6
Splines in Biomedical Image Processing

Slavica Jonic and Carlos Oscar Sanchez Sorzano

6.1
Introduction

The most common definition of splines is that they are piecewise polynomials with pieces smoothly connected together. To obtain a continuous representation of a discrete signal in one or more dimensions, one commonly fits it with a spline. The fit may be exact (interpolation splines) or approximate (least-squares or smoothing splines) [1]. By increasing the spline degree, one progressively switches from the simplest continuous representations (piecewise constant and piecewise linear) to the traditional continuous representation characterized by a bandlimited signal model (for the degree equal to the infinity). The traditional Shannon's sampling theory recommends the use of an ideal low-pass filter (antialiasing filter) when the input signal is not bandlimited. In the spline sampling theory, the Shannon's antialiasing filter is replaced with another filter specified by the spline representation of the signal [2].

The most frequently used splines are B-splines because of the computational efficiency provided by their short support. It has been shown that B-spline basis functions have the minimal support for a given order of approximation [3]. Cubic B-splines offer a good trade-off between the computational cost and the interpolation quality [1]. Also, B-splines are the preferred basis function because of their simple analytical form that facilitates manipulations [1]. Other interesting properties are that they are maximally continuously differentiable and that their derivatives can be computed recursively [1]. Thanks to the separability property of B-splines, the operations on multidimensional data can be performed by a successive processing of one-dimensional (1D) data along each dimension [1]. Besides, they have multiresolution properties that make them very useful for constructing wavelet bases and for multiscale processing [1, 4]. Because of all these properties, many image processing applications take advantage of B-splines.

In the first part of this chapter, we present the main theoretical results about splines which are of use in biomedical imaging applications (Section 6.2). First, we show how interpolation is related to sampling and the posterior reconstruction of the original signal from samples. When talking about spline interpolants, we can

Optical and Digital Image Processing: Fundamentals and Applications, First Edition. Edited by Gabriel Cristóbal, Peter Schelkens, and Hugo Thienpont.

distinguish between tensor product splines and polyharmonic splines. The former is based on the tensor product of two 1D spline functions, while the latter is based on the use of the so-called radial basis functions. Finally, we review the multiscale properties of splines that allow one to address problems in a coarse-to-fine fashion. These multiscale approaches are not only faster but also usually more robust to noise and yield smoother functionals when optimization problems are involved. In the second part of the chapter (Section 6.3), we illustrate the applications of splines in biomedical image processing by showing their use in rigid-body and elastic image and volume registration.

6.2
Main Theoretical Results about Splines

In this section, we present the main results of the spline theory that are used in biomedical imaging applications.

6.2.1
Splines as Interpolants and Basis Functions

6.2.1.1 Tensor Product Splines

The presentation of the theoretical properties of the splines is done in 1D space. However, the extension to two-dimensional (2D) space is readily performed using the tensor product. For instance, if $\varphi_{1D}(x)$ is a 1D spline, the 2D tensor product spline is defined as $\varphi_{2D}(x, y) = \varphi_{1D}(x)\varphi_{1D}(y)$.

The Interpolation Context Given a discrete set of measurements $(x_i,\ y_i)$, interpolating is the art of "filling in the gaps" with a continuous function $y = f(x)$ such that we meet the constraints imposed by our measurements, that is, $y_i = f(x_i)$. In biomedical imaging, our measurements are typically image values, $z_i = I(x_i,\ y_i)$, with z_i being the gray value of our image, and $(x_i,\ y_i)$ being its location in space. For color images, we could decompose the color into three components (red, green, and blue; hue, saturation, and value; etc.); each one would impose a constraint of the same kind as the one for gray values. Interpolating is important to know the value of our image between known pixel values. This is useful for rotations, translations, unwarpings, demosaicking, downsampling, upsampling, and so on.

Given N measurements, we can estimate a polynomial of degree $N-1$ (it has N coefficients and, therefore, N degrees of freedom) passing through all these measurements. In fact, this polynomial is unique and it is given by Newton's general interpolation formula:

$$\begin{aligned} f(x) = {} & a_0 + a_1(x - x_0) + a_2(x - x_0)(x - x_1) + \cdots \\ & + a_{N-1}(x - x_0)(x - x_1)\cdots(x - x_{N-2}) \end{aligned} \tag{6.1}$$

where a_n represents the divided differences defined as

$$\begin{aligned} a_0 &= y_0 \\ a_n &= \sum_{i=0}^{n} \frac{y_i}{\prod_{\substack{j=0 \\ j\neq i}}^{n} (x_i - x_j)} \end{aligned} \tag{6.2}$$

The use of polynomials for interpolation is justified because they are simple to manipulate, differentiate, or integrate; it is also justified by a theorem of Weierstrass that states that any continuous function on a closed interval can be approximated uniformly to any degree of accuracy by a polynomial of a sufficient degree [5].

Regularly Spaced Interpolation: The Generalized Sampling Theorem However, polynomials are not the only functions interpolating the dataset. If the x_i points are regularly distributed ($x_i = iT$, for some integer i and a sampling rate T), then Whittaker showed that the series

$$C(x) = \sum_{i=-\infty}^{\infty} y_i \operatorname{sinc}\left(\frac{x - x_i}{T}\right) \tag{6.3}$$

also interpolates the input measurements [6] $\left(\text{sinc is defined as } \operatorname{sinc}(x) = \frac{\sin(\pi x)}{\pi x}\right)$. $C(x)$ is called the *cardinal function*. Shannon [7] realized that this representation was unique for any function whose maximum frequency is smaller than $\frac{1}{2T}$ Hz. This is the famous Shannon sampling theorem which is valid for any function that is bandlimited.

The sampling theorem can be extended to a larger space of functions: the Hilbert space of L_2 functions, that is, all functions that are square integrable in the Lebesgue sense ($\|f\|^2 = \langle f, f\rangle < \infty$, where the inner product between two real functions is defined as $\langle f, g\rangle = \int_{-\infty}^{\infty} f(x)g(x)\mathrm{d}x$). The set of bandlimited functions is a subset of L_2. The sampling theorem is generally formulated in this space as

$$C(x) = \sum_{i=-\infty}^{\infty} c_i \varphi_i(x), \tag{6.4}$$

where c_i are some coefficients that have to be computed from the y_i input data, and $\varphi_i(x)$ is a shifted version of a basis function $\varphi(x)$ ($\varphi_i(x) = \varphi(x - x_i)$). In the particular case of bandlimited functions, the basis function used is $\varphi(x) = \operatorname{sinc}(\frac{x}{T})$ and the set of all bandlimited functions is spanned by the family of functions $\varphi_i(x)$. In other words, any bandlimited function can be expressed by a linear combination of the infinite set of $\varphi_i(x)$ functions as the sampling theorem shows. From now on, we drop the argument of functions as long as there is no confusion.

Let us consider any function φ in L_2, and the subspace V generated by its translations with step T:

$$V = \left\{ f(x) = \sum_{i=-\infty}^{\infty} c_i \varphi_i(x) : c_i \in l_2 \right\} \tag{6.5}$$

where l_2 is the space of all square-summable sequences. It can be easily proved that if a set of functions φ_i in a Hilbert space is orthonormal (i.e., $\langle \varphi_i, \varphi_j\rangle = \delta_{i-j}$), then

the projection of any function f in L_2 onto the subspace V is

$$\tilde{f} = P_V f = \arg\min_{g \in V} \|f - g\| = \sum_{i=-\infty}^{\infty} \langle f, \varphi_i \rangle \varphi_i \tag{6.6}$$

Note that

$$\langle f, \varphi_i \rangle = \int_{-\infty}^{\infty} f(x)\varphi(x - x_i)\mathrm{d}x = f(x) * \varphi(-x)\big|_{x=x_i} \tag{6.7}$$

that is, the inner product $\langle f, \varphi_i \rangle$ can be easily computed by sampling the linear convolution of $f(x)$ and $\varphi(-x)$ at $x = x_i$. In fact, the antialiasing filter used to effectively limit the frequency of a signal before sampling corresponds exactly to the computation of this inner product. Note that if f is already bandlimited, then the convolution of $f(x)$ with $\varphi(-x)$ (whose frequency response is just a rectangle with a maximum frequency of $\frac{1}{2T}$) is $f(x)$ and, thus, $c_i = y_i$.

The problem of using $\varphi(x) = \mathrm{sinc}\left(\frac{x}{T}\right)$ is that the sinc decays very slowly and the convolutions needed for the computation of the inner product are impractical. Thus, one might set off in search of new functions φ with more computationally appealing properties. First of all, one could relax the orthonormality condition and may ask only that the set $\{\varphi_i\}$ defines a Riesz basis, that is,

$$A \|c_i\|_{l_2}^2 \leq \left\| \sum_{i=-\infty}^{\infty} c_i \varphi_i(x) \right\|_{L_2}^2 \leq B \|c_i\|_{l_2}^2 \quad \forall c_i \in l_2 \tag{6.8}$$

for some constants A and B depending on φ. Orthonormality is a special case when $A = B = 1$. The left-hand side condition assures that the basis functions are linearly independent, while the right-hand side condition assures that the norm of f is bounded and, therefore, the subspace V is a valid subspace of L_2. An important requirement for the basis is that it is able to represent any function with any desired level of accuracy by simply diminishing T. It can be shown that this condition can be reformulated as the partition of unity [1] :

$$\sum_{i=-\infty}^{\infty} \varphi_i(x) = 1 \quad \forall x \tag{6.9}$$

i.e. the sum of all the basis functions (shifted versions of φ) is a constant function.

The function $\varphi(x) = \mathrm{sinc}\left(\frac{x}{T}\right)$ meets all these requirements and is, by far, the most widely known function due to the sampling theorem, besides the fact that it is infinitely supported, making convolutions more complicated. However, other functions also meet these requirements. The shortest function meeting these is the box function

$$\varphi(x) = \beta_0\left(\frac{x}{T}\right) = \begin{cases} 1 & |x| < \frac{T}{2} \\ 0 & |x| > \frac{T}{2} \end{cases} \tag{6.10}$$

which is actually the cardinal B-spline of degree 0. The cardinal B-spline of degree n is simply obtained by convolving $\beta_0(x)$ with itself n times. For instance, $\beta_1\left(\frac{x}{T}\right)$ is the triangular function defined between $-T$ and T, and $\beta_2\left(\frac{x}{T}\right)$ is a parabolic

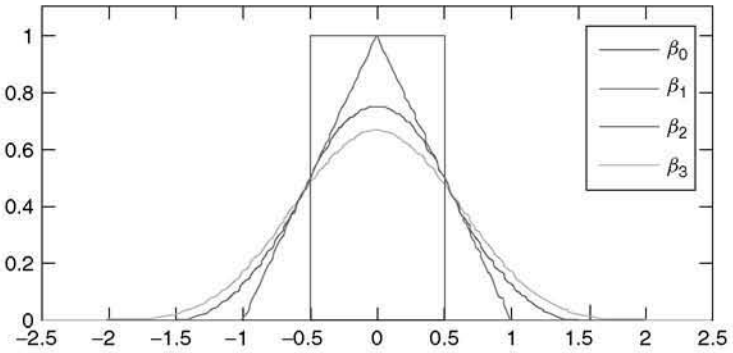

Figure 6.1 B-Splines Plot of the first four B-splines (degrees 0, 1, 2, and 3). (Please find a color version of this figure on the color plates.)

function defined between $-\frac{3}{2}T$ and $\frac{3}{2}T$. In general, $\beta_n\left(\frac{x}{T}\right)$ is an evenly symmetric, piecewise polynomial of degree n defined for $|x| < \frac{n+1}{2}T$. The following equations show the cardinal B-splines of degree 1, 2, and 3 for $T = 1$ (see also Figure 6.1):

$$\beta_1(x) = \begin{cases} 1 - |x| & |x| \leq 1 \\ 0 & |x| > 1 \end{cases} \tag{6.11}$$

$$\beta_2(x) = \begin{cases} \frac{3}{4} - |x|^2 & |x| \leq \frac{1}{2} \\ \frac{1}{2}\left(|x| - \frac{3}{2}\right)^2 & \frac{1}{2} < |x| \leq \frac{3}{2} \\ 0 & |x| > \frac{3}{2} \end{cases} \tag{6.12}$$

$$\beta_3(x) = \begin{cases} \frac{2}{3} + \frac{1}{2}|x|^2(|x| - 2) & |x| \leq 1 \\ \frac{1}{6}(2 - |x|)^3 & 1 < |x| \leq 2 \\ 0 & |x| > 2 \end{cases} \tag{6.13}$$

In general, the cardinal B-spline of degree n can be expressed as

$$\beta_n(x) = \frac{1}{n!}\sum_{k=0}^{n+1}(-1)^k\binom{n+1}{k}\left(x - \left(k - \frac{n+1}{2}\right)\right)_+^n \tag{6.14}$$

where $(x)_+^n$ is the one-sided n-th power of x, $(x)_+^n = \begin{cases} x^n & x \geq 0 \\ 0 & x < 0 \end{cases}$.

Cardinal B-splines are piecewise polynomials, because they can be represented by polynomials of degree n on $|x|$ that are different for each interval i $T \leq |x| < (i+1)$ T. All these functions meet the aforementioned requirements and therefore can be used as basis functions for the representation of signals. Moreover, they are well localized and compactly supported in real space (consequently, infinitely supported in Fourier space), making computations in real space affordable.

The representation of signals using cardinal B-splines is intimately related to interpolation: the use of β_0 is equivalent to nearest neighbor interpolation and the use of β_1 is equivalent to linear interpolation (bilinear interpolation when

the 2D tensor product B-spline is constructed). β_3 has proved to be a good compromise between computational complexity, compact support in real space, and approximation error.

At this point, we have already shown that the sampling theorem is a particular case of an interpolation problem and also shown how it can be generalized for L_2 functions instead of working exclusively with bandlimited functions. The question is how to compute the representation coefficients c_i in Eq. (6.4). We have seen a possibility for orthonormal basis in Eq. (6.7). However, cardinal B-splines are not orthonormal in general (only cardinal B-splines of degree 0 are orthogonal). In the general case, the calculation of the c_i coefficients is similar to the orthonormal case. The only difference is that, instead of using the basis φ_i itself, we have to use the dual basis $\tilde{\varphi}_i$:

$$\tilde{f} = P_V f = \sum_{i=-\infty}^{\infty} \langle f, \tilde{\varphi}_i \rangle \varphi_i \tag{6.15}$$

The dual basis is uniquely defined by the biorthogonality condition $\langle \tilde{\varphi}_i, \varphi_j \rangle = \delta_{i-j}$ and it also inherits the shift invariant property of the original basis function, $\tilde{\varphi}_i(x) = \tilde{\varphi}(x - x_i)$. However, we still do not have a clear way of computing the c_i's because we do not have a close-form formula for the dual basis. It can be proved [2] that the Fourier transform of $\tilde{\varphi}$ is given by

$$\widehat{\tilde{\varphi}}(j\Omega) = \frac{\widehat{\varphi}(j\Omega)}{\sum_{k=-\infty}^{\infty} \left|\widehat{\varphi}(j(\Omega + 2\pi k))\right|^2} \tag{6.16}$$

where $\widehat{f}(j\Omega)$ represents the continuous Fourier transform of the function $f(x)$. Fortunately, Unser and coworkers [8, 9] derived a very efficient way of computing these coefficients using standard digital filters. This is actually the way of producing these coefficients. These filters are derived in 1D. However, they are easily extended to n D. In the case of images, these filters are run individually on each row of the image, producing a new image of coefficients over the horizontal axis, x. Then, they are run on each column of the new image, finally producing the coefficients of the 2D tensor product B-spline. Once the coefficients are produced, images are treated as if they were continuous functions, although they are stored as a discrete set of cardinal B-spline coefficients.

Now, we may wonder if we could design a basis function based on B-splines such that $c_i = y_i$. This would be an interpolating spline and we would be back to a situation similar to the interpolation scheme presented in the sampling theorem, Eq. (6.3). The following function is such an interpolating spline:

$$\varphi_{\text{int}}(x) = \sum_{i=-\infty}^{\infty} q_{\text{int}}[i] \varphi_i(x) \tag{6.17}$$

where $q_{\text{int}}[i]$ is the l_2 sequence defined as the inverse Z transform of $Q_{\text{int}}(z) = \frac{1}{\sum_{k=-\infty}^{\infty} \varphi(kT) z^{-k}}$, and $\varphi(x) = \beta_n\left(\frac{x}{T}\right)$.

Approximation Error: How Far are We from the Truth? An important concept related to the generalized sampling theorem explained above is how well it reproduces any function f. Let us call f_T the approximation with a given sampling rate T. This problem has been studied by the approximation theory, proving that the following three statements are equivalent:

1) Let f be a sufficiently smooth function (f belongs to the Sobolev space W_2^L, i.e., its L first derivatives belong to L_2). Then, as T approaches 0, there exists a constant C independent of f such that $\|f - f_T\| \leq CT^L \|f^{(L)}\|$.
2) The first L moments of φ are constants, that is, $\sum_{i=-\infty}^{\infty} (x - x_i)^m \varphi_i(x) = \mu_m$ for $m = 0, 1, \ldots, L-1$.
3) The first L monomials can be exactly represented, that is, for each monomial x^m ($m = 0, 1, \ldots, L-1$), there exist constants $c_i \in l_2$ such that $x^m = \sum_{i=-\infty}^{\infty} c_i \varphi_i(x)$.

An important consequence of this result is that the approximation error of different basis functions depends mostly on their design, that is, given two polynomials of the same degree, one of them may approach smooth functions more quickly than the other. L is called the *order of approximation* and entirely depends on the moments of φ or the number of monomials that can be represented. In particular, B-splines of degree n have an order of approximation of $L = n + 1$. Another important consequence is that in order to converge to function f as $T \to 0$, the basis function φ must have $L \geq 1$, or in other words, that it fulfills the partition of unity. It is well known [10] that windowed sincs (which are commonly used as a solution to the infinite support of the sinc function) do not meet this condition. One may also try to design the φ family such that they have order of approximation L with a minimum support. This is how the MOMS (maximal-order interpolation of minimal support) set of functions is designed. It turns out that these functions are linear combinations of the cardinal B-spline of degree L and its derivatives [3].

The reader interested in the generalized sampling theorem and this approach to interpolation may gain information from Refs [1, 2, 10–12].

Back to Irregular Interpolation Problems So far, we have already introduced cardinal B-splines, dual splines, and interpolation splines. In fact, splines are a broad family of functions of which we have only seen those used with a regular spacing. In general, a function $S_n(x)$ is a spline of degree n if (i) it is defined by piecewise polynomials of degree at most n; (ii) it is of class C^{n-1}, that is, it has $n - 1$ continuous derivatives even at the points joining the different polynomial pieces [13].

An alternative approach to splines is based on the idea of curve interpolation and knots instead of the idea of sampling. This other approach allows for irregularly spaced samples in a much more direct way. Let us assume that we are interpolating a real function in the interval $[a, b]$ with a piecewise polynomial. We subdivide this interval into N adjacent pieces such that each piece is defined in the interval $[x_i, x_{i+1}]$. The subdivision is such that $x_0 = a$, $x_N = b$, $x_i < x_{i+1}$, and $[a, b] = \bigcup_{i=0}^{N-1} [x_i, x_{i+1}]$. The input samples (x_i, y_i) are called *knots* and they are fixed points

through which the interpolated polynomial must pass. Note that knots need not to be equally spaced in the interval $[a, b]$. We have to be specially careful in selecting the interpolating polynomials such that not only the spline is continuous but also all its derivatives up to the order of $n-1$ are continuous, even at the knots where the function on the left-hand side of the knot is defined by a certain polynomial, and on the right-hand side it is defined by a different polynomial (see Figure 6.2).

With $N+1$ knots, the spline is split into N intervals. A spline of degree n has $n+1$ coefficients in each interval; therefore, the spline has $(n+1)N$ degrees of freedom. Let us call $S_i(x)$ the polynomial of degree n in each interval ($i = 1, 2, \ldots N$). This spline must satisfy the following:

1) Interpolation of the knot values: $2N$ degrees of freedom

$$\begin{aligned} &S_1(x_0) = y_0 \\ &S_i(x_i) = y_i = S_{i+1}(x_i) \quad i = 1, 2, \ldots, N-1 \\ &S_N(x_N) = y_N \end{aligned} \tag{6.18}$$

2) Continuity of the $n-1$ derivatives: $(n-1)(N-1)$ degrees of freedom

$$\begin{aligned} &S_i^{(1)}(x_i) = S_{i+1}^{(1)}(x_i) \qquad i = 1, 2, \ldots, N-1 \\ &S_i^{(2)}(x_i) = S_{i+1}^{(2)}(x_i) \qquad i = 1, 2, \ldots, N-1 \\ &\ldots \\ &S_i^{(n-1)}(x_i) = S_{i+1}^{(n-1)}(x_i) \quad i = 1, 2, \ldots, N-1 \end{aligned} \tag{6.19}$$

However, there are still $n-1$ unfixed degrees of freedom. Depending on the way these degrees of freedom are defined, different kinds of splines are defined. For instance, a natural spline of degree $n = 3$ is one of the most common cases, in which the second derivative at the extremes is set to 0, $S_i^{(2)}(x_0) = S_i^{(2)}(x_N) = 0$. There are efficient algorithms for the solution of the resulting equation system [14].

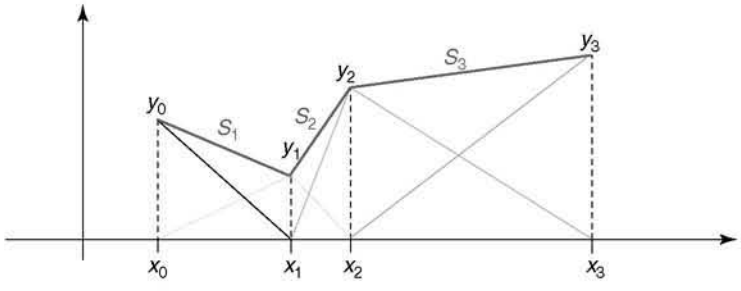

Figure 6.2 Spline interpolation with irregularly spaced samples. Example of B-spline (degree 1) interpolation of irregularly spaced samples. (Please find a color version of this figure on the color plates.)

The expression of the polynomial of degree n in any of the intervals $[x_i,\ x_{i+1})$ can be recursively constructed as

$$\begin{aligned} S_{i,0}(x) &= \begin{cases} 1 & x_i < x \leq x_{i+1} \\ 0 & \text{otherwise} \end{cases} \\ S_{i,l}(x) &= \frac{x - x_i}{x_{i+l} - x_i} S_{i,l-1}(x) + \frac{x_{i+l+1} - x}{x_{i+l+1} - x_i} S_{i+1,l-1}(x) \end{aligned} \tag{6.20}$$

where $S_{i,l}$ is the B-spline of degree l in the interval $[x_i,\ x_{i+1})$. It is common to repeat the extremes of the knots as many times as the finally desired spline degree. For example, the actual list of knots for degree 3 would be $\{x_0,\ x_0,\ x_0,\ x_0,\ x_1,\ x_2,\ \ldots,\ x_{N-1},\ x_N,\ x_N,\ x_N,\ x_N\}$. Because of the repeated knots we admit by convention in the previous formula that $\frac{0}{0} = 0$.

The polynomial in each one of the subintervals is of the general form

$$S_i(x) = \sum_{l=0}^{n} a_l x^l \tag{6.21}$$

In this expansion, we have used the fact that the set $\{1,\ x,\ x^2,\ \ldots,\ x^n\}$ is a basis of the polynomials of degree n. However, this is not an orthonormal basis, which causes numerical instabilities when solving for the a_l coefficients. Alternatively, we could have used any other basis of the polynomials of degree n, $\{P_0(x),\ P_1(x),\ P_2(x),\ \ldots,\ P_n(x)\}$:

$$S_i(x) = \sum_{l=0}^{n} a_l P_l(x) \tag{6.22}$$

Employing different polynomial basis gives rise to different kinds of splines: Bernstein polynomials are used in the Bézier splines; Hermite polynomials are used in the Hermite splines; basic splines are used in the B-splines; and so on.

6.2.1.2 Polyharmonic Splines

The extension to several dimensions can be done by the tensor product of 1D splines as seen in the previous section, or by the specific design of the so-called polyharmonic splines, among which the most famous is the thin-plate spline.

Let us assume that we have a set of input multivariate data points $(\mathbf{x}_i,\ y_i)$ we would like to interpolate $(\mathbf{x}_i \in \mathbb{R}^d,\ y_i \in \mathbb{R})$. The goal would be to find a hypersurface $y = f(\mathbf{x})$ such that the surface contains the input data points.

We look for our interpolant in the Beppo-Levi space $BL^{(2)}(\mathbb{R}^d)$, that is, the space of all L_2 functions from $\mathbb{R}^d$ to $\mathbb{R}$, such that its second derivative is also in L_2. This space is large enough to contain a suitable interpolator. In this space, we can define the rotationally invariant seminorm:

$$\|f\|^2 = \int_{\mathbb{R}^d} \sum_{\eta=1}^{d} \sum_{\xi=1}^{d} \frac{\partial^2 f}{\partial x_\eta \partial x_\xi} \mathrm{d}\mathbf{x} \tag{6.23}$$

Duchon [15] showed that the interpolant minimizing the just introduced seminorm in $BL^{(2)}(\mathbb{R}^d)$ is of the form

$$f(\mathbf{x}) = p_m(\mathbf{x}) + \sum_{i=1}^{N} c_i \phi_{d,k}(\|\mathbf{x} - \mathbf{x}_i\|) \tag{6.24}$$

where $\|\mathbf{x}\|$ denotes the standard Euclidean norm in $\mathbb{R}^d$, N is the number of measurements in the input data, c_i is a set of coefficients that need to be determined, k is any integer such that $2k > d$ (in fact, the approximation order is $k - \frac{d}{2}$ [16]; therefore, on the one hand, it is interesting to choose high values of k although, on the other hand, these result in numerical instabilities in the determination of the spline coefficients as is seen later; a trade-off between these two goals must be achieved), $\phi_{d,k}$ is a radial basis function, and $p_m(\mathbf{x})$ is a polynomial in $\mathbf{x}$ of degree at most m, which is given by $m = k - \left\lceil \frac{d}{2} \right\rceil$. The radial basis function $\phi_{d,k}$ is

$$\phi_{d,k}(r) = \begin{cases} r^{2k-d}\log(r) & \text{for even } d \\ r^{2k-d} & \text{for odd } d \end{cases} \tag{6.25}$$

The function $\phi_{2,2} = r^2 \log(r)$ is the so-called thin-plate spline in $\mathbb{R}^2$ and the minimization of the seminorm in $\mathbb{R}^2$ can be understood as the minimization of the bending energy of a thin sheet of metal that interpolates the input data points. In $\mathbb{R}^3$, $\phi_{3,2}$ is called the *biharmonic spline* and $\phi_{3,3}$ is called the *triharmonic spline*.

The interpolation equations $f(\mathbf{x}_i) = y_i$ do not fully determine the function f (note that we have N input data points, but N coefficients c_i for the radial basis functions and $\sum_{p=0}^{m} \binom{d}{p}$ coefficients for the polynomial). In fact, the polynomial of degree m comes from the fact that it is in the kernel of the seminorm and, therefore, the addition of any polynomial of degree m is "invisible" to the seminorm minimization (for instance, for the thin-plate spline case, $m = 1$, and all polynomials of degree 1 have null second derivatives). In this way, we have to impose extra conditions which generally are

$$\sum_{i=1}^{N} c_i q(\mathbf{x}_i) = 0 \tag{6.26}$$

for all polynomials $q(\mathbf{x})$ of degree at most m (in the case of the thin-plate spline, we would have to use $q(\mathbf{x}) = 1$, $q(\mathbf{x}) = x_1$, and $q(\mathbf{x}) = x_2$). Let us assume that the set of coefficients of the polynomial p are written in vector form as $\mathbf{p}$. Then, the polyharmonic interpolation can be solved by the following equation system

$$\begin{pmatrix} \Phi & \mathbf{P} \\ \mathbf{P}^t & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{c} \\ \mathbf{p} \end{pmatrix} = \begin{pmatrix} \mathbf{y} \\ \mathbf{0} \end{pmatrix} \tag{6.27}$$

where $\mathbf{c}$ and $\mathbf{y}$ are column vectors with the c_i coefficients and the y_i measurements, respectively. Φ is the $N \times N$ system matrix corresponding to the measurements, i.e., $\Phi_{ij} = \phi_{d,k}(\|\mathbf{x}_i - \mathbf{x}_j\|)$ and P is a matrix related to some basis of polynomials up to degree m. Let $\{p_1, p_2, ..., p_l\}$ be such a basis; then $P_{ij} = p_j(\mathbf{x}_i)$. For example, for the thin-plate spline case, we could define $p_1(\mathbf{x}) = 1$, $p_2(\mathbf{x}) = x_1$, and $p_3(\mathbf{x}) = x_2$, but

any other basis of polynomials of degree at most 1 would do. Note that the size of matrix P is $N \times l$ (in the thin-plate spline, $l = 3$).

This equation system is usually ill-conditioned due to the nonlocal nature of the $\phi_{d,k}(r)$ functions (they are not locally supported, instead they grow to infinity with growing r). Moreover, the complexity of solving the equation system depends on the number of sample points, N. Finally, the evaluation of the polyharmonic spline involves as many operations as the input data points (although fast algorithms have been developed for tackling this latter problem [17]). For solving the problem of the ill-conditioning of the equation system, a localization of the polyharmonic spline can be done. This is a process in which the noncompactly supported $\phi_{d,k}(r)$ function is substituted as a weighted sum of compactly supported functions (for instance, B-splines). For further information on this technique the reader is referred to Refs [18, 19].

In the case of noisy data, we can relax the interpolation condition by replacing it by a least-squares approximation condition:

$$f^* = \arg \min_{f \in BL^{(2)}(\mathbb{R}^d)} \lambda \|f\|^2 + \frac{1}{N} \sum_{i=1}^{N} (y_i - f(\mathbf{x}_i))^2 \tag{6.28}$$

It can be proved [20] that the coefficients $\mathbf{c}$ and $\mathbf{p}$ are the solutions of the following linear equation system:

$$\begin{pmatrix} \Phi - 8N\lambda\pi\mathbf{I} & \mathbf{P} \\ \mathbf{P}^t & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{c} \\ \mathbf{p} \end{pmatrix} = \begin{pmatrix} \mathbf{y} \\ \mathbf{0} \end{pmatrix} \tag{6.29}$$

One may wonder why these radial basis functions are called splines; at least, they do not seem to fit our previous definition of piecewise polynomial functions. The solution is a slight modification of our concept of spline (particularly, the requirement of being piecewise). Let us consider the triharmonic ($\phi_{3,3} = r^3$) functions. It is clear that it is a cubic polynomial in r, and its second derivative is continuous everywhere. Therefore, it is a cubic spline.

The reader interested in this topic is referred to Refs [16, 21, 22].

6.2.2 Splines for Multiscale Analysis

The multiscale capabilities of splines come in two different flavors: spline pyramids and spline wavelets. Each one of these approaches exploits a different feature of splines that makes them suitable for multiresolution analysis. Because of the space limitations imposed for the chapter, we only describe the multiresolution spline pyramids here. The reader interested in spline wavelets is referred to Refs [1, 23].

Spline Pyramids Let us assume that we know the representation of a certain function $f(x)$ with B-splines of odd degree n and a sampling rate T:

$$\begin{aligned} f(x) &= \sum_{i=-\infty}^{\infty} c_i \beta_n \left(\frac{x}{T} - i \right) = \left(\sum_{i=-\infty}^{\infty} c_i \delta (x - i) \right) * \beta_n \left(\frac{x}{T} \right) \\ &= c_{\amalg}(x) * \beta_n \left(\frac{x}{T} \right) \end{aligned} \tag{6.30}$$

One may wonder how a finer representation of f would be. For this, we consider the relationship between an odd degree B-spline and its contraction by M to a final sampling rate $\frac{T}{M}$. It can be shown [1] that

$$\beta_n\left(\frac{x}{T}\right) = \sum_{k=-\infty}^{\infty} h_k \beta_n\left(\frac{x}{\frac{T}{M}} - k\right) = h_{\sqcup\!\sqcup}(x) * \beta_n\left(\frac{x}{\frac{T}{M}}\right) \tag{6.31}$$

that is, we can decompose a wide B-spline of degree n as the weighted sum of thinner B-splines of the same degree. For any M and n, the Z transform of the weight sequence h_k is

$$H(z) = z^{\frac{(M-1)(n+1)}{2}} \frac{1}{M^n} \left(\sum_{m=0}^{M-1} z^{-m}\right)^{n+1} \tag{6.32}$$

The case $M = 2$ plays an important role in the design of wavelets, and the corresponding property is called the 2-scale relationship. In case of B-splines of degree 1, the two scale relationship is simply

$$\beta_1\left(\frac{x}{T}\right) = \frac{1}{2}\beta_1\left(\frac{x}{\frac{T}{2}} + 1\right) + \beta_1\left(\frac{x}{\frac{T}{2}}\right) + \frac{1}{2}\beta_1\left(\frac{x}{\frac{T}{2}} - 1\right) \tag{6.33}$$

and, in general, for splines of odd degree n and $M = 2$, we have

$$h_k = \begin{cases} 2^{-n}\binom{n+1}{k+\frac{n+1}{2}} & |k| \leq \frac{n+1}{2} \\ 0 & |k| > \frac{n+1}{2} \end{cases} \tag{6.34}$$

Substituting the expression of the wide B-splines by the weighted sum of fine B-splines, we obtain

$$f(x) = \left(\left(\uparrow_M \left\{c_{\sqcup\!\sqcup}(x)\right\} * h\right)_{\sqcup\!\sqcup}(x)\right) * \beta_n\left(\frac{x}{\frac{T}{M}}\right) \tag{6.35}$$

In other words, to obtain a finer representation of a signal, we simply have to upsample its B-spline coefficients and convolve them with a finite weight sequence depending on the scaling factor M and the spline degree. Note that the function represented with splines at the finer resolution is exactly the same as the original one. No interpolation has been performed on the way.

Creating a coarser representation of the function $f(x)$ is a little bit more involved since we cannot have exactly the same function but an approximation to it:

$$\tilde{f}(x) = \tilde{c}_{\sqcup\!\sqcup}(x) * \beta_n\left(\frac{x}{MT}\right) \tag{6.36}$$

and we have to devise a way of estimating the coefficients $\tilde{c}$ from the c coefficients. The easiest way is to look for the $\tilde{c}$ that minimize the L_2 norm of the error $\left\|f - \tilde{f}\right\|$. It can be proved [24] that the solution to this least-squares problem is

$$\tilde{c} = \frac{1}{2}\left(\left(b_k^{2n+1}\right)^{-1} * \downarrow_M \left\{h_k * b_k^{2n+1} * c_k\right\}\right), \tag{6.37}$$

with $\downarrow_M \{\cdot\}$ being the downsampling operator, c_k being the B-spline coefficients of the function f with sampling rate T, h_k being the sequence described in Eq. (6.34), and b_k^{2n+1} being the sequence $b_k^{2n+1} = \beta_{2n+1}(k)$. Note that $\left(b_k^{2n+1}\right)^{-1}$ is the inverse of this sequence. It can be easily understood by inverting the Z transform of the sequence b_k^{2n+1} and then performing an inverse Z transform. While b_k^{2n+1} is compactly supported, $\left(b_k^{2n+1}\right)^{-1}$ is not, and convolution with this sequence requires the design of an IIR (infinite impulse response) filter [9, 25].

6.3 Splines in Biomedical Image and Volume Registration

In this section, we show two examples of the use of splines for image and volume registration, which is one of the most challenging tasks in biomedical image processing.

The intensity-based registration can be viewed in an optimization framework in which the registration problem consists in searching for a geometric transformation of a source image/volume that gives the image/volume that best matches a target image/volume, under a chosen similarity measure. The restriction of the motion to rigid-body motion means that the distance between the points of the object is the same in the registered source and target images/volumes. Elastic registration is frequently employed in medical image analysis to combine data that describe anatomy, both because biological tissues are in general not rigid and because anatomy varies between individuals. In the intensity-based elastic registration techniques, the solution of the registration problem is the deformation field that warps the source image/volume so that the resulting image/volume best matches the target image/volume. The registration is achieved by minimizing a cost function, which represents a combination of the cost associated with the image/volume similarity and the cost associated with the smoothness of the transformation (regularization term).

Many authors have proposed to use linear combinations of B-splines placed on a regular grid to model the transformation [27–31]. The available techniques differ in the form of employed regularization term, as well as in the employed image/volume similarity metrics and the optimization method. They produce good results but have a high computational cost. The computation can be accelerated using multiscale image/volume processing, and spline pyramids provide a convenient tool for this. Moreover, spline model can be used for all computation aspects of the registration (image pyramid, transform, and the gradient of the optimization criterion) as shown in Refs [4, 31, 32].

Many examples of rigid-body image registration can be found in 3D electron microscopy [34]. Indeed, the structure of a macromolecular complex can be computed in three dimensions from a set of parallel-beam projection images of the same complex acquired in a microscope [34]. For the so-called single-particle analysis, images of the sample containing thousands of copies of the same complex are collected. Ideally, the copies have the same structure. Their orientation in the

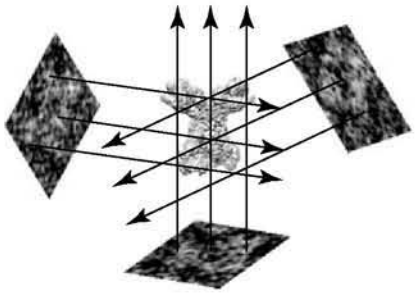

Figure 6.3 3D electron microscopy 3D electron microscopy. Experimental images corresponding to three arbitrarily chosen projection directions and a 3D model of the phosphorylase kinase 26. The arrows represent the directions of back projection in 3D space.

3D space is random and unknown, and the position of the center of each complex is unknown. These parameters have to be estimated before applying a method for 3D reconstruction. Given a first guess for the 3D model, one estimates the unknown parameters by aligning the images with the 3D model (reference model) [30] (Figure 6.3). A new reconstruction is then computed using the images and the estimated orientations and positions. The new model can be used as the reference model to resume the alignment in the next iteration of the iterative refinement of the estimated parameters [34]. It has been shown that such procedures can yield 3D models of subnanometer resolution [26].

An illustration of elastic image registration is given in Figure 6.4. One of the major difficulties in the analysis of electrophoresis 2D gels is that the gels are affected by spatial distortions due to run-time differences and dye-front deformations, which results in images that significantly differ in the content and geometry. The method proposed in Ref. [33] models the deformation field using B-splines, the advantage of which is that the model can be adapted to any continuous deformation field simply by changing the spacing between splines. The method computes quasi-invertible deformation fields so that the source image can be mapped onto the target image and vice versa, which helps the optimizer to reduce the chance of getting trapped in a local minimum and allows the simultaneous registration of any number of images.

6.4 Conclusions

In this chapter, we reviewed spline interpolation and approximation theory by presenting two spline families: tensor product splines and polyharmonic splines. Also, we presented the multiscale properties of splines. Finally, we illustrated biomedical image processing applications of splines by showing their use in image/volume registration.

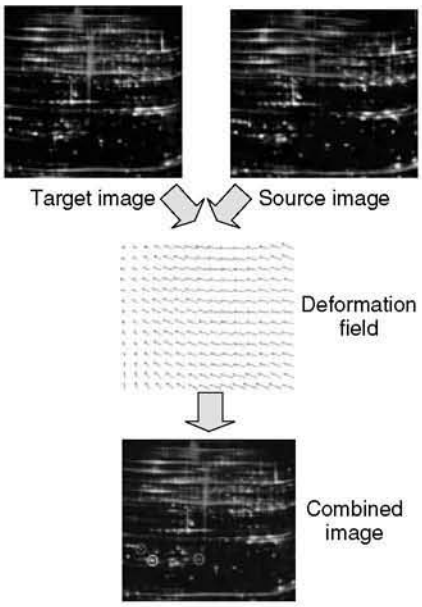

Figure 6.4 Elastic registration of 2D gels Elastic registration of two 2D protein gels (source and target) using the method proposed in [33]. The deformation field represents the continuous deformation required to convert the source into the target. The combined image shows the target image in the red channel and the warped source in the green channel. Note that there are red spots (proteins expressed in the target image and not expressed in the source image), green spots (just the opposite), and yellow spots (proteins equally expressed in both images). (Please find a color version of this figure on the color plates.)

References

1. Unser, M. (1999) Splines: a perfect fit for signal and image processing. *IEEE Signal Process. Mag.*, **16**, 22–38.
2. Unser, M. (2000) Sampling 50 years after Shannon. *Proc. IEEE*, **88**, 569–587.
3. Blu, T., Thevenaz, P., and Unser, M. (2001) MOMS: maximal-order interpolation of minimal support. *IEEE Trans. Image Process.*, **10**, 1069–1080.
4. Thevenaz, P., Ruttimann, U., and Unser, M. (1998) A pyramid approach to subpixel registration based on intensity. *IEEE Trans. Med. Imaging*, **7**, 27–41.
5. Weierstrass, K. (1884) Über die analytische Darstellbarkeit sogenannter willkürlicher Funktionen einer reellen Veränderlichen. *Sitzungsber. Königlich Preußischen Akad. Wiss. Berlin*, 633–789.
6. Whittaker, E.T. (1914) On the functions which are represented by expansion of the interpolation theory. *Proc. Roy. Soc. Edinb.*, **A35**, 181–194.
7. Shannon, C.E. (1949) Communication in the presence of noise. *Proc. Inst. Radio Eng.*, **37**, 10–21.
8. Unser, M., Aldroubi, A., and Eden, M. (1993) B-Spline signal processing: part I theory. *IEEE Trans. Signal Process.*, **41**, 821–832.
9. Unser, M., Aldroubi, A., and Eden, M. (1993) B-Spline signal processing: part II efficient design and applications. *IEEE Trans. Signal Process.*, **41**, 834–848.
10. Meijering, E. (2002) A chronology of interpolation: from ancient astronomy to modern signal and image

processing. *Proc. IEEE*, **90**, 319–342, DOI: 10.1109/5.993400.

11. Eldar, Y. and Michaeli, T. (2009) Beyond bandlimited sampling. *IEEE Signal Process. Mag.*, **26**, 48–68, DOI: 10.1109/MSP.2009.932125.
12. Thevenaz, P., Blu, T., and Unser, M. (2000) Interpolation revisited. *IEEE Trans. Med. Imaging*, **19**, 739–758.
13. Schoenberg, I.J. (1946) Contributions to the problem of approximation of equidistant data by analytic functions. *Quart. Appl. Math.*, **4**, 45–99, 112–141.
14. Piegl, L. and Tiller, W. (1997) *The NURBS Book*, Springer.
15. Duchon, J. (1977) Splines minimizing rotation-invariant semi-norms in Sobolev spaces, in *Constructive Theory of Functions of Several Variables. Lecture Notes in Mathematics*, **571** (eds. W. Schempp and K., Zeller), Springer-Verlag, pp. 85–100.
16. Iske, A. (2003) On the approximation order and numerical stability of local Lagrange interpolation by polyharmonic splines. Proceedings of 5th International Conference on Modern Developments in Multivariate Approximation, Witten-Bommerholz, Germany.
17. Beatson, R.K., Powell, M.J.D., and Tan, A.M. (2007) Fast evaluation of polyharmonic splines in three dimensions. *IMA J. Numer. Anal.*, **27**, 427–450.
18. Kybic, J. (2001) Elastic image registration using parametric deformation models. Ph.D. thesis, EPFL.
19. Van de Ville, D., Blu, T., and Unser, M. (2005) Isotropic polyharmonic B-splines: scaling functions and wavelets. *IEEE Trans. Image Process.*, **14**, 1798–1813.
20. Wahba, G. (1990) *Spline Models for Observational Data*, SIAM, Philadelphia
21. Carr, J. *et al.* (2001) *Reconstruction and representation of 3D objects with radial basis functions*. Proceedings of SIGGRAPH, pp. 67–76.
22. Madych, W.R. and Nelson, S.A. (1990) Polyharmonic cardinal splines. *J. Approx. Theory*, **60**, 141–156.
23. Strang, G. and Nguyen, T. (1997) *Wavelets and Filter Banks*, Wellesley-Cambridge Press.
24. Thevenaz, P. and Millet, P. (2001) Multiresolution imaging of in vivo ligand-receptor interactions. Proceedings of the SPIE International Symposium on Medical Imaging, San Diego.
25. Unser, M., Aldroubi, A., and Eden, M. (1993) The L_2 polynomial spline pyramid. *IEEE Trans. Pattern Anal. Mach. Intell.*, **15**, 364–378.
26. Venien-Bryan, C. *et al.* (2009) The structure of phosphorylase kinase holoenzyme at 9.9 angstroms resolution and location of the catalytic subunit and the substrate glycogen phosphorylase. *Structure*, **17**, 117–127.
27. Rueckert, D. *et al.* (1999) Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Trans. Med. Imaging*, **18**, 712–721.
28. Studholme, C., Constable, R., and Duncan, J. (2000) Accurate alignment of functional EPI data to anatomical MRI using a physics-based distortion model. *IEEE Trans. Med. Imaging*, **19**, 1115–1127.
29. Kybic, J. *et al.* (2000) Unwarping of unidirectionally distorted EPI images. *IEEE Trans. Med. Imaging*, **19**, 80–93.
30. Jonic, S. *et al.* (2005) Spline-based image-to-volume registration for three-dimensional electron microscopy. *Ultramicroscopy*, **103**, 303–317.
31. Jonic, S. *et al.* (2006) An optimized spline-based registration of a 3D CT to a set of C-arm images. *Int. J. Biomed. Imaging*, 12, article ID 47197, DOI: 10.1155/IJBI/2006/47197.
32. Jonic, S. *et al.* (2006) Spline-based 3D-to-2D image registration for image-guided surgery and 3D electron microscopy, in *Biophotonics for Life Sciences and Medicine* (eds. M. Faupel, P. Smigielski, A. Brandenburg, and J. Fontaine), Fontis Media, Lausanne VD, Switzerland, pp. 255–273.
33. Sorzano, C. *et al.* (2008) Elastic image registration of 2-D gels for differential and repeatability studies. *Proteomics*, **8**, 62–65.
34. Frank, J. (1996) *Three-Dimensional Electron Microscopy of Macromolecular Assemblies*, Academic Press.