Part I

Parametric image registration

4

LANDMARK-BASED REGISTRATION

4.1 General remarks

In this first part we will discuss image registration techniques which are based on a finite set of parameters and/or a finite set of so-called *image features.* The basic idea is to determine the transformation such that for a finite number of features, any feature of the template image is mapped onto the corresponding feature of the reference image.

Typical features are, for example, "hard" or "soft" landmarks in the images; see Fig. 4.1 below for an illustration. A landmark is the location of a typically outstanding feature of an image, e.g., the tip of a finger or the point of maximal curvature. *Hard* landmarks or *prospective* landmarks are so-called *fiducial markers* which are positioned before imaging at certain spatial positions on a patient. Typically, the spatial position of these landmarks can be deduced from the images with high accuracy; see, e.g., Maurer & Fitzpatrick (1993) and references therein. However, this type of landmark might be very uncomfortable for the patient. In contrast, *soft* landmarks or *retrospective* landmarks are deduced from the images themselves. The spatial location of these "anatomical" landmarks requires expert knowledge and/or sophisticated image analysis tools for automatic detection; see, e.g., Rohr (2001). Other features based on global intensity knowledge will be discussed in Chapter 5.

To make the feature-based registration idea slightly more formal, let $\mathcal{F}(R,j)$ and $\mathcal{F}(T,j)$ denote the j^{th} feature in the reference image R and the template image T, respectively, $j = 1,\dots,m$, where $m \in \mathbb{N}$ denotes the number of features. The registration problem reads as follows.

Problem 4.1 *Let $m \in \mathbb{N}$ and the features $\mathcal{F}(R,j)$ and $\mathcal{F}(T,j)$, $j = 1,\dots,m$, be given. Find a transformation $\varphi : \mathbb{R}^d \to \mathbb{R}^d$, such that*

$$\mathcal{F}(R,j) = \varphi(\mathcal{F}(T,j)), \quad j = 1,\dots,m. \tag{4.1}$$

As we will see subsequently, the interpolation problem 4.1 might also be replaced by the following approximation problem 4.2. For convenience and later usage, we define the *distance measure*

$$\mathcal{D}^{\mathrm{LM}}[\varphi] := \sum_{j=1}^{m} \|\mathcal{F}(R,j) - \varphi(\mathcal{F}(T,j))\|_f^2 , \tag{4.2}$$

where $\|\cdot\|_f$ denotes a norm on the feature space, e.g., $\|\cdot\|_f = \|\cdot\|_{\mathbb{R}^d}$, if the features are locations of points.

Problem 4.2 *Let $\mathcal{D}^{\mathrm{LM}}$ be as in eqn (4.2). Find a transformation $\varphi : \mathbb{R}^d \to \mathbb{R}^d$, such that $\mathcal{D}^{\mathrm{LM}}[\varphi] = \min$.*

Suitability means that the transformation is either an element of a typically finite-dimensional space, for example, spanned by polynomials, splines, or wavelets (see Section 3.3.1) or it is required to be smooth in a certain sense to be discussed later. Note that any of these choices may be viewed as a regularization of Problem 4.2.

4.2 Landmark-based parametric registration

In this approach, the features to be mapped are the locations of a number of user-supplied landmarks, i.e., spatial positions,

$$\mathcal{F}(R,j) = x^{R,j}, \quad \mathcal{F}(T,j) = x^{T,j}, \qquad j = 1,\ldots,m, \quad m \in \mathbb{N}.$$

Figure 4.1 illustrates the location of a set of six landmarks in a reference image and the corresponding set in a template image.

Using the Euclidean norm in eqn (4.2), we obtain

$$\mathcal{D}^{\mathrm{LM}}[\varphi] = \sum_{j=1}^{m} \left\|x^{R,j} - \varphi(x^{T,j})\right\|_{\mathbb{R}^d}^2 . \tag{4.3}$$

For the first approach, we expand the transformation $\varphi = (\varphi_1,\ldots,\varphi_d)^\top$ in terms of some basis functions ψ_k, i.e., with some coefficients $\alpha_{\ell,k}$, we have

$$\varphi_\ell = \sum_{k=1}^{n} \alpha_{\ell,k}\psi_k, \qquad \alpha_{\ell,k} \in \mathbb{R}, \quad \psi_k : \mathbb{R}^d \to \mathbb{R}, \quad n \in \mathbb{N}, \quad \ell = 1,\ldots,d. \tag{4.4}$$

For simplicity, we assume that the ψ_k's as well as the lengths of the expansions are the same for all $\ell = 1,\ldots,d$ and do not depend on ℓ.

With the expansion (4.4), Problem 4.2 can be rewritten as follows.

Problem 4.3 *Find parameters $\alpha_{\ell,k} \in \mathbb{R}$, $k = 1,\ldots,n$, $\ell = 1,\ldots,d$, such that $\mathcal{D}^{\mathrm{LM}}[\varphi] = \min$, where $\mathcal{D}^{\mathrm{LM}}$ is given by eqn (4.3) and $\varphi = (\varphi_1,\ldots,\varphi_d)^\top$ is given by eqn (4.4).*

A computation gives

$$\mathcal{D}^{\mathrm{LM}}[\varphi] = \sum_{j=1}^{m}\sum_{\ell=1}^{d}\left(x_\ell^{R,j} - \sum_{k=1}^{n}\alpha_{\ell,k}\psi_k(x^{T,j})\right)^2 = \sum_{\ell=1}^{d}\|y_\ell - \Psi a_\ell\|_{\mathbb{R}^d}^2 ,$$

where for $\ell = 1,\ldots,d$,

$$y_\ell = (x_\ell^{R,1},\ldots,x_\ell^{R,m})^\top \in \mathbb{R}^m,$$
$$\Psi = \left(\psi_k(x^{T,j})\right)_{\substack{j=1,\ldots,m\\k=1,\ldots,n}} \in \mathbb{R}^{m\times n},$$
$$\text{and} \quad a_\ell = (\alpha_{\ell,1},\ldots,\alpha_{\ell,n})^\top \in \mathbb{R}^n.$$

Thus, the problem of determining the optimal parameters $\alpha_{\ell,k}$ in Problem 4.3 decouples with respect to the spatial dimension ℓ. The optimal parameters can be obtained by solving d least squares problems, where, under the above assumption, the matrix Ψ does not depend on ℓ.

Assuming that Ψ has full rank, $\operatorname{rank}(\Psi) = n$, the solution of Problem 4.3 is unique. A numerical solution can be obtained by using a QR-factorization of Ψ; see, e.g., Golub & van Loan (1989, §5.3.4).

If in particular $\|y_\ell - \Psi a_\ell\|_{\mathbb{R}^d} = 0$ for all $\ell = 1,\ldots,d$, we have a one-to-one correspondence between the landmarks in the reference and template images; see also Fig. 4.1.

For the important case of a linear transformation, we have $n = d+1$ and

$$\varphi = \begin{pmatrix} \sum_{k=1}^{d+1}\alpha_{1,k}\psi_k \\ \vdots \\ \sum_{k=1}^{d+1}\alpha_{d,k}\psi_k \end{pmatrix} \in \Pi_1^d(\mathbb{R}^d),$$

where $\psi_1(x) = 1$, $\psi_{\ell+1}(x) = x_\ell$, $\ell = 1,\ldots,d$. The restriction $\operatorname{rank}(\Psi) = n = m$ is equivalent to $x^{T,j}$ not being co-linear.

However, the theorem of Mairhuber & Curtis (1956, 1958) (for a modern formulation see, e.g., Braess (1980)) gives a disappointing answer to the question whether an interpolation problem in multi-dimensional spaces has a unique solution or not. The following corollary states that a set of points can be found, such that the Vandermonde matrix becomes singular. Thus, additional conditions are required to make the problem well-posed. For example, using an affine linear interpolation, the interpolation points are not allowed to be co-linear.

Corollary 4.1 *Let $d, m \in \mathbb{N}$, ψ_k, $k = 1, \dots, m$, be continuous functions, $\psi_k : \mathbb{R}^d \to \mathbb{R}$. If $d > 1$ and $m > 1$, then there exists a set of points $x_j \in \mathbb{R}^d$, $j = 1, \dots, m$, such that $\det(\psi_k(x_j)_{j,k=1,\dots,n}) = 0$.*

Proof Assume there exists a set of points $x_j \in \mathbb{R}^d$, $j = 1, \dots, m$, such that $\det(\psi_k(x_j)_{j,k=1,\dots,n}) = c \neq 0$ and hence $x_j \neq x_k$ for $k \neq j$. Since $n \geq 2$, we can find a closed, non-intersecting curve $\gamma : [0,1] \to \mathbb{R}^d$ with $x_1 = \gamma(0) = \gamma(1)$, $\gamma(1/2) = x_2$, and $x_k \notin \gamma(0,1])$ for $k = 3, \dots, m$. For the determinant of

$$\Psi(t) := \begin{pmatrix} \psi_1(\gamma(t)) & \psi_2(\gamma(t)) & \cdots & \psi_m(\gamma(t)) \\ \psi_1(\gamma(t+\frac{1}{2})) & \psi_2(\gamma(t+\frac{1}{2})) & \cdots & \psi_m(\gamma(t+\frac{1}{2})) \\ \psi_1(x_3) & \psi_2(x_3) & \cdots & \psi_m(x_3) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_1(x_m) & \psi_2(x_m) & \cdots & \psi_m(x_m) \end{pmatrix} \in \mathbb{R}^{m \times m},$$

we have $\det \Psi(0) = -\det \Psi(\frac{1}{2})$, since $\Psi(\frac{1}{2})$ can be obtained from $\Psi(0)$ by interchanging the first two rows. Thus there exists a $\xi \in]0, \frac{1}{2}[$, such that $\det \Psi(\xi) = 0$ and thus the Vandermonde matrix with respect to $\gamma(\xi), \gamma(\xi + \frac{1}{2}), x_3, \dots, x_m$ is singular. □

Note that for $d = 1$, the proof does not apply since we cannot find a non-intersecting curve. This shows the particular quality of one-dimensional interpolation.

The main disadvantage is, however, that the transformation from the parametric approach is in general not diffeomorphic. Figure 4.1 shows the results for a linear ($\varphi \in \Pi_1^d(\mathbb{R}^d)$) and a quadratic ($\varphi \in \Pi_2^d(\mathbb{R}^d)$) parametric registration. As is apparent from this figure, the linear approach yields satisfactory results, though the fit of the landmarks is not perfect. After quadratic registration, all landmarks are mapped perfectly. However, since φ is a quadratic polynomial, the map is not diffeomorphic and leads to a "mirrored" image, which is certainly not a satisfactory registration. Note that in this example, the landmarks are chosen such that the interpolation problem is well-posed.

4.3 Landmark-based smooth registration

As already seen in Fig. 4.1, the parametric approach presented in Section 4.2 has some severe drawbacks. Figure 4.2 (LEFT) illustrates these drawbacks for dimension one. The figure shows the results of approximating some monotonic data a linear and a quadratic polynomial. Although the quadratic polynomial

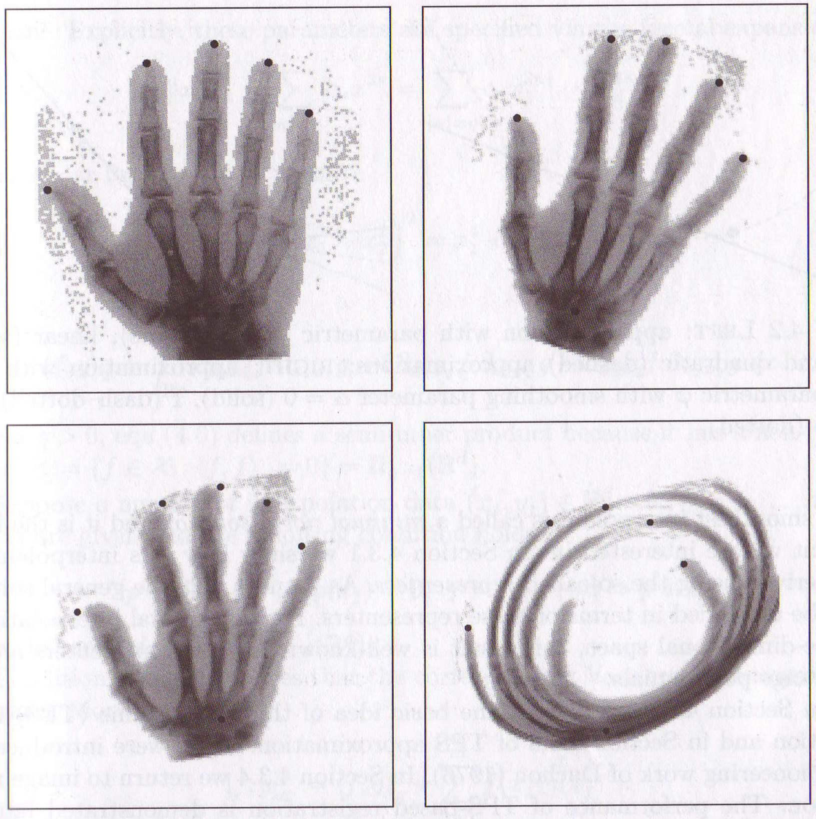

FIG. 4.1 Landmark-based image registration. TOP LEFT: reference, and TOP RIGHT: template with landmarks (black dots), BOTTOM LEFT: parametric linear registered, BOTTOM RIGHT: parametric quadratic registered.

is optimal with respect to the data, it is not preferable for registration. This is because the quadratic is not bijective, manifests oscillation, and does not reflect the monotonicity of the data.

Instead of tuning parameters in an expansion of the transformation in terms of some more or less artificial basis functions, we introduce additional smoothness restrictions to the transformation. These restrictions are expressed by a functional $\mathcal{S}$. Roughly speaking, smoothness is measured in terms of curvature. It turns out, somewhat surprisingly, that the minimizer of this regularized approach is again parameterized: it is a linear combination of shifts of a *radial basis function* plus some polynomial corrections.

In order to provide a detailed insight into the underlying interpolation concepts, we present a general treatment following Light (1995). To begin with, we are looking for an interpolant $\psi : \mathbb{R}^d \to \mathbb{R}$ which is smooth in a certain sense.

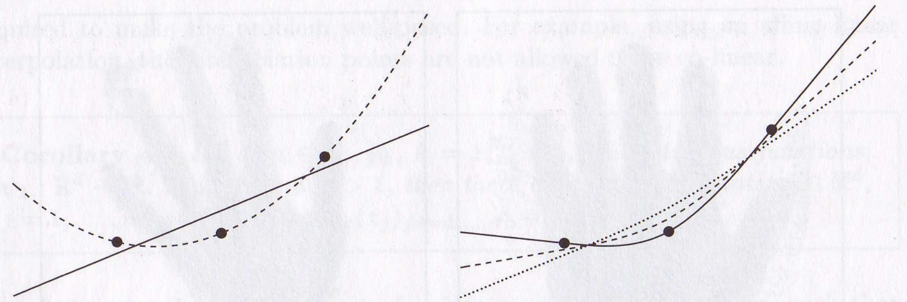

FIG. 4.2 LEFT: approximation with parametric φ, data (dots), linear (solid), and quadratic (dashed) approximations; RIGHT: approximation with non-parametric φ with smoothing parameter $\alpha = 0$ (solid), 1 (dash–dotted), and 5 (dotted).

The smoothest interpolant is called a *minimal norm solution* and it is the interpolant we are interested in. In Section 4.3.1 we show how this interpolant can be derived using the so-called *representers*. As it turns out, the general solution can be expanded in terms of these representers. For polynomial interpolation in a one-dimensional space, this result is well-known and the representers are the Lagrange polynomials.

In Section 4.3.2 we present the basic idea of thin plate spline (TPS) interpolation and in Section 4.3.3 of TPS approximation. TPSs were introduced by the pioneering work of Duchon (1976). In Section 4.3.4 we return to image registration. The performance of TPS-based registration is demonstrated later by Fig. 4.4.

4.3.1 *Minimal norm solutions for interpolation problems*

Following Light (1995) we treat the problem of finding a minimal norm element in a space with interpolation restrictions. To this end, we consider the L_2 inner product

$$\langle f, g\rangle_0 := \langle f, g\rangle_{L_2} := \int_{\mathbb{R}^d} f(x)g(x)dx \qquad (4.5)$$

and a semi-inner product of order q,

$$\langle\cdot,\cdot\rangle_q : \mathcal{X}\times\mathcal{X}\to\mathbb{R}, \quad \langle f, g\rangle_q := \sum_{|\kappa|=q} c_\kappa \langle D^\kappa f, D^\kappa g\rangle_0, \qquad (4.6)$$

where $\kappa \in \mathbb{N}_0^d$, $|\kappa| = \kappa_1 + \cdots + \kappa_d$, $D^\kappa f = \left(\frac{\partial}{\partial x_1}\right)^{\kappa_1}\cdots\left(\frac{\partial}{\partial x_d}\right)^{\kappa_d} f$, $\mathcal{X} = H^q \cap C(\mathbb{R}^d) \cup \Pi_{q-1}(\mathbb{R}^d)$, and H^q denotes the Sobolev space of order q. The set of coefficients $\{\, c_\kappa : |\kappa| = q \,\}$ is chosen such that the semi-norm is rotationally invariant. Explicitly, these parameters are specified via the formal expansion

$$\|x\|_{\mathbb{R}^d}^{2q} = \sum_{|\kappa|=q} c_\kappa x^{2\kappa} = \sum_{|\kappa|=q} c_\kappa x_1^{2\kappa_1}\cdots x_d^{2\kappa_d}. \qquad (4.7)$$

In particular for $d = q = 2$, we have

$$\|x\|_{\mathbb{R}^2}^4 = \left(x_1^2 + x_2^2\right)^2 = x_1^4 + 2x_1^2x_2^2 + x_2^4$$

and

$$\langle f, g\rangle_2 = \int_{\mathbb{R}^d} \partial_{x_1x_1} f\,\partial_{x_1x_1} g + 2\partial_{x_1x_2} f\,\partial_{x_1x_2} g + \partial_{x_2x_2} f\,\partial_{x_2x_2} g\,dx.$$

For $q > 0$, eqn (4.6) defines a semi-inner product because it has a non-trivial kernel $K = \{f \in \mathcal{X} : \langle f, f\rangle_q = 0\} = \Pi_{q-1}(\mathbb{R}^d)$.

Suppose a number of interpolation data $(x_j, y_j) \in \mathbb{R}^d \times \mathbb{R}$, $j = 1, \ldots, m$, for $m \in \mathbb{N}$ are given and the following condition holds,

$$[p \in \Pi_{q-1}(\mathbb{R}^d) \wedge p(x_j) = 0, \ \ j = 1, \ldots, m] \Rightarrow p \equiv 0. \qquad (4.8)$$

Note that $m \geq d_q := \dim(\Pi_{q-1}(\mathbb{R}^d))$.

Condition (4.8) guarantees that the corresponding Vandermonde matrix has full rank. Moreover,

$$[f, g]_q := \langle f, g\rangle_q + \sum_{j=1}^m f(x_j)g(x_j) \qquad (4.9)$$

is an inner product on $\mathcal{X}$; cf., e.g., Light (1995).

Our goal is to derive an explicit expression for the *minimal norm solution* ψ, i.e.,

$$\psi = \arg\min\{[f, f]_q,\ f \in H \text{ and } f(x_j) = y_j,\ j = 1, \ldots, m\},$$

where $H := (\mathcal{X}, [\cdot,\cdot]_q)$ is a Hilbert space. The minimizer is unique since the norm is convex. In order to compute the minimizer, we construct a representer $R_x \in H$ for $x \in \mathbb{R}^d$, where the representer is characterized by $f(x) = [f, R_x]_q$ for all $f \in H$.

Under the additional assumption

$$\forall x \in \mathbb{R}^d\ \exists K_x \in \mathbb{R}_{\geq 0} :$$

$$f(x_1) = \cdots = f(x_m) = 0 \Rightarrow |f(x)| \leq K_x\sqrt{\langle f, f\rangle_q}, \qquad (4.10)$$

the point evaluation functionals (cf., Definition 3.5) are continuous functionals on the Hilbert space $H := (\mathcal{X}, [\cdot,\cdot]_q)$. An important fact is that if

$$2q > d,$$

then H consists of continuous functions on $\mathbb{R}^d$, and the point evaluation functionals are elements of the dual H^*; see, e.g., Light (1995).

We now define a subspace $H_0 \subset H$ with integrated interpolation conditions,

$$H_0 := \{ f \in H : \; f(x_1) = \cdots = f(x_{d_q}) = 0 \} ,$$

where $x_1, \ldots, x_{d_q}$ are chosen such that condition (4.8) holds for $m = d_q$. We then construct a representer $R_{0,x}$ for a fixed $x \in \mathbb{R}^d$ on the space H_0 and finally extend $R_{0,x}$ to R_x, a representer for x on H.

For $f \notin H_0$, a projection $\mathcal{Q} : H \to H_0$ has to be used. Let $L_1, \ldots, L_{d_q}$ be a Lagrange basis for $\Pi_{q-1}(\mathbb{R}^d)$ and $\mathcal{P} : H \to \Pi_{q-1}(\mathbb{R}^d)$ with

$$\mathcal{P}f := \sum_{j=1}^{d_q} f(x_j) L_j .$$

Hence, $\mathcal{Q} = \mathcal{I} - \mathcal{P}$ maps H onto H_0.

For $u, v \in H$ we define the inner product

$$[u, v]_q = \langle u, v \rangle_q + \sum_{j=1}^{d_q} u(x_j) v(x_j) .$$

Note that $[u, f]_q = \langle u, f \rangle_q$ for all $u \in H$ and $f \in H_0$.

For a characterization of $R_{0,x}$, we assume that for a fixed $x \in \mathbb{R}^d$ the function $R_{0,x}$ is a representer on H_0. Thus, we have

$$\begin{aligned} f(x) - \mathcal{P}f(x) &= [f - \mathcal{P}f, R_{0,x}]_q = \langle f - \mathcal{P}f, R_{0,x} \rangle_q = \langle f, R_{0,x} \rangle_q \\ &= \sum_{|\kappa|=q} c_\kappa \int_{\mathbb{R}^d} (D^\kappa f)(y)(D^\kappa R_{0,x})(y) dy \\ &= \sum_{|\kappa|=q} c_\kappa \langle D^\kappa f, D^\kappa R_{0,x} \rangle_0 \\ &= \left\langle f, (-1)^q \sum_{|\kappa|=q} c_\kappa D^{2\kappa} R_{0,x} \right\rangle_0 , \end{aligned} \tag{4.11}$$

where Green's formula (see, e.g., Buck (1978, §9.4)) has been used. On the other hand

$$f(x) - \mathcal{P}f(x) = f(x) - \sum_{j=1}^{d_q} f(x_j) L_j(x) = \left\langle f, \delta_x - \sum_{j=1}^{d_q} L_j(x) \delta_{x_j} \right\rangle_0 . \tag{4.12}$$

Equations (4.11) and (4.12) show that $R_{0,x}$ is the solution of the distributional differential equation

$$(-1)^q \sum_{|\kappa|=q} c_\kappa D^{2\kappa} R_{0,x} = \delta_x - \sum_{j=1}^{d_q} L_j(x) \delta_{x_j} . \tag{4.13}$$

Note that the right hand side of eqn (4.13) is a linear combination of point evaluation functionals. Thus, the solution can be derived from combinations of shifted versions of a fundamental solution or radial basis function.

Let $\rho_x(y) := \rho(\|y - x\|_{\mathbb{R}^d})$, where ρ is the *radial basis function* or Green's function; see Theorem 4.2 and Rohr (2001).

Theorem 4.2 *The radial basis function for* $(-1)^q \sum_{|\kappa|=q} c_\kappa D^{2\kappa}$ *is given by*

$$\rho(r) := c_q^d \begin{cases} r^{2q-d} \log r, & d \text{ even}, \\ r^{2q-d}, & d \text{ odd}, \end{cases} \tag{4.14}$$

where

$$c_q^d = \begin{cases} \frac{(-1)^{q+1+d/2}}{2^{2q-1}\pi^{d/2}(q-1)!(q-d/2)!}, & d \text{ even}, \\ \frac{\Gamma(d/2-q)}{2^{2q}\pi^{d/2}(q-1)!}, & d \text{ odd}. \end{cases}$$

With this radial basis function we have

$$(-1)^q \sum_{|\kappa|=q} c_\kappa D^{2\kappa} \rho_x = \delta_x$$

and thus a particular solution of eqn (4.13) is given by

$$\widetilde{R}_{0,x} = \rho_x - \sum_{j=1}^{d_q} L_j(x) \rho_{x_j} .$$

Since $\widetilde{R}_{0,x}$ is not necessarily in H_0 we define

$$\begin{aligned} R_{0,x} &:= (\mathcal{I} - \mathcal{P}) \widetilde{R}_{0,x} \\ &= \rho_x - \sum_{j=1}^{d_q} L_j(x) \rho_{x_j} - \sum_{j=1}^{d_q} \rho_x(x_j) L_j - \sum_{j,k=1}^{d_q} \rho_{x_j}(x_k) L_j(x) L_k . \end{aligned} \tag{4.15}$$

It remains to extend $R_{0,x}$ to the full space H. To this end we make use of the following lemma.

Lemma 4.3 (Light (1995, Th. 2.1)) *Let x_j, $j = 1, \dots, d_q$, be distinct points in $\mathbb{R}^d$ satisfying condition (4.8) for $m = d_q = \dim(\Pi_{q-1}(\mathbb{R}^d))$. Then the Lagrange polynomial L_k is a representer for x_k in H, $k = 1, \dots, d_q$.*

Proof For any $f \in H$ we have

$$[f, L_k]_q = \langle f, L_k \rangle_q + \sum_{j=1}^{d_q} f(x_j) L_k(x_j) = \langle f, L_k \rangle_q + f(x_k) = f(x_k),$$

since $L_k \in \Pi_{q-1}(\mathbb{R}^d)$ belongs to the kernel of the semi-inner product. □

Now we are in a position to construct the representer R_x in H. For $f \in H$ we have

$$\begin{aligned}[f, R_{0,x}]_q &= [f - \mathcal{P}f, R_{0,x}]_q + [\mathcal{P}f, R_{0,x}]_q = f(x) - \mathcal{P}f(x) + 0 \\ &= f(x) - \sum_{j=1}^{d_q} f(x_j) L_j(x) = f(x) - \sum_{j=1}^{d_q} [f, L_j]_q L_j(x)\end{aligned}$$

or

$$f(x) = [f, R_{0,x}]_q + \sum_{j=1}^{d_q} [f, L_j]_q L_j(x) = \left[f, R_{0,x} + \sum_{j=1}^{d_q} L_j(x) L_j \right]_q,$$

showing that

$$R_x := R_{0,x} + \sum_{j=1}^{d_q} L_j(x) L_j$$

is a representer in H.

The following theorem gives an explicit expansion of the minimal norm solution in terms of representers.

Theorem 4.4 (Light (1995, Th. 1.1)) *Let $f \in H$ with $f(x_j) = y_j$, $j = 1, \dots, m$, and let $\psi \in C_f$ be the minimal norm solution, where*

$$C_f := \left\{ v \in H : [v, v]_q \le [f, f]_q \wedge v(x_j) = y_j,\ j = 1, \dots, m \right\}.$$

Let R_x be the representer for $x \in \mathbb{R}^d$. Then there exist $\theta_1, \dots, \theta_m \in \mathbb{R}$, such that $\psi = \sum_{j=1}^m \theta_j R_{x_j}$. The coefficients $\theta_1, \dots, \theta_m$ are determined by the equations

$$y_k = \psi(x_k) = [\psi, R_{x_k}]_q = \sum_{j=1}^m \theta_j \left[R_{x_j}, R_{x_k} \right]_q.$$

However, a characterization explicitly based on coefficients in an expansion with respect to basis functions is of particular interest. The next Theorem 4.5 characterizes the minimal norm solution with respect to the radial basis functions.

Theorem 4.5 *Let $d, m, q \in \mathbb{N}$ and $d_q := \dim(\Pi_{q-1}(\mathbb{R}^d))$ and let $x_j \in \mathbb{R}^d$, $y_j \in \mathbb{R}$, $j = 1, \dots, m$, be given interpolation data. The minimal norm solution*

$$\psi = \arg\min \{ [f, f]_q,\ f \in H \text{ and } f(x_j) = y_j,\ j = 1, \dots, m \}$$

is characterized by

$$\psi = \sum_{j=1}^m \theta_j \rho_{x_j} + \sum_{j=1}^{d_q} \beta_j p_j, \tag{4.16}$$

where $\rho_{x_j} = \rho(\| \cdot - x_j \|_{\mathbb{R}^d})$ (see Theorem 4.2) and $p_1, \dots, p_{d_q}$ is a basis for $\Pi_{q-1}(\mathbb{R}^d)$. The coefficients $\theta := (\theta_1, \dots, \theta_m)^\top \in \mathbb{R}^m$ and $\beta := (\beta_1, \dots, \beta_{d_q})^\top \in \mathbb{R}^{d_q}$ are determined by the following system of linear equations,

$$\begin{pmatrix} K & B^\top \\ B & 0 \end{pmatrix} \begin{pmatrix} \theta \\ \beta \end{pmatrix} = \begin{pmatrix} y \\ 0 \end{pmatrix}, \tag{4.17}$$

$$y := (y_1, \dots, y_m)^\top \in \mathbb{R}^m,$$

$$K := (\rho(\|x_j - x_k\|_{\mathbb{R}^d}))_{j,k=1,\dots,m} \in \mathbb{R}^{m \times m} \tag{4.18}$$

$$\text{and} \quad B := (p_j(x_k))_{\substack{j=1,\dots,d_q \\ k=1,\dots,m}} \in \mathbb{R}^{d_q \times m}. \tag{4.19}$$

Proof In the expansion $\psi = \sum_{j=1}^m \mu_j R_{x_j}$ (see Theorem 4.4) any representer is a linear combination of translated radial basis functions and polynomials. Thus, with some coefficients θ_j and β_j together with a basis $p_1, \dots, p_{d_q}$ for $\Pi_{q-1}(\mathbb{R}^d)$ we have the characterization given by eqn (4.16), which satisfies the equations

$$y_k = \psi(x_k) = \sum_{j=1}^m \theta_j \rho_{x_j}(x_k) + \sum_{j=1}^{d_q} \beta_j p_j(x_k), \quad k = 1, \dots, m, \tag{4.20}$$

or $K\theta + B^\top \beta = y$.

Moreover, since $\Pi_{q-1}(\mathbb{R}^d)$ is the kernel of $\langle \cdot, \cdot \rangle_q$, we have for any $p \in \Pi_{q-1}(\mathbb{R}^d)$,

$$\begin{aligned}
0 = \langle \psi, p \rangle_q &= \sum_{|\kappa|=q} c_\kappa \int_{\mathbb{R}^d} (D^\kappa \psi)(D^\kappa p)\, dx \\
&= \sum_{j=1}^m \theta_j \int_{\mathbb{R}^d} (-1)^q \sum_{|\kappa|=q} c_\kappa (D^{2\kappa} \rho_{x_j}) p\, dx \\
&= \sum_{j=1}^m \theta_j \left\langle (-1)^q \sum_{|\kappa|=q} c_\kappa D^{2\kappa} \rho_{x_j}, p \right\rangle_0 \\
&= \sum_{j=1}^m \theta_j \langle \delta_{x_j}, p \rangle_0 = \sum_{j=1}^m \theta_j p(x_j).
\end{aligned} \tag{4.21}$$

Expanding p with respect to $p_1, \ldots, p_{d_q}$, we also find $B\theta = 0$. □

For a modest number of interpolation data, the system of linear equations (4.17) can be solved using standard techniques. Some results are shown in Section 4.3.4, where we also compare this approach with one based on approximation. For a higher number of interpolation data this is still a topic of research; cf., e.g., Schaback (1997).

4.3.2 *Example: splines and TPSs*

Now we study the important cases $d = 1 \wedge q = 2$ and $d = q = 2$ in detail. With the semi-inner product (4.6) and the abbreviation

$$\mathcal{S}^{\mathrm{TPS}}[\psi] := \tfrac{1}{2} \langle \psi, \psi \rangle_q^2 \tag{4.22}$$

we consider the following problem.

> **Problem 4.4** *Find* $\psi : \mathbb{R}^d \to \mathbb{R}$, *such that*
>
> $$\mathcal{S}^{\mathrm{TPS}}[\psi] = \min \ \textit{subject to} \ \psi(x_j) = y_j, \ j = 1, \ldots, m.$$

We will derive an explicit solution for this problem. The Lagrange function for Problem 4.4 reads

$$L[\psi, \lambda] = \mathcal{S}^{\mathrm{TPS}}[\psi] + \sum_{j=1}^m \lambda_j (\delta_{x_j}[\psi] - y_j),$$

where λ_j are Lagrange multipliers and δ_z are the point evaluation functionals (cf., Definition 3.5) located at position z, i.e.,

$$\delta_z[\psi] := \int_{\mathbb{R}^d} \delta_z(x) \psi(x) dx = \psi(z).$$

We start by computing the Gâteaux derivative of δ_z,

$$\begin{aligned}
d\delta_z[\psi; \zeta] &:= \lim_{h \to 0} \frac{1}{h} (\delta_z[\psi + h\zeta] - \delta_z[\psi]) \\
&= \lim_{h \to 0} \frac{1}{h} (\psi(z) + h\zeta(z) - \psi(z)) = \delta_z[\zeta],
\end{aligned}$$

and the Gâteaux derivative of $\mathcal{S}^{\mathrm{TPS}}$,

$$\begin{aligned}
d\mathcal{S}^{\mathrm{TPS}}[\psi; \zeta] &:= \lim_{h \to 0} \frac{1}{h} (\mathcal{S}^{\mathrm{TPS}}[\psi + h\zeta] - \mathcal{S}^{\mathrm{TPS}}[\psi]) \\
&= \sum_{|c_\kappa|=q} c_\kappa \lim_{h \to 0} \frac{1}{2h} \int_{\mathbb{R}^d} (D^\kappa[\psi + h\zeta])^2 - (D^\kappa \psi)^2 dx \\
&= \sum_{|c_\kappa|=q} c_\kappa \int_{\mathbb{R}^d} (D^\kappa \psi)\, (D^\kappa \zeta) dx \\
&= (-1)^q \sum_{|c_\kappa|=q} c_\kappa \int_{\mathbb{R}^d} (D^{2\kappa} \psi)\, \zeta\, dx,
\end{aligned}$$

for suitable perturbations ζ. Here, we used Green's formula (cf., e.g., Buck (1978, §9.4)) and, as an implicit necessary condition for a minimizer, $\psi(x) \to 0$ for $\|x\|_{\mathbb{R}^d} \to \infty$.

For $q = 2$, we have

$$(-1)^q \sum_{|c_\kappa|=q} c_\kappa D^{2\kappa} \psi = \Delta^2 \psi = \begin{cases} \psi^{(\mathrm{IV})}, & d = 1 \\ \partial_{1111}\psi + 2\partial_{1122}\psi + \partial_{2222}\psi, & d = 2; \end{cases}$$

cf., eqn (4.7). A minimizer for Problem 4.4 must satisfy the condition

$$dL[\psi, \lambda; \zeta] = \int_{\mathbb{R}^d} \left(\Delta^2 \psi(x) + \sum_{j=1}^m \lambda_j \delta_{x_j}(x) \right) \zeta(x) dx = 0$$

for all perturbations ζ. Hence by variation of ζ, we obtain the distributional differential equation

$$\Delta^2 \psi(x) + \sum_{j=1}^m \lambda_j \delta_{x_j} = 0. \tag{4.23}$$

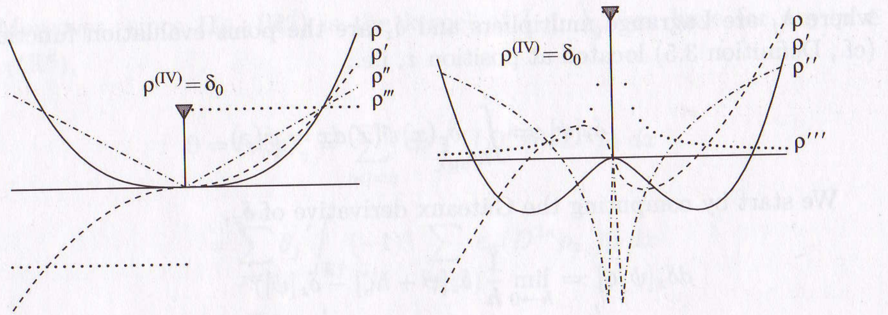

FIG. 4.3 Fundamental solution ρ and derivatives for $d = 1$ (LEFT) and $d = 2$ (RIGHT).

A fundamental solution of $\Delta^2\psi(x) + \delta_0 = 0$ is the Green's function ψ_0, with

$$\psi_0(x) = \rho(\|x\|_{\mathbb{R}^d}) + p_3(x), \quad p_3 \in \Pi_3(\mathbb{R}^d),$$

where with $r := \|x\|_{\mathbb{R}^d}$,

$$\rho(r) = \begin{cases} \frac{1}{12}r^3, & d = 1 \\ \frac{1}{8\pi}r^2 \log r, & d = 2; \end{cases}$$

see also Fig. 4.3.

For $d = 1$ this is verified straighforwardly; for $d = 2$ and $x \neq 0$,

$$\begin{aligned} \nabla\rho(x) &= (8\pi)^{-1}(2\log r + 1)x, \\ \Delta\rho(x) &= (2\pi)^{-1}(\log r + 1), \\ \nabla(\Delta\rho)(x) &= (2\pi r^2)^{-1}x, \\ \Delta^2\rho(x) &= (\pi r^3)^{-1}\langle\nabla r, x\rangle + (2\pi r^2)^{-1}\operatorname{div} x = 0. \end{aligned}$$

Thus the general solution of Problem 4.4 takes the form

$$\psi = \sum_{j=1}^{m} \theta_j \rho_{x_j} + \sum_{j=1}^{3} \beta_j p_j, \quad \operatorname{span}\{p_1, p_2, p_3\} = \Pi_1(\mathbb{R}^d),$$

$$\psi(x_j) = y_j, \quad \text{for } j = 1, \dots, m,$$

which is, of course, in accordance with Theorem 4.4 and Theorem 4.5.

For the case $d = 1$ we rediscovered the well-known interpolating *cubic splines*, and the solution for $d = 2$ is also known as *thin plate splines*; cf., e.g., Rohr (2001). Note that for the one-dimensional case there exists an $\mathcal{O}(m)$ algorithm based on moments (second order derivatives of the spline in the interpolation points); cf., e.g., Piegl & Tiller (1997, §2.6).

4.3.3 *Smooth minimal norm approximations*

We now slightly extend the theory derived in Section 4.3.1. To this end, we consider the inner product

$$[u, v]_{q,\alpha} := \alpha \langle u, v\rangle_q + \sum_{j=1}^{m} (u(x_j) - y_j)(v(x_j) - y_j), \tag{4.24}$$

where $\alpha > 0$ serves as a regularizing parameter. A necessary condition for a minimal norm solution ψ now reads

$$\alpha(-1)^q \sum_{|\kappa|=q} c_\kappa D^{2\kappa}\psi + \sum_{j=1}^{m} \left(\langle\delta_{x_j}, \psi\rangle_0 - y_j\right)\delta_{x_j} = 0.$$

Exploiting the expansion

$$\psi = \sum_{j=1}^{m} \theta_j \rho_{x_j} + \sum_{j=1}^{d_q} \beta_j p_j \tag{4.25}$$

(cf., Theorem 4.5) we have

$$\begin{aligned} 0 &= \alpha(-1)^q \sum_{|\kappa|=q} c_\kappa D^{2\kappa}\psi + \sum_{j=1}^{m} \left(\langle\delta_{x_j}, \psi\rangle_0 - y_j\right)\delta_{x_j} \\ &= \alpha \sum_{j=1}^{m} \theta_j \left((-1)^q \sum_{|\kappa|=q} c_\kappa D^{2\kappa}\rho_{x_j}\right) + \sum_{j=1}^{m} \left(\langle\delta_{x_j}, \psi\rangle_0 - y_j\right)\delta_{x_j} \\ &= \sum_{j=1}^{m} \left(\alpha\theta_j + \left(\langle\delta_{x_j}, \psi\rangle_0 - y_j\right)\right)\delta_{x_j}. \end{aligned}$$

Hence, for $k = 1, \dots, m$,

$$0 = \alpha\theta_k + \left(\langle\delta_{x_k}, \psi\rangle_0 - y_k\right)$$

or, equivalently, $\psi(x_k) + \alpha\theta_k = y_k$. This shows that the optimal parameters θ and β in the expansion (4.25) are determined by

$$\begin{pmatrix} K + \alpha I_m & B^\top \\ B & 0 \end{pmatrix} \begin{pmatrix} \theta \\ \beta \end{pmatrix} = \begin{pmatrix} y \\ 0 \end{pmatrix}, \tag{4.26}$$

where K and B are as in Theorem 4.5 and I_m denotes the identity matrix. Interpolation thus turns out to be a special case of approximation if we choose $\alpha = 0$ in the above equation.

Figure 4.2 (RIGHT) demonstrates the role of the regularization parameter α. For $\alpha = 0$, the function interpolates the data. Varying $\alpha = 1, 5$, we found that the function ψ becomes more and more linear (less curved).

4.3.4 *Smooth transformations*

We now have the ingredients to solve the registration problems 4.5 and 4.6, respectively. We again make use of the abbreviation

$$\mathcal{S}^{\mathrm{TPS}}[\varphi] := \frac{1}{2}\sum_{\ell=1}^{d} \langle \varphi_\ell, \varphi_\ell \rangle_q ,$$

where $\langle \cdot, \cdot \rangle_q$ is the semi-inner product defined by eqn (4.6) with kernel $\Pi_{q-1}(\mathbb{R}^d)$.

Let $x^{T,j}$ and $x^{R,j}$, $j = 1, \ldots, m$, be the given landmarks in the reference and template image, respectively, and let ρ denote the radial basis function with respect to the semi-inner product. Moreover, we define the interpolation space

$$\mathrm{IS}_{q,m} := \mathrm{span}\{\rho(\|x - x^{T,j}\|_{\mathbb{R}^d}),\ j = 1, \ldots, m\}^d \cup \Pi_{q-1}^d(\mathbb{R}^d). \quad (4.27)$$

Problem 4.5 *Find a transformation* $\varphi : \mathbb{R}^d \to \mathbb{R}^d$*, such that*

$$\mathcal{S}^{\mathrm{TPS}}[\varphi] \longrightarrow \min$$

subject to $\varphi(x^{T,j}) = x^{R,j}$, $j = 1, \ldots, m$.

Problem 4.6 *Given* $\alpha > 0$*, find a transformation* $\varphi : \mathbb{R}^d \to \mathbb{R}^d$, $\varphi \in \mathrm{IS}_{q,m}$*, such that*

$$\alpha\mathcal{S}^{\mathrm{TPS}}[\varphi] + \mathcal{D}^{\mathrm{LM}}[\varphi] \longrightarrow \min,$$

where $\mathcal{D}^{\mathrm{LM}}$ *is defined by eqn (4.3) and* $\mathrm{IS}_{q,m}$ *by eqn (4.27).*

The solution of both problems has the form

$$\varphi_\ell = \sum_{j=1}^{m} \theta_{\ell,j}\rho_{x_j} + \sum_{j=1}^{d_q} \beta_{\ell,j} p_j, \quad \ell = 1, \ldots, d,$$

where ρ is the radial basis function corresponding to the semi-inner product, $d_q := \dim(\Pi_{q-1}^d)$, and $p_1, \ldots, p_{d_q}$ is a basis of Π_{q-1}^d. For $\ell = 1, \ldots, d$, the coefficients $\theta_\ell := (\theta_{\ell,1}, \ldots, \theta_{\ell,m})^\top \in \mathbb{R}^m$ and $\beta_\ell := (\beta_{\ell,1}, \ldots, \beta_{\ell,d_q})^\top \in \mathbb{R}^{d_q}$ are determined by

$$\begin{pmatrix} K + \alpha I & B^\top \\ B & 0 \end{pmatrix} \begin{pmatrix} \theta_\ell \\ \beta_\ell \end{pmatrix} = \begin{pmatrix} y_\ell \\ 0 \end{pmatrix}, \quad (4.28)$$

where $y_\ell := (x_\ell^{R,1}, \ldots, x_\ell^{R,m})^\top \in \mathbb{R}^m$ and the matrices K and B are defined in Theorem 4.5. Note that K and B do not depend on ℓ.

Again, the solution of Problem 4.5 might be viewed as a particular solution of Problem 4.6 for $\alpha = 0$.

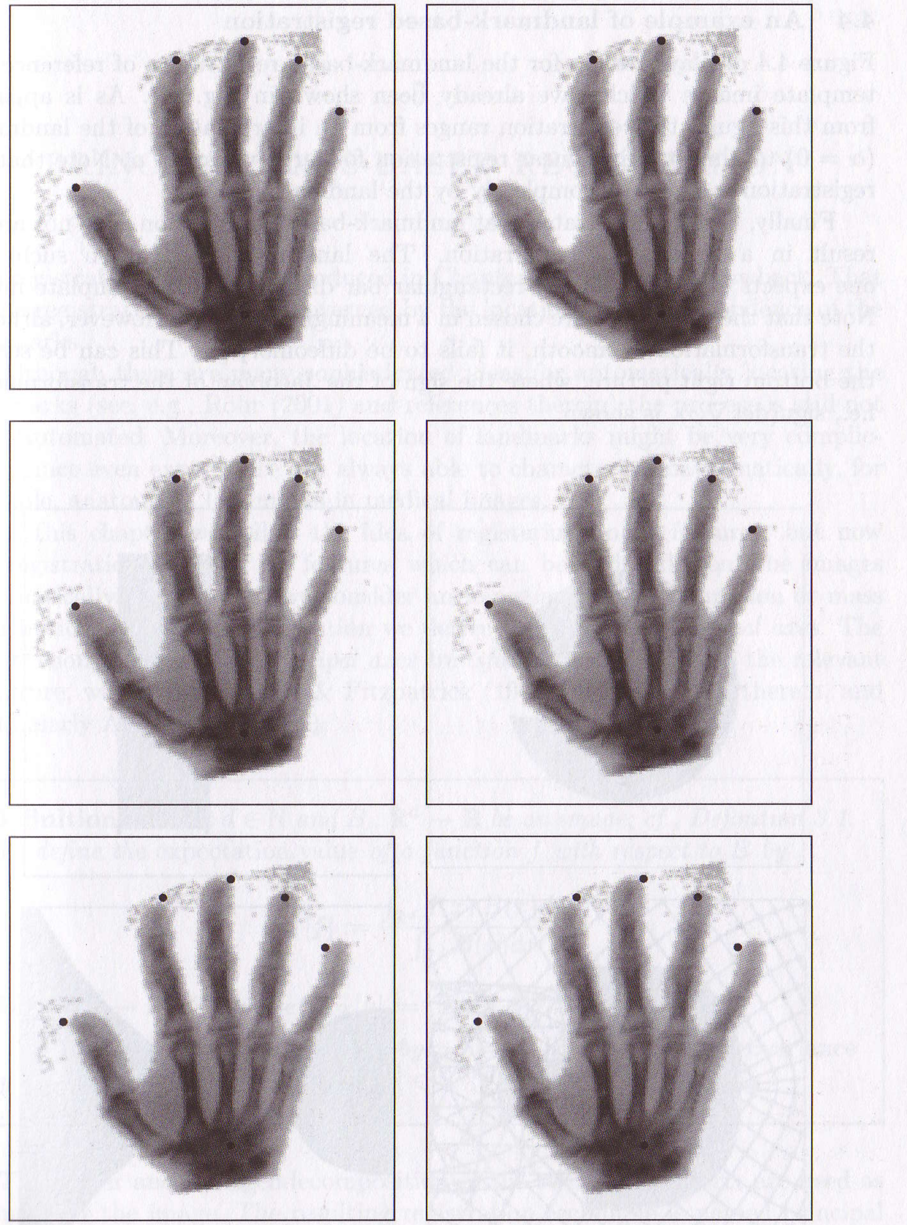

FIG. 4.4 TPS-based registration for $\alpha = 0, 10, 10^2, 10^3, 10^4$, and 10^5 and (LEFT to RIGHT and TOP to BOTTOM) respectively.

4.4 An example of landmark-based registration

Figure 4.4 displays results for the landmark-based registration of reference and template images which have already been shown in Fig. 4.1. As is apparent from this figure, the registration ranges from an interpolation of the landmarks ($\alpha = 0$) to almost affine linear registration for large values of α. Note that the registration is governed completely by the landmarks.

Finally, Fig. 4.5 illustrates that landmark-based registration does not always result in a meaningful registration. The landmarks are chosen such that one expects a bending of the rectangular bar displayed in the template image. Note that the landmarks are chosen in a meaningful ordering. However, although the transformation is smooth, it fails to be diffeomorphic. This can be seen in the bottom right picture, where the sign of the Jacobian of the transformation, i.e., sign(det $\nabla\varphi$), is shown.

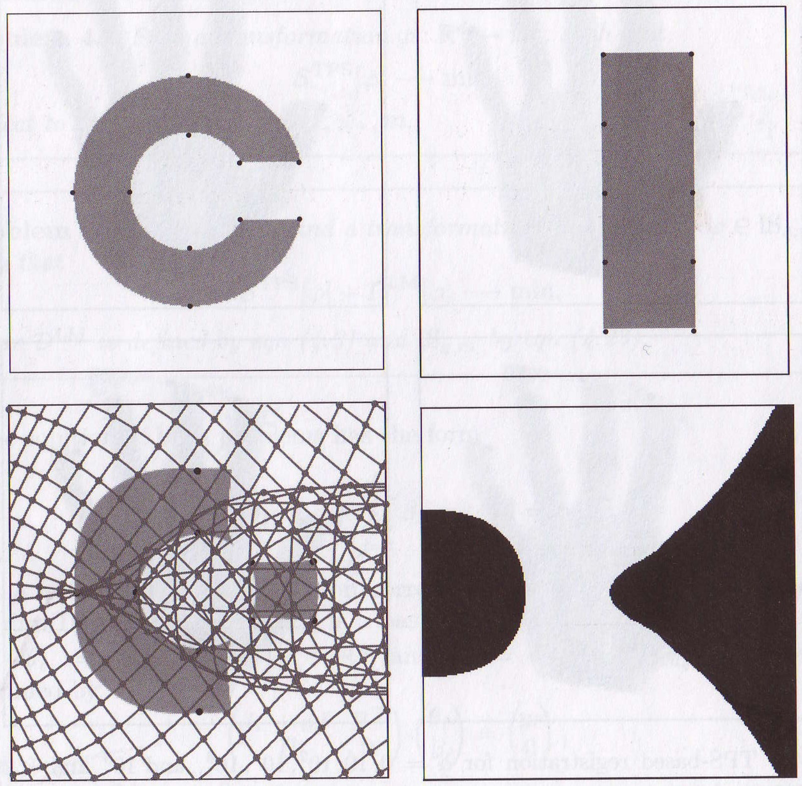

FIG. 4.5 TOP LEFT: reference image with landmarks (dots), TOP RIGHT: template image with landmarks (dots), BOTTOM LEFT: template image after TPS registration with landmarks and deformed grid, BOTTOM RIGHT: sign of determinant of Jacobian of φ (white: det $\nabla\varphi \geq 0$, black: det $\nabla\varphi < 0$).

5

PRINCIPAL AXES-BASED REGISTRATION

The registration technique introduced in Chapter 4 has a major drawback. That is, the registration process is governed by the location and correspondence of the landmarks.

Although there are many sophisticated ideas for automatically locating the landmarks (see, e.g., Rohr (2001) and references therein) the process is still not fully automated. Moreover, the location of landmarks might be very complicated, since even experts are not always able to characterize mathematically, for example, anatomical landmarks in medical images.

In this chapter we follow the idea of registering image features, but now the registration is based on features which can be deduced from the images automatically. To this end, we consider an image as a density function or mass distribution. From this distribution we derive the so-called *principal axes*. The registration is also called *principal axes transformation* (PAT). For the relevant literature, we refer to Maurer & Fitzpatrick (1993) and references therein, and particularly Alpert et al (1990).

Definition 5.1 *Let $d \in \mathbb{N}$ and $B : \mathbb{R}^d \to \mathbb{R}$ be an image; cf., Definition 3.1. We define the* expectation value *of a function f with respect to B by*

$$\mathbb{E}_B[f] := \frac{\int_{\mathbb{R}^d} f(x)\, B(x) dx}{\int_{\mathbb{R}^d} B(x) dx}.$$

For $u : \mathbb{R}^d \to \mathbb{R}^{m\times n}$, we set $\mathbb{E}_B[u] := (\mathbb{E}_B[u_{j,k}])_{\substack{j=1,\dots,m\\ k=1,\dots,n}} \in \mathbb{R}^{m\times n}$.

The center *of an image is defined by $c_B := \mathbb{E}_B[x] \in \mathbb{R}^d$ and the* covariance *by $\mathrm{Cov}_B := \mathbb{E}_B\left[(x-c_B)(x-c_B)^\top\right] \in \mathbb{R}^{d\times d}$.*

The center and an eigendecomposition of the covariance matrix are used as features for the image. The resulting registration technique is named principal axes transformation (PAT); see, e.g., Alpert et al (1990).

Since the covariance matrix is real, symmetric, and positive semi-definite, it permits an orthogonal eigendecomposition

$$\mathrm{Cov}_B = D_B \Sigma_B^2 D_B^\top, \tag{5.1}$$