15:233-251.

Bookstein, F. L.,Sampson, P. D., Connor, P. D. and Streissguth, A. P. (2002b). The midline corpus callosum is a neuroanatomical focus of fetal alcohol exposure. *Anatomical Record – The New Anatomist*, 269:162-174.

Brook, A. H. (1984). A unifying aetiological explanation for anomalies of human tooth number and size. *Archives of Oral Biology*, 29: 373-378.

Brook, A.H., editor (2001). *Dental Morphology 2001: 12th International Symposium on Dental Morphology, Sheffield, UK*. Sheffield Academic Press.

Brook, A. H., Pitts, N. B., and Renson, C. E. (1983). Determination of tooth dimensions from study casts using an image analysis system. *Journal of the International Association of Dentistry for Children.* 14:55-60

Brook, A. H., Smith, R. N., Elcock, C., Al-Sharood, M. H., Shah, A. A. and Karmo, M. (1998). The measurement of tooth morphology: development and validation of a new image analysis system. In Mayhall, J.T. and Heikkinen, T., editors, *Dental Morphology 1998: Proceedings of the 11th International Symposium on Dental Morphology, Oulu, Finland,* pages 380-387. Oulu University Press.

Carbonell, V. M. (1963). Variations in the frequency of shovel-shaped incisors in different populations. In Brothwell, D. R., editor, *Dental Anthropology*, pages 211-234. Pergamon Press, London.

Carpenter, K. E., Sommer, J. V., Marcus, L. F. (1996). Converting truss inter-landmark distances to Cartesian coordinates. In Marcus, L. F., Corti, M., Loy, A., Naylor, G. J. P. and Slice, D. E., editors, *Advances in Morphometrics*, pages 103-112, NATO ASI Series A: Life Sciences, 284. Plenum: New York.

Cheverud, J. M. and Richtsmeier, J. T. (1986). Finite element scaling applied to sexual dimorphism in rhesus macaque (macaca mulatta) facial growth. *Systematic Zoology*, 35:381-399.

Cole, T. M. (1996). Historical note: Early anthropological contributions to "geometric morphometrics". *American Journal of Physical Anthropology*, 101:291-296.

Cooley, J. W. and Tukey, O. W. (1965). An algorithm for the machine calculation of Complex Fourier Series. *Mathematical Computing*, 19:297-301.

Cootes, T. F., Taylor, C. J., Cooper, D. H., and Graham, J. (1992). Training models of shape from sets of examples. In Hogg, D. C. and Boyle, R. D., editors, *British Machine Vision Conference*, pages 9-18. Berlin: Springer-Verlag.

Cressie, N. A. C. (1993). *Statistics for Spatial Data*. Wiley, New York.

Crimmins, T. R. (1982). A complete set of Fourier descriptors for two-dimensional shapes. *IEEE Transactions on Systems, Man and Cybernetics*, 12:848:855.

Dahlberg, A. A. (1945). The changing dentition of man. *Journal of the American Dental Association*, 32: 676-690.

Dahlberg, A. A. (1949). The dentition of the American Indians. In Laughlin, W. S., editor, *Papers on the Physical Anthropology of the American Indians*, pages 138-176. The Viking Fund, Inc.

Dempsey, P. J., Townsend, G. C., Martin, N. G., and Neale, M. C. (1995). Genetic covariance structure of incisor crown size in twins. *Journal of Dental Research*, 74(7):1389-1398.

Dennis, J. E., Gay, D. M., and Welsch, R. E. (1981), An adaptive nonlinear least-squares algorithm. *ACM Transactions on Mathematical Software*, 7:348-368.

Dennis, J. E. and Mei, H. H. W. (1979). Two new unconstrained optimization algorithms which use function and gradient values. *Journal of Optimization Theory and Applications*, 28:453-483.

Donner, A. and Eliasziw. M. (1987). Sample size requirements for reliability studies. *Statistics in Medicine*, 6:441-448.

Doris, J. M., Bernard, B. W., and Kuftinec, M. M. (1981). A biometric study of tooth size and dental crowding. *American Journal of Orthodontics*, 79:326-336.

Dryden, I. L. and Mardia, K. V. (1991). General shape distributions in a plane. *Advances in Applied Probability*, 23:259-276.

Dryden, I. L. and Mardia, K. V. (1993). Multivariate shape analysis. *Sankhya Series A*, 55:460-480.

Dryden, I. L. and Mardia, K. V. (1998). *Statistical Shape Analysis*. Wiley, Chichester.

Dryden, I. L. and Walker, G. (1997). Shape analysis using highly robust regression. In *Proceedings of the ISI 51st Session*, Vol. 2, pages 127-130, Istanbul, Turkey. ISI.

Dryden, I. L. and Walker, G. (1999). Highly resistant regression and object matching. *Biometrics*, 55:820-825.

Duchon, J. (1976). Interpolation des fonctions de deux variables suivant la principe de la flexion des plaques minces. *RAIRO Analyse Numerique*, 10:5-12.

Enciso, R., Lewis, J. P., Neuman, U., and Mah, J. (2003). 3D tooth shape from radiographs using thin plate splines. *Studies in Health Technology and Informatics*, 94:62-64.

Fearne, J. M. and Brook, A. H. (1993). Small primary tooth-crown size in low birthweight children. *Early Human Development*, 33(2):81-90.

Ferrario, V. F., Sforza, C., tartaglia, C. M., Colombo, A., and Serrao, G. (1999). Size and shape of the human first permanent molar: A Fourier analysis of the occlusal and equatorial outlines. *American Journal of Physical Anthropology*, 102:281-294.

Ferson, S. F., Rohlf, F. J. and Koehn, R. K. (1985). Measuring shape variation of two-dimensional outlines. *Systematic Zoology*, 34:59-68.

Fitzpatrick, J. M., West, J. B., and Maurer, C. R. (1998). Predicting error in rigid-body, point-based image registration. IEEE Transactions in Medical Imaging, 17:694-702.

Fliess, J. L. (1986). *The Design and Analysis of Clinical Experiments*. Wiley, Chichester.

Fliess, J. L. and Shrout, P. E. (1977), The effects of measurement errors on some multivariate procedures. *American Journal of Public Health*, 67:1188-1191.

Garn, S.M., Lewis, A.B. and Kerewsky, R.S. (1965). Genetic, nutritional, and maturational correlates of dental development. *Journal of Dental Research*, 44: 228-242.

Garn, S. M., Dahlberg, A. A., Lewis, A. B., and Kerewsky, R. S. (1966a). Groove pattern, cusp number and tooth size. *Journal of Dental Research*, 45(3): 970.

Garn, S. M., Lewis, A. B., and Kerewsky, R. S. (1966b). Bilateral asymmetry and concordance in cusp number and crown morphology of the mandibular first molar. *Journal of Dental Research*, 45: 1820.

Garn, S. M., Lewis, A. B., and Kerewsky, R. S. (1967a). Sex difference in tooth shape. *Journal of Dental Research*, 46: 1470.

Garn, S. M., Lewis, A. B., Swindler, D. R., and Kerewsky, R. S. (1967b). Genetic control of sexual dimorphism in tooth size. *Journal of Dental Research*, Supplement No. 5, 46: 963-972.

Garn, S. M., Osborne, R. H., McCabe, K. D. (1979). The effect of prenatal factors on crown dimensions. *American Journal of Physical Anthropology*, 51(4):665-678.

Glasbey, C. A. and Horgan, G. W. (1995). *Image analysis for the Biological Sciences*. Wiley, Chichester.

Glaseby, C. A., McRae, D. C. and Fleming, J. (1988). The size distribution of potato tubers and its application to grading schemes. *Annals of Applied Biology*, 113:579-587.

Goodall, C. R. (1991). Procrustes methods in the statistical analysis of shape (with discussion). *Journal of the Royal Statistical Society, Series B*, 53:285-339.

Goodall, C. R. (1995). Procrustes methods in the statistical analysis of shape revisited. In Mardia, K. V. and Gill, C. A., editors, *Current Issues in Statistical Shape Analysis*, pages 18-33. University of Leeds Press.

Goodall, C. R. and Green, P. B. (1986). Quantitative analysis of surface growth. *Botanical Gazette*, 147:1-15.

Goodall, C. R. and Mardia, K. V. (1993). Multivariate aspects of shape theory. *Annals of Statistics*, 21:848-866.

Goose, D. H. (1963). Dental measurement: An assessment of its value in anthropological studies. In Brothwell, D. R., editor, *Dental anthropology*, pages 125-148. Pergamon Press, London.

Gower, J. C. (1975). Generalized Procrustes analysis. *Psychometrika*, 40:33-50.

Granlund, G. H. (1972). Fourier pre-processing for hand print character recognition. *IEEE Transactions in Computing*. C-21(2):195-201.

Green, W. D. K. (1995). Spline-based deformable models. In Melter, R. A., Wu, A. Y., Bookstein, F. L., and Green, W. D. K., editors, *Vision Geometry IV, vol. 2573 of SPIE Proceedings*, pages 290-301. SPIE.

Green, W. D. K. (1996). The thin-plate spline and images with curving features. In Mardia. K. V., Gill, C. A., and Dryden, I. L., editors, *Proceedings in Image Fusion and Shape Variability*, pages 79-87. University of Leeds Press.

Green, P. J. and Silverman, B. W. (1994). *Non-parametric Regression and Generalized Linear Models: A Roughness Penalty Approach*. Chapman and Hall, London.

Haar, A. (1910). Zur Theorie der orthogonalen Funktionensysteme. *Annals of Mathematics*, 69:331-371.

Heasman, P. (1997). *Periodontology (Colour Guides)*. Churchill Livingstone.

Hill, A, and Taylor, C. J. (1992). Model-based images interpretation using genetic algorithms. *Image and Vision Computing*. 10:295-300.

Hillson, S. (1996). *Dental Anthropology*. Cambridge University Press.

Horgan, G. W., Creasey, A., and Fenton, B. (1992). Superimposing two-dimensional gels to study genetic variation in malaria parasites. *Electrophoresis*, 13:871-875.

Hu, M. K. (1962). Visual pattern recognition by invariant moments. *IRE Transactions on Information Theory*, 8:179-187.

Hurley, J. R. and Cattell, R. B. (1962). The Procrustes program: producing direct rotation to test a hypothesised factor structure. *Behavioural Science*, 7:258-262.

Huxley, J. S. (1924). Constant differential growth ratios and their significance. *Nature*, 114:895-896.

Huxley, J. S. (1932). *Problems of Relative Growth*. Methuen, London.

Johnson, D. R., O'Higgins, P., McAndrew, T. J., Adams, L. M., and Flinn, R. M. (1985). Measurement of biological shape: a general method applied to mouse vertebrae. *Journal of Embryology and Experimental Morphology*, 90:363-377.

Kaesler, R. L. and Waters, J. A. (1972). Fourier analysis of the ostracod margin. *Bulletin of the Geological Society of America*, 83:1169-1178.

Keene, H. J. (1966). Relationship between maternal age and parity, birth weight and hypodontia in navel recruits. *Journal of Dentistry for Children*, 33: 135-147.

Keiser, J. A. (1990). *Human Adult Ondontometrics*. Cambridge University Press.

Kendall, D. G. (1977). The diffusion of shape. *Advances in Applied Probability*, 9:428-430.

Kendall, D. G. (1984). Shape manifolds, Procrustean metrics and complex projective spaces. *Bulletin of the London Mathematical Society*, 16:81-121.

Kendall, D. G., Barden, D., Carne, T. K., and Le, H. L. (1999). *Shape and Shape Theory*. Wiley, Chichester.

Kent, J. T. (1994). The complex Bingham distribution and shape analysis. *Journal of the Royal Statistical Society, Series B*, 56:285-299.

Kent, J. T. and Mardia, K. V. (1994). The link between kriging and thin-plate splines. In Kelly, F. P., editor, *Probability, Statistics and Optimization: a Tribute to Peter Whittle*, pages 325-339. Wiley, Chichester.

Khalaf, K., Elcock, C., Robinson, D. L., Brook, A. H., and Smith, R. N. (2001). Tooth size and shape in patients with supernumerary teeth. *Journal of Dental Research*, 79(5):286.

Klingenberg, C. P. and Bookstein, F. L., editors (1998). *Acta Zoologica Academiae Scientiarum Hungaricae*, 44(1/2). Hungarian Natural History Museum, Budapest.

Kristof, W. and Wingersky, B. (1971). Generalization of the orthogonal Procrustes rotation procedure to more than two matrices. In *Proceedings of the 79th Annual Convention of the American Psychological Association*, pages 89-90.

Kuhl, F. P. and Giardina, C. R. (1982). Elliptic Fourier features of a closed contour. *Computer Graphics and Image Processing*, 18:236-258.

Kuroda, T., Motohashi, N., Tominaga, R., and Iwata, K. (1996). Three-dimensional dental casts analyzing system using laser scanning. *American Journal of Orthodontics & Dentofacial Orthopedics*, 110(4): 365-369.

Lamb, D. J. (1988). *Appearance and Aesthetics in Denture Practice*. Bristol Wright.

Lavelle, C. L. (1968). Anglo-Saxon and modern British teeth. *Journal of Dental Research*, 47: 811-815.

Lavelle, C. L. (1970). Comparison of the deciduous teeth between Caucasoid, Negroid and Mongoloid population samples. *Dental Practitioner*, 21(4): 121-124.

Lavelle, C. L. (1972). Maxillary and mandibular tooth size in different racial groups and in different occlusal categories. *American Journal of Orthodontics*, 61(1):29-37.

Lavelle, C. L., Ashton, E. H. and Flinn, R. M. (1970). Cusp pattern, tooth size and third molar agenesis in the human mandibular dentition. *Archives of Oral Biology*, 15:227-37.

Lele, S. R. (1993). Euclidean distance matrix analysis (EDMA): estimation of mean form and mean form difference. *Mathematical Geology*, 25:573-602.

Lele, S. R. and Cole, T. M. (1995). Euclidean distance matrix analysis: a statistical review. In Mardia, K. V. and Gill, C. A., editors, *Current Issues in Statistical Shape Analysis*, pages 49-53. University of Leeds Press.

Lele, S. R. and Cole, T. M. (1996). A new test for shape differences when variance-covariance matrices are unequal. *Journal of Human Evolution*. 31:193:212.

Lele, S. R. and Richtsmeier, J. T. (1991). Euclidean distance matrix analysis: a coordinate-free approach for comparing biological shapes using landmark data. *American Journal of Physical Anthropology*, 86:415-427.

Lele, S. R. and Richtsmeier, J. T. (2001). *An Invariant Approach to Statistical Analysis of Shapes*. Chapman & Hall, London.

Lestrel, P.E. (1974). Some problems in the assessment of morphological size and shape differences. *Yearbook of Physical Anthropology*, 18:140-162.

Lestrel, P. E. (1997). *Fourier Descriptors and their Applications in Biology*. Cambridge University Press.

Lin, C. C. and Challappa, R. (1987). Classification of partial 2-D shapes using Fourier descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(5):686-690.

Lindhe, J. (1990). *Textbook of Clinical Periodontology*. Munksgaard, Denmark.

Lohmann, G. P. (1983). Eigenshape analysis of microfossils: A general morphometric procedure for describing changes in shape. *Mathematical Geology*, 15:659-672.

Lougheed, S. C., Arnold, T. W., and Bailey, R. C. (1991). Measurement error of external and skeletal variables in birds and its effect on principal components. *Auk*, 108:432-436.

McKeown, H. F., Robinson, D. L., Elcock, C., Al-Sharood, M. H., and Brook, A. H. (2002). Tooth dimensions in hypodontia patients, their unaffected relatives and a control group measured by a new image analysis system. *European Journal of Orthodontics*, 24: 131-141.

Marcus, L.F. (1990). Traditional morphometrics. In Rohlf, F. J. and Bookstein, F.L., editors, *Proceedings of the Michigan Morphometrics Workshop*, pages 77-122. Special Publication No. 2, University of Michigan Museum of Zoology, Ann Arbor.

Marcus, L. F. (1996). Introduction to data acquisition. In Marcus, L. F., Corti, M., Loy, A., Naylor, G. J. P. and Slice, D. E., editors, *Advances in Morphometrics*, pages 1-13, NATO ASI Series A: Life Sciences, 284. Plenum: New York.

Marcus, L. F. and Corti, M. (1996). Overview of the new, or geometric morphometrics. In Marcus, L. F., Corti, M., Loy, A., Naylor, G. J. P. and Slice, D. E., editors, *Advances in Morphometrics*, pages 1-13, NATO ASI Series A: Life Sciences, 284. Plenum: New York.

Marcus, L. F., Bello, E., and García-Valdecasas, A., editors (1993). *Contributions to Morphometrics*. Museo Nacional de Ciencias Naturales Monografias, Madrid.

Marcus, L. F., Corti, M., Loy, A., Naylor, G. J. P., and Slice, D. E., editors (1996). *Advances in Morphometrics*. NATO ASI Series A: Life Sciences, 284. Plenum: New York.

Mardia, K. V. and Dryden, I. L. (1989a). Shape distributions for landmark data. *Advances in Applied Probability*, 21:742-755.

Mardia, K. V. and Dryden, I. L. (1989b). The statistical analysis of shape data. *Biometrika*, 76:71-282.

Mardia, K. V. and Dryden, I. L. (1998). The complex Watson distribution and shape analysis. Technical report, Department of Statistics, University of Leeds.

Mardia, K. V. and Gill, C. A., editors (1995). *Current Issues in Statistical Shape Analysis*. Proceedings of the Leeds Annual Statistics Research Workshop, Leeds. University of Leeds Press.

Mardia, K. V., Gill, C. A., and Dryden, I. L., editors (1996a). *Image Fusion and Shape Variability: Proceedings of the Leeds Annual Statistics Research Workshop*. University of Leeds Press.

Mardia, K. V., Kent, J. T., Goodall, C. R., and Little, J. L. (1996b). Kriging and splines with derivative informations. *Biometrika*, 83:207-221.

Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979). *Multivariate Analysis*. Academic Press, London.

Mardia, K. V., Rabe, S., and Kent,J. T. (1995). Statistics, shape and images. In *IMA Proceedings on Complex Stochastic Systems and Engineering Applications*, pages 85-103, Oxford. Clarendon Press.

Marshall, S. (1989). Review of shape coding techniques. *Image and Vision Computing*, 7:281-294.

Matheron, G. (1981). Splines and kriging: their formal equivalence. In Merriam, D. F., editor, *Down-to-Earth Statistics: Solutions looking for Geological Problems*, pages 77-95. Syracuse University Geology Contributions.

Mayhall, J.T. and Heikkinen, T., editors (1998). *Dental Morphology 1998: Proceedings of the 11th International Symposium on Dental Morphology, Oulu, Finland.* Oulu University Press.

Meinguet, J. (1979). Multivariate interpolation at arbitrary points made simple. *Zeitschrift Angewandte Mathematik und Physik*, 30:292-304.

Miethke, R. R. (1972). *Zhanbreiten und Zahnbreitenkorrelationen*, Berlin.

Mitterrocker, P. and Gunz, P. (2002). Semi-landmarks on curves and surfaces in three dimensions. *American Journal of Physical Anthropology, Supplement*, 34:114-115.

Molnar, S (1971). Human tooth wear, tooth function and cultural variability. *American Journal of Physical Anthropology.* 34(2):175-189.

Moorrees, C. F. A., Thomsen, S. O., Jensen, E., and Yen, P. K. (1957). Mesio-distal crown diameters of the deciduous and permanent teeth in individuals. *Journal of Dental Research*, 36: 39-47.

Moorrees, C. F. A. and Reed, R.B. (1964) Correlations among crown diameters of human teeth. *Archives of Oral Biology.* 115:685-697.

Mosier, C. I. (1939). Determining a simple structure when loadings for certain tests are known. *Psychometrika*, 4:149-162.

Mosimann, J. E. (1970). Size allometry: Size and shape variables with characterizations of the lognormal and generalized gamma distributions. *Journal of the American Statistical Association*, 65:930-948.

Moskona, D., Vainder, M., Hershkovitz, I. and Kobyliansky, E. (1997). Dental morphology (non-metric) traits in human isolates – South Sinai Bedouin tribes. *Homo*, 48:227-284.

Nason, G.P. and Silverman, B.W. (2000). Wavelets for regression and other statistical problems. In Schimek, M. G., editor, *Smoothing and Regression: Approaches, Computation and Application,* Wiley, Chichester.

Osbourne, R. H., Horowitz, S. L., and DeGeorge, F. V. (1958). Genetic variation of tooth dimensions: A twin study of the permanent anterior teeth. *American Journal of Human Genetics*, 10: 350-356.

Pavlidis, T. (1977). Polygonal Approximations by Newton's method, *IEEE Transactions on Computers,* 26:800-807.

Peck, H. and Peck, S. (1972). An index for assessing tooth shape deviations as applied to the mandibular incisors. *American Journal of Orthodontics*, 61(4): 384-401.

Pedersen P. O., Dahlberg, A. A., and Alexandersen, V. (1967). International Symposium on Dental Morphology, Fredensborg, Denmark, September 27-29, 1965. *Journal of Dental Research*, 46:769-992.

Peretz, B. and Smith, P. (1993). Morphometric variables of developing primary mandibular second molars. *Archives of Oral Biology*, 38:745–749.

Peretz B., Shapira, J., Farbstein, H., Arieli, E. and Smith P,. (1998). Modified cuspal relationships of mandibular molar teeth in children with Down's syndrome. *Journal of Anatomy*, 193(4):529-33.

Powell, M. J. D. (1987). Radial basis functions for multivariate interpolation: a review. In Mason, J. C. and Cox, M. G., editors, *Algorithms for Approximations*, pages 143-167, Oxford, Clarendon Press.

Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (1992). *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd edition.* Cambridge University Press.

Radlinski, R. J. and Renz, H., editors (1995). *Proceedings of the 10th International Symposium on Dental Morphology, Berlin.* "M" Marketing Services, Berlin.

Ralston, A. (1965). *A First Course in Numerical Analysis*. McGraw Hill, New York.

Rao, C. R. and Suryawanshi, S. (1996). Statistical analysis of shape of objects based on landmark data. *Proceedings of the National Academy of Science USA*, 93:12132-12136.

Reig, S. (1998). 3D digitising precision and sources of error in the geometric analysis of weasel skulls. . In Klingenberg, C. P. and Bookstein, F. L., editors, *Acta Zoologica Academiae Scientiarum Hungaricae*, Vol. 44(1/2), pages 61-72. Hungarian Natural History Museum, Budapest.

Reinhardt, G. A. (1983). Attrition and the edge-to-edge bite: An anthropological study. *Angle Orthodontist*. 53(2):157-164.

Rieger, S. (1993). *Untersuchungen zur Geschlechtsbestimmung merowinderzeitlicher Individuen aus Baden-Wurttemberg anhand der Zahnmabe*. Med Diss, Freiburg.

Remane, A. (1930). Zur Mesztechnik der Primatenzahne. In Abderhanden, E., editor, *Hanbuch der biologischen Arbeitsmethoden*, Vol. 7, pages 609-635. Urban & Schwarzenberg, Berlin.

Renaud, S. (1995). Shape analysis of outlines through Fourier methods using curvature measures. In Mardia, K. V. and Gill, C. A., editors, *Current Issues in Statistical Shape Analysis*, pages 181-182. University of Leeds Press.

Reyment, R. A., Blackith, R. E., and Campbell, N. A. (1984). *Multivariate Moryhometrics*, 2nd. Edition. Academic Press, New York.

Richmond, S. (1987). Recording the dental cast in three dimensions. *American Journal of Orthodontics & Dentofacial Orthopedics*, 92: 199-206.

Robinson, D. L., Blackwell, P. G., Stillman, E. C. and Brook, A. H. (2001). Planar Procrustes Analysis of Tooth Shape. *Archives of Oral Biology*, 46(3):191-199.

Robinson, D. L., Blackwell, P. G., Stillman, E. C. and Brook, A. H. (2002). Impact of landmark reliability on the planar Procrustes analysis of tooth shape. *Archives of Oral Biology*, 47(1):545-554.

Rohlf, F.J. (1996). Introduction to outlines. In Marcus, L. F., Corti, M., Loy, A., Naylor, G. J. P. and Slice, D. E., editors (1996). *Advances in Morphometrics*. pages 209-210, NATO ASI Series A: Life Sciences, 284. Plenum: New York.

Rohlf, F. J. (1999). Shape statistics: Procrustes superimpositions and tangent spaces. *Journal of Classification*. 16:197-223.

Rohlf, F. J. (2000a). Statistical power comparisons among alternative morphometric methods. *American Journal of Physical Anthropology*. 111:463:478.

Rohlf, F. J. (2000b). One the use of shape spaces to compare morphometric methods. *Hystrix*. 11:9-25.

Rohlf, F. J. and Archie, J. (1984). A comparison of Fourier methods for the description of wing shape in mosquitoes. *Systematic Zoology*, 33:302-317.

Rohlf, F. J. and Bookstein, F.L., editors (1990). *Proceedings of the Michigan Morphometrics Workshop*. Special Publication No. 2, University of Michigan Museum of Zoology, Ann Arbor.

Rohlf, F. J. and Slice, D. (1990). Extensions of the Procrustes method for the optimal superimposition of landmarks. *Systematic Zoology*, 39(1):40-59.

Rohr, K. (1998). Image registration based on thin-plate splines and local estimates of anisotropic landmark localization uncertainties. In Wells, W. M., Colchester, A., and Delp, S., editors, *Proceedings of the International Conference on Medical Imaging and Computer Assisted Intervention (MICCAI'98), Cambridge, MA, USA. Lecture notes in Computer Science*, 1496:1174-1183. Springer: Berlin.

Rohr, K., Fornefett, M., and Stiehl, H. S. (1999). Approximating thin-plate splines for elastic registration: Integration of landmark errors and orientation attributes. In Kuba, A., Samal, M., and Todd-Pokropek, A., editors, *Lecture Notes in Computer Science*, 1613:252-267. Springer: Berlin.

Rohr, K., Sprengel, R., Stiehl, H. S. (1997). Incorporation of landmark error ellipsoids for image registration based on approximating smoothing splines. In Lemke, H. U., Vannier, M. V., and Inamura, k., editors, *Proceedings in Computer Assisted Radiology and Surgery (CAR '97), Berlin, Germany*, pages 234-239. Elsevier: Amsterdam.

Rohr, K., Stiehl, H. S., Sprengel, R., Beil, W., Buzug, T. M., Weese, J., and Kuhn, M. H. (1996). Poinr based elastic registration of medical image data using approximating thin-plate splines. In Hohne, K. H. and Kikinis, R., editors, *Proceedings of the 4th International Conference on Visualization in Biomedical Computing (VCB '96), Hamburg, Germany. Lecture notes in Computer Science* 1131:297:306.

Rousseeuw, P. J. (1984). Least median of squares regression. *Journal of the American Statistical Association,* 79:871-880.

Rousseeuw, P. J. and Yohai, V. J. (1984). Robust regression by means of S-estimators. In Franke, J., Hardle, W., and Martin, R. D., editors, *Robust and Non-linear Time Series Analysis*, pages 256-272, New York. Springer-Verlag.

Sampson, P. D., Bookstein, F. L., Sheehan, F. H. and Bolston, E. L. (1996). Eiegenshape analysis of left ventricular outlines from contrast ventriculograms. In Marcus, L. F., Corti, A., Loy, A., Naylor, G. J. P., and Slice, D., editors, *Advances in Morphometrics*, pages 211-233, New York. NATO ASI Series A, Vol. 284, Plenum.

Schoenberg, I. J. (1946) Contributions to the problem of approximation of equidistant data by analytic functions, *Quarterly of Appied Mathematics, 4:45–99.*

Seber, G. A. F. (1977). *Linear Regression Analysis*. Wiley, New York.

Shah, A. A., Elcock, C., and Brook, A. H. (2003). Incisor crown shape and crowding. *American Journal of Othodontics and Dentofacial Orphopedics*. 123(5):562-7.

Siegel, A. F. and Benson, R. H. (1982). A robust comparison of biological shapes. *Biometrics*, 38:341-350.

Slice, D. E. (2001). Landmark coordinates aligned by Procrustes analysis do not lie in Kendall's shape space. *Systematic Biology*, 50:141-149.

Slice, D. E., Bookstein, F. L., Marcus, L. F., and Rohlf, F. J. (1996). A glossary for geometric morphometrics. In Marcus, L. F., Corti, A., Loy, A., Naylor, G. J. P., and Slice, D., editors, *Advances in Morphometrics*, pages 531-551, New York. NATO ASI Series A, Vol. 284, Plenum.

Small, C. G. (1996). *The Statistical Theory of Shape*. Springer, New York.

Sneath, P. H. A. (1967). Trend surface analysis of transformation grids. *Journal of Zoology*, London, 151:65-122.

Strang, G. (1993). Wavelet transforms versus Fourier transforms. *Bulletin of the. American Mathematical Society.* 28:288-305.

Taverne, P. P., Amesz-Voorhoeve, W. H., Leertouwer, H. L. (1979). A photogrammetric method which provides geometric data in dental morphology. *Zeitschrift fur Morphologie und Anthropologie.* 70(2):163-173.

Ten Berge, J. M. F. (1977). Orthogonal Procrustes rotation for two or more matrices. *Psychometrika*, 42:267-276.

Townsend, G. C. (1980). Heritability of deciduous tooth size in Australian aboriginals. *American journal of physical Anthropology*, 50: 297-300.

Townsend, G. C. (1983). Tooth size in children and young adults with Trisomy 21 (Down) Syndrome. *Archives of Oral Biology*, 28(2): 159-166.

Townsend, G. C. and Alvesalo, L. (1985). Tooth size in 47 XYY males: Evidence for a direct effect of the Y chromosome on growth. *Australian Dental Journal*, 30: 268-272.

Thompson, D. W. (1917). *On Growth and Form*. Cambridge University Press, Cambridge.

Van Der Linden, F. P. G. M., Boersma, H., Zelders, T., Peters, K. A., and Raaben, J. H. (1972). Three-dimensional analysis of dental casts my means of the optocom. *Journal of Dental Research*, 51 (4): 1100.

Wahba, G. (1990). *Spline Models for Observational Data*. Society for Industrial and Applied Mathematics, Philadelphia.

Waite, I. M. and Strahan, D. J. (1989). *A Colour Atlas of Periodontology*. Mosby.

Walker, J. A. (2000). Ability of geometric morphometric methods to estimate a known covariance matrix. *Systematic Biology*, 49(4):686-696.

Weatherhill, G. B., editor (1986). *Regression Analysis with Applications*. London: Chapman and Hall.

Wheeler, R. C. (1962). *An Atlas of Tooth Form*. W.B. Saunders company, Philadelphia.

Wheeler, R. C. (1974). *Dental Anatomy, Physiology and Occlusion*, W.B. Saunders company, Philadelphia.

Wilson, T. G. and Kornman, K. S. (1996). *Fundamentals of Periodontics*. Quintessence Publishing Co, Inc.

Wolpoff, M. H. (1971). *Metric Trends in Hominid Dental Evolution*. Cleveland: Case Western Reserve University Press.

Wood, B. A. and Engleman, C. A. (1988). Analysis of the dental morphology of Plio-Pleistocene hominids. V. Maxillary post-canine tooth morphology. *Journal of Anatomy*, 161:1-35.

Zahn, C. T. and Roskies, R. Z. (1972). Fourier descriptors for plane closed curves. *IEEE Transactions on Computers*, C21:269-281.

Zelditch, M. L, Swiderski, D. L., Sheets, H. D., and Fink, W. L. (2004). *Geometric Morphometrics for Biologists: A Primer*. Elsevier Academic Press.

Zienkiewicz, O. C. (1971). *The Finite Element Method in Engineering Science*. McGraw-Hill, London.

Appendix

On the enclosed CD is a text file containing the S-plus routines 'semi.it' and 'semi.it.new', as referred to in sections 5.5.3 and 7.5 respectively. Descriptions of the arguments which may be supplied (including default values) can be found in these sections, along with details of the various graphical and data output options which may be specified.

Both functions call on many other sub-routines, some of which have been adapted from I.L. Dryden's 'shape' package (see section 2.5). These are also included within the text file, along with several other functions which have been used within this thesis.

To install within S-plus (Windows version), simply cut and paste the contents of the text file into the package's command window.

Also included on the disc are several of the datasets used for this work.

THESIS CONTAINS CD ROM