Generalized Sampling: A Variational Approach—Part I: Theory

Jan Kybic, *Member, IEEE*, Thierry Blu, *Member, IEEE*, and Michel Unser, *Fellow, IEEE*

***Abstract*—We consider the problem of reconstructing a multidimensional vector function $\mathbf{f}_{\text{in}}: \mathbb{R}^m \rightarrow \mathbb{R}^n$ from a finite set of linear measures. These can be irregularly sampled responses of several linear filters. Traditional approaches reconstruct in an *a priori* given space, e.g., the space of bandlimited functions. Instead, we have chosen to specify a reconstruction that is optimal in the sense of a quadratic plausibility criterion J. First, we present the solution of the generalized interpolation problem. Later, we also consider the approximation problem, and we show that both lead to the same class of solutions.**

Imposing generally desirable properties on the reconstruction largely limits the choice of the criterion J. Linearity leads to a quadratic criterion based on bilinear forms. Specifically, we show that the requirements of translation, rotation, and scale-invariance restrict the form of the criterion to essentially a one-parameter family. We show that the solution can be obtained as a linear combination of generating functions. We provide analytical techniques to find these functions and the solution itself. Practical implementation issues and examples of applications are treated in a companion paper.

***Index Terms*—Reconstruction, sampling, thin-plate splines, variational criterion.**

I. Introduction

A. Sampling and Reconstruction

RECONSTRUCTING a signal from its samples is one of the most fundamental tasks in signal processing. The classical sampling theorem presented by Shannon [2] states that a bandlimited function f_{in} (whose frequency spectrum is limited by the Nyquist frequency $\omega_{\max} = \pi/T$) can be reconstructed perfectly from its regularly-spaced (ideal) samples $s_j = f_{\text{in}}(jT)$ by convolution with a sinc kernel

$$f_{\text{out}}(x) = f_{\text{in}}(x) = \sum_{j\in\mathbb{Z}} s_j \operatorname{sinc}(x/T - j) \quad \text{where } \operatorname{sinc}(x) = \frac{\sin(\pi x)}{\pi x}. \tag{1}$$

In 1977, Papoulis [3] showed that it was also possible to recover f_{in} from the output of q linear shift-invariant filters sampled at $(1/q)$th the Nyquist rate. This has generalized Shannon's theory to *multichannel* and *nonideal* generalized sampling that is closer to reality than Shannon's ideal one. Papoulis' theory has been extended further to multidimensional [4] ($m > 1$) and vector [5] ($n > 1$) bandlimited functions $\mathbf{f}_{\text{in}}: \mathbb{R}^m \rightarrow \mathbb{R}^n$. Recent applications of generalized sampling include, among others, deinterlacing [6], [7], and super-resolution [8], [9] reconstruction.

Unser and Aldroubi [10] replaced perfect reconstruction by the weaker condition of *consistency*, requiring that the reconstructed signal f_{out} provides exactly the same measurements as the original signal f_{in} when run through the measurement system. The reconstruction should also be unique; this depends on the reconstruction space V, the measurement system, and sampling locations. Their reconstruction formula is a generalized case of (1). See [11] for the multichannel case. A generic linear measurement system (generalized sampling) with a consistent reconstruction is shown in Fig. 1.

In nonuniform sampling, the location of measurement points is irregular, either because of the lack of control of the measurement process or because some domain needs more attention. Examples include shape reconstruction [12] or landmark interpolation [13]–[16]. The reconstruction can be done within the class of bandlimited functions [17], [18] or more general wavelet and spline-like spaces [19].

For an extensive review on sampling, see [20] and [21].

B. Related Work

The work presented in this paper can be seen as an extension of the theory of radial-basis function approximation [22], [23], especially Duchon's thin-plate splines [24], [25] to vector functions, nonideal (generalized) sampling, and generating functions that need not be radial. An alternative extension of the thin-plate splines and multiquadrics theory is found in [26] and [27], including error bounds.

There is also a close link with the variational formulation of splines [28], [29], which can be derived from the presented theory in the one-dimensional case. The related case of multichannel sampling in spline spaces is treated in [11] and [30], where tempered splines were also used [31]. Generalized sampling has been studied in the wavelet [32] and spline [10] bases in the case of nonuniform sampling locations as well [19]. Related techniques include nonseparable wavelets [33], vector-valued wavelets [34], or box-splines [35]; however, we are not aware of them having been applied explicitly to sampling.

C. Variational Reconstruction

The reconstruction method presented in this paper has been designed to be as general as possible from several possible viewpoints. It can handle multidimensional and vector functions.

Manuscript received April 11, 2001; revised May 6, 2002. The associate editor coordinating the review of this paper and approving it for publication was Prof. Fredrik Gustafsson.

J. Kybic was with the Biomedical Imaging Group, DMT/IOA, Swiss Federal Institute of Technology Lausanne, Switzerland. He is now with Robotvis, INRIA, Sophia Antipolis, France (e-mail: kybic@ieee.org).

T. Blu and M. Unser are with the Biomedical Imaging Group, DMT/IOA, Swiss Federal Institute of Technology Lausanne, Switzerland.

Publisher Item Identifier 10.1109/TSP.2002.800391.

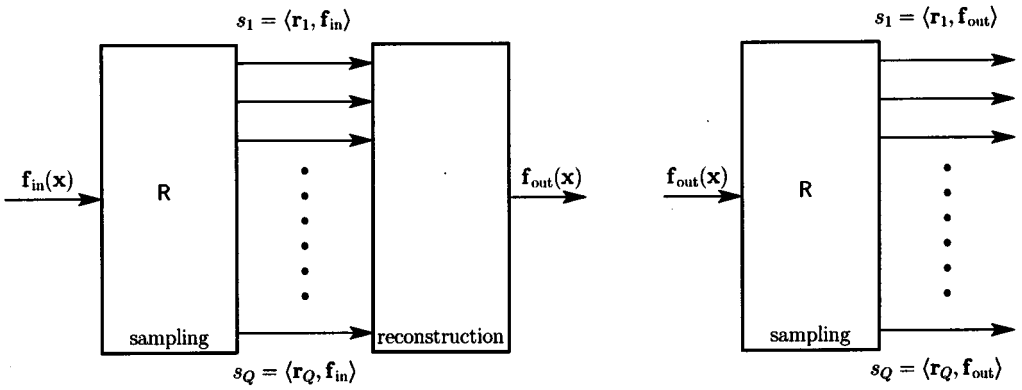

Fig. 1. Generalized sampling converts a vector input function $\mathbf{f}_{\text{in}}(\mathbf{x})$ into a set of scalar measures s_i by taking scalar products with measurement devices $\mathbf{R} = [\mathbf{r}_1 \cdots \mathbf{r}_Q]$. These measures are input into the reconstruction algorithm (which we are trying to develop) to produce a reconstruction $\mathbf{f}_{\text{out}}(\mathbf{x})$. The consistency statement requires that the sampling applied on $\mathbf{f}_{\text{out}}(\mathbf{x})$ and $\mathbf{f}_{\text{in}}(\mathbf{x})$ provide the same measures.

It can use arbitrary linear measurements (nonideal samples), which may, for example, be obtained by sampling the output of a multichannel filterbank (like sampling of Papoulis [3]). We put no bandlimiting restrictions on the input signal. The reconstruction is stable and unique for a large class of sampling configurations.

We retain the idea of a consistent reconstruction. However, we will not specify the reconstruction space beforehand. Rather, this space will be determined naturally from the problem at hand based on a continuous *regularization*. We introduce a non-negative *smoothness criterion* (penalty function) $J(\mathbf{f})$, which gets smaller as the function $\mathbf{f}$ gets smoother. We then seek a function $\mathbf{f}$ minimizing this criterion under the consistency constraints (introduced in Section I-A), e.g., passing through given points. In other words, we replace the subspace constraint $\mathbf{f} \in V$ by a variational formulation. The criterion $J(\mathbf{f})$ provides the regularization needed to overcome the ambiguity of the reconstruction problem. It may also represent an *a priori* knowledge in the Bayesian framework, quantifying our confidence that a particular function $\mathbf{f}$ is close to the input $\mathbf{f}_{\text{in}}$ [36]. Our regularization is completely specified in the continuous domain, unlike alternative methods that often use discretized version of the regularization operator [37]–[39].

The basic problem is therefore to reconstruct a signal from a series of linear measurements. This leads to a functional minimization problem under linear constraints. We will concentrate on minimizing quadratic energy functionals as this yields a vector space characterization of the solution as a linear combination of basis functions. The key feature here is that the basis functions themselves are the result of a mathematical optimization. Consequently, they are optimally tailored to the problem at hand.

In this paper, we present the mathematical foundations of the method. More practical aspects of generalized sampling and variational reconstruction are treated elsewhere [1]. This companion paper presents the computational recipes for the method and provides a number of examples illustrating the wide applicability of the present formulation. It may, therefore, also be a good starting point for those more interested in results than in mathematical derivations.

D. Motivation

This paper has four primary goals.

1) *To provide a precise mathematical formulation of generalized sampling in a variational setting.* This is done in Section II-C, where we also state our assumptions and list some of the general properties of the solution.
2) *To understand and control the key properties of the solution through an appropriate selection of the regularization criterion.* To this end, we investigate quadratic shift-invariant criteria and their corresponding bilinear forms. In Section III, we provide the corresponding convolutional kernel representation in both time and frequency domains. Our strategy is to impose some desirable properties on the solution (enumerated in Section II-D) and to infer the corresponding class of criteria. We find that a small set of perfectly justifiable requirements, such as rotation and scale invariance, essentially limits the degrees of freedom to a one-parameter family of criteria. This is formalized in Theorem 1 at the end of Section IV.
3) *To solve our generalized sampling problem under hard constraints (consistency requirement).* The general solution is derived in Section IV and described in Theorem 2. We show how to construct a basis for the solution space. The critical step involves finding the Green's functions of the operator associated with the bilinear form of the criterion. The solution usually includes an additional polynomial term whose main effect is to make the reconstruction well behaved far from the sampling points. These results lead to the specification of the linear system of equations that yields the optimal coefficients for the solution.
4) *To solve our generalized sampling problem under soft constraints.* The idea here is to consider a cost function that is the sum of a nonlinear data term and the same regularization criterion as before. In Section V, we prove that the solution of this approximation problem—irrespective of the form of the data term—lies in the same subspace as in the previous case (hard constraints) (cf. Theorem 3). We also work out an explicit formula for the least-squares case. Interestingly, this solution can be obtained by an almost trivial modification of the hard-constrained equations (addition of constant diagonal term to the system

matrix). The advantage of this approach is that it stabilizes the reconstruction. It is also better suited for noisy situations where it is often counterproductive to impose hard constraints.

II. Formulation

A. Notation

We denote vectors by bold letters and consider them as columns, that is, $\mathbf{x} = [x_1 \cdots x_m]^T$. Matrices will be denoted by upright letters ($\mathbf{X}$) with elements $(\mathbf{X})_{i,j} = x_{i,j}$.

We define a scalar product of two multivariate vector functions as $\langle \mathbf{f}, \mathbf{g} \rangle = \int_{\mathbb{R}^m} \mathbf{f}(\mathbf{x})^T \mathbf{g}(\mathbf{x})\, \mathrm{d}\mathbf{x}$. By extension, the notation $\langle \mathbf{X}, \mathbf{f} \rangle$ applied on a matrix $\mathbf{X}$ and a vector $\mathbf{f}$ is a vector of scalar products between columns of $\mathbf{X}$ and $\mathbf{f}$. Similarly, we define a convolution of vector and matrix functions following the usual rules for matrix multiplication; for example, $\mathbf{X} * \mathbf{y} = \mathbf{z}$ means $\sum_j x_{ij} * y_j = z_i$.

We will denote $\hat{f}$, $\mathcal{F}f$ as the Fourier transform of f, $\hat{f}(\boldsymbol{\omega}) = \int f(\mathbf{x}) e^{-i\boldsymbol{\omega}^T \mathbf{x}}\, \mathrm{d}\mathbf{x}$. The scalar product $\langle \mathbf{f}, \mathbf{g} \rangle$ corresponds to an L_2 norm $\|\mathbf{f}\|_{L_2} = \langle \mathbf{f}, \mathbf{f} \rangle^{1/2}$.

B. Distributions

Many results in this paper are obtained through calculations in the sense of *distributions*. The basic reference here is in [40], but a more accessible introduction is in [41]. A distribution u is a function-like object defined indirectly through its scalar products $\langle u, v \rangle$ with arbitrary *test functions* v from the space $\mathcal{D}$ of compactly supported and infinitely differentiable functions. Since there will be no ambiguity, we use the same symbol $\mathcal{D}$ for vector test functions as well. We say that two distributions u, v are equal if for all $w \in \mathcal{D}$ we have $\langle u, w \rangle = \langle v, w \rangle$. This is weaker than the usual point-wise equality. Distributions are generalizations of functions and can often be operated on using the same rules except, and this is noteworthy, the multiplication. The majority (but not all) of the practically used functions are indeed distributions. The best-known example of a distribution that is not a function is Dirac's δ, which is defined as $\langle \delta, v \rangle = v(0)$. Similarly, its derivative δ' gives $\langle \delta', v \rangle = -v'(0)$. Consequently, convolving $\delta * v$ yields v, whereas $\delta' * v = v'$.

We use an extension of the Fourier transform to a subset of distributions called *tempered* (such as polynomials) through the definition $\langle \hat{u}, v \rangle = \langle u, \hat{v} \rangle$.

C. Problem Definition

The variational problem we consider consists of finding a vector function $\mathbf{f}\colon \mathbb{R}^m \to \mathbb{R}^n$ minimizing a non-negative functional criterion $J(\mathbf{f})$ under a finite number Q of constraints $\langle \mathbf{r}_i, \mathbf{f} \rangle = s_i$, where $i = 1 \cdots Q$ and $\mathbf{r}_i$ correspond to sampling devices. The expression $\langle \mathbf{r}_i, \mathbf{f} \rangle$ linearly maps functions to real scalars. As $\mathbf{r}_i$ is a distribution, most linear forms can be written in this form.

When $\mathbf{f}$ satisfies all the Q constraints $\langle \mathbf{r}_i, \mathbf{f} \rangle = s_i$, we write

$$\langle \mathbf{R}_{n \times Q}, \mathbf{f} \rangle = \mathbf{s}. \tag{2}$$

We only consider functions from a space F induced by the criterion J and measurable by the devices $\mathbf{R}$

$$F = \{\mathbf{f}\colon \mathbb{R}^m \to \mathbb{R}^n;\ \ J(\mathbf{f}) < \infty \ \text{ and } \ \|\langle \mathbf{R}, \mathbf{f} \rangle\| < \infty\} \tag{3}$$

where $\|\cdot\|$ is the usual Euclidean norm of vectors in $\mathbb{R}^Q$.

Definition 1 (Generalized Interpolation Problem): We say that $\mathbf{f}_{\text{out}}$ solves the problem $\mathcal{P}(J, \mathbf{R}, \mathbf{s})$ iff $\mathbf{f}_{\text{out}}$ minimizes J in F under constraints (2). □

Note that there can be more than one of such functions $\mathbf{f}_{\text{out}}$ with the same value of $J(\mathbf{f}_{\text{out}})$.

Generalized interpolation problems arise whenever we need to reconstruct a continuous function from linear measures. The companion paper [1] gives examples of several such problems. There, we will look at a more structured system where the measurements are obtained from the samples of a multichannel filterbank, i.e., $s_{ij} = (\mathbf{h}_i^T * \mathbf{f})(\mathbf{x}_{ij})$. For the time being, however, we prefer to work with the more general formulation (2), which simplifies the notation and the mathematical derivations. Later on, will turn back to the multichannel system and take advantage of the convolutional form of the measurement process to simplify the description and implementation of the solution (cf. [1, Sect. II]).

D. Properties of the Solution

In order for our variational approach to be useful in the context of sampling and reconstruction [1], the solution should satisfy a certain number of properties, which will in turn impose constraints on the criterion J and the devices $\mathbf{r}_i$. We will see that the properties detailed later help us to specify an essentially one-parameter family of criteria. We will be able to give a constructive theorem concerning the existence property, obtain unicity in the majority of useful cases, and guarantee the invariance and linearity of the solution in the sense we detail further on.

For each property, we give an indication of how it can be verified or guaranteed. Note, however, that the conditions we give are only sufficient, but not necessary, because searching for necessary conditions proved to be extremely difficult and of small practical interest. On the other hand, we will see in the forthcoming sections that our conditions yield a sufficiently general family of criteria.

Property 1 (Existence and Uniqueness): There is exactly one solution $\mathbf{f}_{\text{out}}$. □

The motivation of the existence requirement is clear: We want our method to give us at least one solution for any possible measurements s_i. There are various reasons why the problem $\mathcal{P}$ might not have a solution, e.g., when the constraints are contradictory or when the solution space is not complete with respect to J. That is to say, if for any sequence of functions $\mathbf{f}_i$ satisfying the constraints such that the criterion $J(\mathbf{f}_i)$ is decreasing, this sequence does not converge in F. A typical example might be a sequence of continuous functions converging toward a discontinuous one, under a derivative criterion.

We also want the solution to be unique. For the uniqueness, it is useful for J to be discriminative so that as few functions $\mathbf{f}$ as possible have the same criterion value J.

In practice, we verify existence and uniqueness *a posteriori*. We first construct a function and then verify that it solves the problem and that no other function does. In some cases, the work on the *a priori* analysis has been already done [23], [42].

Property 2 (Vector Space of Solutions): If $\mathbf{f}$ solves $\mathcal{P}(J, \mathbf{R}, \mathbf{s})$ and $\mathbf{g}$ solves $\mathcal{P}(J, \mathbf{R}, \mathbf{t})$, then $\alpha\mathbf{f} + \beta\mathbf{g}$ solves $\mathcal{P}(J, \mathbf{R}, \alpha\mathbf{s} + \beta\mathbf{t})$ for $\alpha, \beta \in \mathbb{R}$. □

This scalar linearity property ensures that the solution space is a vector space and that consequently, every solution can be expressed as a linear combination of Q basis functions (where Q is the number of constraints). We will see later (see Section III-A) that this is ensured if the criterion J is a quadratic criterion.

Property 3 (Matrix Linearity): If $\mathbf{f}$ solves $\mathcal{P}(J, A^T\mathbf{R}, \mathbf{s})$, then $\mathbf{A}\mathbf{f} + \mathbf{b}$ solves $\mathcal{P}(J, \mathbf{R}, \mathbf{s} + \langle \mathbf{R}, \mathbf{b}\rangle)$, where $\mathbf{A}$ is an arbitrary matrix. □

This comprises the cases of rotating, scaling, shifting, permuting, inverting, and otherwise linearly deforming the "output" coordinate system of the function $\mathbf{f}$. We want the solution to be invariant with respect to these changes. The matrix linearity property is guaranteed if J is *pseudo-invariant* with respect to these changes, which means that the criterion value for $\mathbf{A}\mathbf{f} + \mathbf{b}$ is proportional to the criterion value for $\mathbf{f}$, namely

$$J(\mathbf{A}\mathbf{f} + \mathbf{b}) = c(\mathbf{A}, \mathbf{b})J(\mathbf{f}) \tag{4}$$

where c is a continuous function of $\mathbf{A}$ and $\mathbf{b}$, independent of $\mathbf{f}$. For the remainder of this paper, we consider only matrices satisfying

$$\mathbf{A}\mathbf{A}^T = k\mathbf{I} \tag{5}$$

(where $\mathbf{I}$ is the identity matrix), which corresponds to orthogonal transformations and uniform scaling. See Section IV-D for details.

Property 4 (Geometric Invariance): If $\mathbf{g}$ solves $\mathcal{P}(J, \mathbf{R}, \mathbf{s})$, then $\mathbf{f}$ solves $\mathcal{P}(J, \mathbf{R}', \mathbf{s})$, where $\mathbf{g}(\mathbf{x}) = \mathbf{f}(\mathbf{A}\mathbf{x} + \mathbf{b})$, provided that $\langle \mathbf{R}', \mathbf{f}\rangle = \langle \mathbf{R}, \mathbf{g}\rangle$ for all $\mathbf{f}$. □

This encompasses the cases of rotating, scaling, and shifting the coordinate system of $\mathbf{x}$. We want our solution to be invariant with respect to these changes. The new filter $\mathbf{R}'$ can be written in the functional form as $\mathbf{R}(\mathbf{x}) = (\det \mathbf{A})\mathbf{R}'(\mathbf{A}\mathbf{x}+\mathbf{b})$. Consequently, we want $\mathbf{A}$ to be an invertible matrix. Similarly to Property 3, the geometric invariance can be ensured by pseudo-invariance with respect to the geometric transformations, i.e.,

$$J\left(\mathbf{f}(\mathbf{A}\mathbf{x} + \mathbf{b})\right) = c(\mathbf{A}, \mathbf{b})J\left(\mathbf{f}(\mathbf{x})\right) \tag{6}$$

where c is a continuous function of $\mathbf{A}$ and $\mathbf{b}$, independent of $\mathbf{f}$, and with no connection to c in (4).

We will impose geometric invariance only with respect to scaled orthogonal matrices $\mathbf{A}$, satisfying $\mathbf{A}\mathbf{A}^T = k\mathbf{I}$. See Section IV-A for details.

Property 5 (Density): The solution space F contains all test functions $\mathbf{u}$ from $\mathcal{D}$. For any function $\mathbf{f} \in F$, there is a sequence of test functions $\mathbf{u}_1, \mathbf{u}_2, \ldots$ such that $\lim_{i\to\infty} J(\mathbf{f} - \mathbf{u}_i) = 0$. □

This property is indeed somewhat technical but its significance can be readily grasped. The first part ensures that the solution space is large, i.e., that it contains as many "good" functions as possible. It guarantees that at least all test functions can be measured using the criterion J. The second part concerns the behavior of J for functions on the closure of $\mathcal{D}$, that is to say, for functions that are not in $\mathcal{D}$ but can be expressed as a limit of a sequence of test functions. It specifies the density of $\mathcal{D}$ in F. Consequently, we can do most of our reasoning in the space of test functions and then extend the result to the whole of F using a limiting process.

In practice, Property 5 is always satisfied by the quadratic semi-norms we will be considering, in particular, by the semi-norms of Duchon. (This originates from the density of $\mathcal{D}$ in Sobolev spaces.)

III. Bilinear Forms

From now on, we consider exclusively those criteria that can be expressed using a nonnegative *bilinear form*. A bilinear form $B(\mathbf{f}, \mathbf{g})$ maps pairs of functions $\mathbf{f}$, $\mathbf{g}$ onto $\mathbb{R}$. It is *symmetric* $(B(\mathbf{f}, \mathbf{g}) = B(\mathbf{g}, \mathbf{f}))$ and *linear* $(B(\alpha\mathbf{f} + \beta\mathbf{g}, \mathbf{h}) = \alpha B(\mathbf{f}, \mathbf{h}) + \beta B(\mathbf{f}, \mathbf{g}))$ with respect to both its arguments. It is *non-negative* iff $B(\mathbf{f}, \mathbf{f}) \geq 0$ for all $\mathbf{f} \in F$. We associate B with a criterion

$$J(\mathbf{f}) = B(\mathbf{f}, \mathbf{f}) \tag{7}$$

which we call a *quadratic criterion*. Conversely, given a quadratic criterion J, the associated bilinear form B can be obtained as

$$B(\mathbf{f}, \mathbf{g}) = \tfrac{1}{4}\left(J(\mathbf{f}+\mathbf{g}) - J(\mathbf{f}-\mathbf{g})\right). \tag{8}$$

The square-root $\sqrt{J}$ is a *semi-norm*, i.e., it satisfies the triangular inequality and semi-linearity $(\sqrt{J(\lambda\mathbf{f})} = |\lambda|\sqrt{J(\mathbf{f})})$. Unlike for a norm, there might be more than one $\mathbf{f}$ satisfying $J(\mathbf{f}) = 0$. Such functions define a *kernel* $\mathcal{K}$. The criterion J is convex. The important Cauchy–Schwartz inequality $|B(\mathbf{f}, \mathbf{g})| \leq \sqrt{J(\mathbf{f})J(\mathbf{g})}$ holds as well; the equality is reached iff $\exists\ \lambda, \mu \in \mathbb{R}, |\lambda| + |\mu| \neq 0$; $\lambda\mathbf{f} + \mu\mathbf{g} \in \mathcal{K}$.

A. Variational Problem With a Quadratic Criterion and Linear Constraints

The restriction to bilinear forms is justified, namely, in view of satisfying Property 2, which yields a useful vector space structure for the solution space. The proof that a quadratic criterion J implies that Property 2 can be found in Appendix A.

Because of the convexity of J, if there is a local minimum, it is also the global minimum. Moreover, if two functions $\mathbf{f}_1, \mathbf{f}_2$ solve the problem $\mathcal{P}$, then their difference $\mathbf{f}_1 - \mathbf{f}_2$ necessarily belongs to $\mathcal{K}$. (See Appendix B for a proof.) Therefore, if the constraints (2) cannot be met by two distinct functions differing by an element from the kernel, the solution is unique. This is easy to check because in most cases of interest, the kernel $\mathcal{K}$ is fairly small. We will see later that it mostly consists of low-order polynomials.

B. Operator Kernel of a Bilinear Form

Any bilinear form satisfying very mild conditions (see [40]) can be written in the form of a scalar product

$$B(\mathbf{f}, \mathbf{g}) = \int_{\mathbb{R}^{2m}} \mathbf{f}^T(\mathbf{x})\mathbf{V}(\mathbf{x}, \mathbf{y})\mathbf{g}(\mathbf{y})\,\mathrm{d}\mathbf{x}\,\mathrm{d}\mathbf{y} \tag{9}$$

where $\mathbf{V}$ is a $n \times n$ matrix of distributions called an *operator kernel* of the bilinear form. Technically, the existence of the integral is not guaranteed unless both $\mathbf{f}$ and $\mathbf{g}$ are from the class $\mathcal{D}$ of infinitely differentiable and compactly supported *test functions*.

Without any loss of generality, we can assume $\mathbf{V}$ to be symmetric ($\mathbf{V}(\mathbf{x}, \mathbf{y}) = \mathbf{V}(\mathbf{y}, \mathbf{x})$) because the operator kernel can always be symmetrized as $1/2\,(\mathbf{V}(\mathbf{x}, \mathbf{y}) + \mathbf{V}(\mathbf{y}, \mathbf{x}))$ without affecting the associated bilinear form B. By exchanging $\mathbf{f}$ and $\mathbf{g}$, we also find that $\mathbf{V}$ must have a matrix symmetry $\mathbf{V} = \mathbf{V}^T$. The implications of (4) and (6) on the properties of $\mathbf{V}$ are studied in Section IV.

C. Convolutional Kernel

If $\mathbf{V}$ is translation-invariant, it can be written using a single-parametric distribution matrix $\mathbf{U}(\mathbf{x} - \mathbf{y}) = \mathbf{V}(\mathbf{x}, \mathbf{y})$. This transforms (9) to

$$B(\mathbf{f}, \mathbf{g}) = \int_{\mathbb{R}^m} \mathbf{f}^T(\mathbf{x}) \mathbf{U}(\mathbf{x} - \mathbf{y}) \mathbf{g}(\mathbf{y})\, \mathrm{d}\mathbf{x}\, \mathrm{d}\mathbf{y} \tag{10}$$

for test functions $\mathbf{f}$, $\mathbf{g}$. We recognize the convolution here:

$$B(\mathbf{f}, \mathbf{g}) = \langle \mathbf{f}, \mathbf{U} * \mathbf{g} \rangle, \qquad \text{for } \mathbf{g} \in \mathcal{D} \tag{11}$$

$$= \langle \mathbf{U} * \mathbf{f}, \mathbf{g} \rangle \tag{12}$$

where the restriction of $\mathbf{g}$ to test functions is useful to ensure that $\langle \mathbf{f}, \mathbf{U} * \mathbf{g} \rangle$ exists. We call $\mathbf{U}$ the *convolutional kernel* of the bilinear form. Because of the symmetries of $\mathbf{V}$, we have the same symmetries on $\mathbf{U}$, i.e., $\mathbf{U}(\mathbf{x}) = \mathbf{U}(-\mathbf{x}) = \mathbf{U}^T(\mathbf{x})$. [This has also simplified (12).]

D. Fourier Form

Both (9) and (10) can be also calculated in the Fourier domain. For this, we need the Fourier transforms $\hat{\mathbf{f}}$, $\hat{\mathbf{g}}$, and $\hat{\mathbf{U}}$ (see Sections II-A and II-B for a definition). For example, the expression (10) can be written using Parseval's theorem as

$$B(\mathbf{f}, \mathbf{g}) = \frac{1}{(2\pi)^m} \int_{\mathbb{R}^m} \hat{\mathbf{f}}^H(\boldsymbol{\omega}) \hat{\mathbf{U}}(\boldsymbol{\omega}) \hat{\mathbf{g}}(\boldsymbol{\omega})\, \mathrm{d}\boldsymbol{\omega} \tag{13}$$

where $\hat{\mathbf{f}}^H = (\hat{\mathbf{f}}^T)^*$ is the Hermite transpose of $\hat{\mathbf{f}}$.

E. Extending and Factorizing the Bilinear Form

The original equations (9) and (10) define $B(\mathbf{f}, \mathbf{g})$ only for test functions $\mathcal{D}$. However, later, we will need to evaluate B also for $\mathbf{f}$ from some larger class $\overline{F} \supset F$, conserving all the properties of the bilinear form. Already, (9) retains a meaning if $g \in \mathcal{D}$ and $\mathbf{f}$ belongs to the dual (distribution) space of $G = \{\int \mathbf{V}(\mathbf{x}, \mathbf{y}) \mathbf{g}(\mathbf{y})\, \mathrm{d}\mathbf{y};\ g \in \mathcal{D}\}$. In particular, if we define B through (12), it allows us to consider any distribution $\mathbf{f}$, provided that $\mathbf{U}$ is compactly supported. The extensions of B coincide for test functions but might give different results when evaluated for other (nontest) functions.

An alternative, symmetric definition of B is

$$B(\mathbf{f}, \mathbf{g}) = \langle \mathbf{L}^T * \mathbf{f}, \mathbf{L}^T * \mathbf{g} \rangle \qquad \text{for } \mathbf{f}, \mathbf{g} \in \mathcal{D} \tag{14}$$

which leads to a very simple expression for J:

$$J(\mathbf{f}) = \|\mathbf{L}^T * \mathbf{f}\|^2 = \sum_{i=1}^{p} \left\| \mathbf{l}_i^T * \mathbf{f} \right\|^2. \tag{15}$$

The convolutional operator $\mathbf{L}_{n \times p} = [\mathbf{l}_1 \cdots \mathbf{l}_p]$ has an adjoint $\mathbf{L}^\dagger = \check{\mathbf{L}}^T$ [where the notation $\check{f}(\mathbf{x})$ stands for $f(-\mathbf{x})$]. We obtain an equivalence between (14) and (12) by setting $\mathbf{U} = \check{\mathbf{L}} * \mathbf{L}^T$.

There are generally many possible factorizations, leading to many extensions as detailed in the previous section. To illustrate this point, we consider the example of the scalar distribution in two dimensions: $u = (\partial^4/\partial x^4)\,\delta + 2(\partial^4/\partial x^2 \partial y^2)\,\delta + (\partial^4/\partial y^4)\,\delta$. It can be factorized either with the 1-D (scalar) operator $\mathbf{L} = [(\partial^2/\partial x^2)\,\delta + (\partial^2/\partial y^2)\,\delta]$ or, alternatively, with the three-dimensional (3-D) vector operator $\mathbf{L} = [(\partial^2/\partial x^2)\,\delta\ \sqrt{2}\,(\partial^2/\partial x \partial y)\,\delta\ \ (\partial^2/\partial y^2)\,\delta]$.

The latter factorization leads to the Duchon's semi-norm (see Section IV-G)

$$\|f\|_{D_2}^2 = \int_{\mathbb{R}^2} \left(\frac{\partial^2 f}{\partial x^2}\right)^2 + 2\left(\frac{\partial^2 f}{\partial x \partial y}\right)^2 + \left(\frac{\partial^2 f}{\partial y^2}\right)^2 \mathrm{d}x\, \mathrm{d}y \tag{16}$$

whereas the former gives a semi-norm based on the Laplacian:

$$\|f\|_{\Delta_2}^2 = \int_{\mathbb{R}^2} \left(\frac{\partial^2 f}{\partial x^2}\right)^2 + 2\frac{\partial^2 f}{\partial x^2}\frac{\partial^2 f}{\partial y^2} + \left(\frac{\partial^2 f}{\partial y^2}\right)^2 \mathrm{d}x\, \mathrm{d}y \tag{17}$$

which is *not* strictly equivalent to (16). An example is $f = xy$, which gives $\|f\|_{\Delta_2}^2 = 0$ but $\|f\|_{D_2}^2 = \infty$. An important case where the expressions (17) and (16) *are* equivalent is when f is a test function. Then, by integration by parts

$$\int_{\mathbb{R}^2} \frac{\partial^2 f}{\partial x^2}\frac{\partial^2 f}{\partial y^2}\, \mathrm{d}x\, \mathrm{d}y = \int_{\mathbb{R}^2} \left(\frac{\partial^2 f}{\partial x \partial y}\right)^2 \mathrm{d}x\, \mathrm{d}y. \tag{18}$$

These may sound like technicalities, but they should not be overlooked; otherwise, one may easily formulate problems that are not well defined mathematically (as was, for example, the case in [43]).

Coming back to the general formulation with $\mathbf{f}$ and $\mathbf{g}$ in $\mathcal{D}$, we write the Fourier domain equivalent of (14) and (15):

$$B(\mathbf{f}, \mathbf{g}) = \frac{1}{(2\pi)^m} \left\langle \hat{\mathbf{L}}^T \hat{\mathbf{f}}, \hat{\mathbf{L}}^T \hat{\mathbf{g}} \right\rangle \tag{19}$$

$$= \frac{1}{(2\pi)^m} \int_{\mathbb{R}^m} \left(\hat{\mathbf{L}}^T \hat{\mathbf{f}}\right)^H (\boldsymbol{\omega}) \left(\hat{\mathbf{L}}^T \hat{\mathbf{g}}\right)(\boldsymbol{\omega})\, \mathrm{d}\boldsymbol{\omega} \tag{20}$$

with an associated criterion

$$J(\mathbf{f}) = \frac{1}{(2\pi)^m} \int_{\mathbb{R}^m} \left\| \hat{\mathbf{L}}^T \hat{f} \right\|^2 \mathrm{d}\boldsymbol{\omega} \tag{21}$$

where $\hat{\mathbf{U}} = \hat{\mathbf{L}}^* \hat{\mathbf{L}}^T$. Note that the phase of $\hat{\mathbf{L}}$ can be freely chosen in addition to the freedom demonstrated in the time-domain factorization. The phase of $\hat{\mathbf{L}}$ may represent the shift of $\mathbf{L}$ in the time domain; more generally, it corresponds to applying an allpass (unitary-gain) filter to $\mathbf{L}$.

IV. Imposing Invariance Properties

The intent of this section is to apply the first principles from Section II-D to come up with a constrained form of the variational criterion that is consistent with our invariance requirements. We will end up with what is essentially a one-parameter family of criteria (cf. Theorem 1).

As we have seen, sufficient conditions to ensure Properties 3 and 4 are given by (4) and (6), respectively. We now show how

(4) and (6) constrain our choice of the kernel $\mathbf{V}$ of the bilinear form B. It is useful to realize that if

$$\int \mathbf{f}^T(\mathbf{x})\mathbf{V}(\mathbf{x},\mathbf{y})\mathbf{f}(\mathbf{y})\,d\mathbf{x}\,d\mathbf{y} = \int \mathbf{f}^T(\mathbf{x})\mathbf{W}(\mathbf{x},\mathbf{y})\mathbf{f}(\mathbf{y})\,d\mathbf{x}\,d\mathbf{y}, \qquad \text{for all } \mathbf{f}\in\mathcal{D} \quad (22)$$

then by considering $\mathbf{f}+\mathbf{g}$ and $\mathbf{f}-\mathbf{g}$ instead of $\mathbf{f}$, we get

$$\int \mathbf{f}^T(\mathbf{x})\mathbf{V}(\mathbf{x},\mathbf{y})\mathbf{g}(\mathbf{y})\,d\mathbf{x}\,d\mathbf{y} = \int \mathbf{f}^T(\mathbf{x})\mathbf{W}(\mathbf{x},\mathbf{y})\mathbf{g}(\mathbf{y})\,d\mathbf{x}\,d\mathbf{y}, \qquad \text{for all } \mathbf{f},\mathbf{g}\in\mathcal{D} \quad (23)$$

which is equivalent to saying that $\mathbf{V}=\mathbf{W}$ in the distributional sense (see Section II-B). The converse also holds by substituting $\mathbf{f}=\mathbf{g}$. Therefore, (4) and (6) on the criterion translate into equations for the distributional kernel $\mathbf{V}$ as

Matrix linearity:

$$\mathbf{A}^T\mathbf{V}(\mathbf{x},\mathbf{y})\mathbf{A} = c(\mathbf{A},\mathbf{b})\mathbf{V}(\mathbf{x},\mathbf{y}) \quad (24)$$

and

$$\int \mathbf{V}(\mathbf{x},\mathbf{y})\,d\mathbf{y} = 0. \quad (25)$$

Geometric invariance:

$$\mathbf{V}(\mathbf{x},\mathbf{y}) = c(\mathbf{A},\mathbf{b})\mathbf{V}(\mathbf{A}\mathbf{x}+\mathbf{b},\mathbf{A}\mathbf{y}+\mathbf{b})(\det\mathbf{A})^2. \quad (26)$$

A. Translation Invariance

From (25), we directly see that

$$B(\mathbf{f},\mathbf{b}) = 0 \qquad \text{for any constant } \mathbf{b} \text{ and } \mathbf{f}\in\mathcal{D} \quad (27)$$

and thus, there exists an extension of J to functions outside of $\mathcal{D}$ such that $J(\mathbf{b})=0$. In other words, the criterion J must give zero for constant functions.

We can now consider geometric translation invariance (in the domain of $\mathbf{x}$) by setting $\mathbf{A}=-\mathbf{I}$ (reflection about the origin) and $\mathbf{b}=\mathbf{x}+\mathbf{y}$ in (26), which simplifies to

$$\mathbf{V}(\mathbf{x},\mathbf{y}) = c(-\mathbf{I},\mathbf{x}+\mathbf{y})\mathbf{V}(\mathbf{y},\mathbf{x}). \quad (28)$$

Using the symmetry of $\mathbf{V}$, this implies $c(-\mathbf{I},\mathbf{x}')=1$ for all $\mathbf{x}'$. Letting $\mathbf{b}=\mathbf{y}$ in (26) leads to

$$\mathbf{V}(\mathbf{x},\mathbf{y}) = \mathbf{U}(\mathbf{y}-\mathbf{x}) \quad (29)$$

where we have substituted $\mathbf{V}(\mathbf{y}-\mathbf{x},\mathbf{0})=\mathbf{U}(\mathbf{y}-\mathbf{x})$. This means that we can use the simpler expression (10) instead of (9). By virtue of (29), $\mathbf{U}$ is symmetric and even and the hypothesis in Section III-C applies. Equation (26) then becomes $\mathbf{U}(\mathbf{x}-\mathbf{y}) = c(\mathbf{A})\mathbf{U}(\mathbf{A}(\mathbf{x}-\mathbf{y}))$, and consequently

$$\mathbf{U}(\mathbf{x}) = c(\mathbf{A})\mathbf{U}(\mathbf{A}\mathbf{x}) \quad (30)$$

for any matrix $\mathbf{A}$, where we have incorporated $(\det\mathbf{A})^2$ into $c(\mathbf{A})$.

B. Rotational Invariance

Another special case of geometrical transformations are rotations and symmetries, i.e., matrices that satisfy the orthogonality condition $\mathbf{A}\mathbf{A}^T=\mathbf{I}$. Applying (30) twice yields

$$c(\mathbf{A}\mathbf{B}) = c(\mathbf{A})c(\mathbf{B}). \quad (31)$$

It is useful to consider this equation for a Householder matrix $\mathbf{A}_\mathbf{v}=\mathbf{I}-2\mathbf{v}\mathbf{v}^T$, where $\mathbf{v}^T\mathbf{v}=1$, since any orthogonal matrix can by factorized using a finite product of Householder matrices only. As $\mathbf{A}_\mathbf{v}^2=\mathbf{I}$, from (31), we have $|c(\mathbf{A}_\mathbf{v})|=1$. Furthermore, as $J\geq 0$, we necessarily have $c(\mathbf{A}_\mathbf{v})\geq 0$. Thus, $c(\mathbf{A})=1$ for all orthogonal matrices $\mathbf{A}$. Equation (30) becomes $\mathbf{U}(\mathbf{x})=\mathbf{U}(\mathbf{A}\mathbf{x})$. It is always possible to choose $\mathbf{A}$ such that $\mathbf{A}\mathbf{x}=\|\mathbf{x}\|\mathbf{e}_1$, where $\mathbf{e}_1$ is the first basis vector; see [44]. Consequently, the distribution $\mathbf{U}$ must be radial

$$\mathbf{U}(\mathbf{x}) = \mathbf{U}_0(\rho), \qquad \text{where } \rho=\|\mathbf{x}\|. \quad (32)$$

It is easy to verify that thanks to the orthonormality of $\mathbf{A}$, rotating $\mathbf{x}\to\mathbf{A}\mathbf{x}$ does not change ρ.

C. Scale Invariance

The last remaining class of geometrical transformations we consider is uniform scaling. Using (30) as before yields $\mathbf{U}_0(\mathbf{x}) = c(\lambda)\mathbf{U}_0(\lambda\mathbf{x})$, where λ is a real scaling factor, and where we have accommodated the Jacobian λ^{2m} into $c(\lambda)$. We use the rotation-invariant form (32), which gives $\mathbf{U}_0(\rho)=c(\lambda)\mathbf{U}_0(\lambda\rho)$ for $\lambda>0$. Note that $c(\lambda\lambda')=c(\lambda)c(\lambda')$. Repetitive scaling by λ yields $c(\lambda)^k=c(\lambda^k)$. This implies $c(\lambda)^{1/q}=c(\lambda^{1/q})$ and $c(\lambda)^{p/q}=c(\lambda^{p/q})$. By continuity, $c(\lambda)^x=c(\lambda^x)$ as well for real x. Consequently, we have $c(\lambda)=\lambda^\gamma$ and

$$\mathbf{U}_0(\rho) = \lambda^\gamma\mathbf{U}_0(\lambda\rho). \quad (33)$$

In the case where the radial form of the convolutional kernel $\mathbf{U}_0(\rho)$ is a function, the preceding equation implies $\mathbf{U}_0(\rho)=c\rho^{-\gamma}$. Note that when $\gamma\leq 2m$, then $\mathbf{U}(\mathbf{x})$ is not locally integrable over $\mathbb{R}^m\times\mathbb{R}^m$. Therefore, we need to consider the equations in the sense of distributions.

The corresponding expression in the Fourier domain is

$$\hat{\mathbf{U}}(\boldsymbol{\omega}) = c\|\boldsymbol{\omega}\|^{2\alpha} \quad (34)$$

where $2\alpha=\gamma-m$, and the factor 2 is for future convenience and notational consistency with [24].

D. Matrix Linearity

We have already studied the effect of $\mathbf{b}$ in (24). Let us now concentrate on the implications of $\mathbf{A}$. Substituting (29) yields $\mathbf{A}^T\mathbf{U}(\mathbf{x})\mathbf{A} = c(\mathbf{A})\mathbf{U}(\mathbf{x})$; thus, $c(\mathbf{A}\mathbf{B})=c(\mathbf{A})c(\mathbf{B})$. We show that $c(\mathbf{A})=1$ by the same proof as in Section IV-B. Thus, $\mathbf{U}(\mathbf{x})$ commutes with an arbitrary orthogonal matrix $\mathbf{A}$:

$$\mathbf{U}\mathbf{A} = \mathbf{A}\mathbf{U}. \quad (35)$$

It can be easily seen that $\mathbf{U}$ is a multiple of the identity matrix and is completely determined by a scalar distribution $u(\mathbf{x})$: $\mathbf{U}(\mathbf{x})=u(\mathbf{x})\cdot\mathbf{I}$. To prove this, it suffices to consider Householder matrices $\mathbf{A}_\mathbf{v}=\mathbf{I}-2\mathbf{v}\mathbf{v}^T$; substituting into (35) yields $\mathbf{v}\mathbf{v}^T\mathbf{U}=\mathbf{U}\mathbf{v}\mathbf{v}^T$; right-multiplying by $\mathbf{v}$ shows that $\mathbf{v}(\mathbf{v}^T\mathbf{U}\mathbf{v})=\mathbf{U}\mathbf{v}$, which means that any vector $\mathbf{v}$ is an eigenvector of $\mathbf{U}$ and completes the demonstration.

E. Form of the Criterion

A direct consequence of the results from the preceding sections is the following theorem.

Theorem 1 (Form of the Criterion): Let $J(\mathbf{f})$ be a quadratic criterion. Then, any associated variational problem $\mathcal{P}$ satisfies Property 2. Furthermore, $\mathcal{P}$ satisfies Properties 3 and 4 if and only if J can be expressed in the following form:

$$J(\mathbf{f}) = c \int_{\mathbb{R}^m} \sum_{i=1}^{n} \underbrace{\|\boldsymbol{\omega}\|^{2\alpha}}_{(\hat{\mathbf{v}})_{ii}} |\hat{f}_i(\boldsymbol{\omega})|^2 \, \mathrm{d}\boldsymbol{\omega} \tag{36}$$

for any function $\mathbf{f} \in \mathcal{D}$.

The bilinear form associated with (36) is

$$B(f, g) = \int_{\mathbb{R}^m} \sum_{i=1}^{n} \|\boldsymbol{\omega}\|^{2\alpha} \hat{f}_i(\boldsymbol{\omega}) \hat{g}_i^*(\boldsymbol{\omega}) \, \mathrm{d}\boldsymbol{\omega}. \tag{37}$$

Note that the criterion value for a vector function is a sum of the criterion values for its components

$$J(\mathbf{f}) = \sum_{i=1}^{n} J(f_i) \tag{38}$$

which permits us to concentrate on the scalar case for simplicity. We now consider two possible extensions of B to nontest functions.

F. Laplacian Semi-Norm

The criterion defined in the Fourier domain by (36) is easily associated (in the scalar case) to an equivalent semi-norm in the time domain using an iterated m-dimensional Laplacian for even α

$$J_{\Delta_\alpha}(f) = \|f\|_{\Delta_\alpha}^2 = \int_{\mathbb{R}^m} \left| \Delta^{\alpha/2} f(\mathbf{x}) \right|^2 \, \mathrm{d}\mathbf{x}. \tag{39}$$

G. Duchon's Semi-Norms

The principal disadvantage of (39) is that its kernel $\mathcal{K}_\Delta$ is too large. For example, for $\alpha = 2$, it contains every function that satisfies the Laplace equation, such as the real part of an analytical function, for example, $(x + iy)^k + (x - iy)^k$. Therefore, the variational problem with this criterion will typically have an infinite number of solutions.

Fortunately, it turns out that there are other time-domain forms that correspond to (36) and do not have this problem. Namely, we now present the family of semi-norms introduced by Duchon [24]. He first defines a differential operator D as a vector of all possible partial derivatives of f of order M

$$D^M f = \left[\frac{\partial^M f}{\partial x_1^M}, \ldots, \frac{\partial^M f}{\partial x_{k_1} \cdots \partial x_{k_M}}, \ldots, \frac{\partial^M f}{\partial x_m^M} \right]^T \tag{40}$$

with $k_1, \ldots, k_M \in \{1, \ldots, m\}^M$. For example, for $m = 2$, $M = 2$, we get

$$D^2 f(x, y) = \left[\frac{\partial^2 f}{\partial x^2}, \frac{\partial^2 f}{\partial x \partial y}, \frac{\partial^2 f}{\partial y \partial x}, \frac{\partial^2 f}{\partial y^2} \right]^T. \tag{41}$$

Then, he defines a semi-norm by taking the sum of the squares of all the elements and integrating it over the space $\mathbb{R}^m$

$$\|f\|_{D_M} = \left(\int_{\mathbb{R}^m} \|D^M f\|^2 \, \mathrm{d}\mathbf{x} \right)^{1/2} \tag{42}$$

where $\|\cdot\|$ is an Euclidean norm in $\mathbb{R}^{m^M}$. More explicitly (using the commutativity of the partial derivatives)

$$\|f\|_{D_M}^2 = \sum_{|\mathbf{l}|=M} \frac{M!}{\mathbf{l}!} \int_{\mathbb{R}^m} \left(\frac{\partial^M f}{\partial \mathbf{x}^{\mathbf{l}}} \right)^2 \mathrm{d}\mathbf{x} \tag{43}$$

where $l_1, \ldots, l_m \in \{0, \ldots, M\}$, $\mathbf{l}! = \prod_{s=1}^m l_s!$, $|\mathbf{l}| = \sum_{s=1}^m l_s$, and $\partial \mathbf{x}^{\mathbf{l}} = \partial x_1^{l_1} \cdots \partial x_m^{l_m}$. Following our example for $m = 2$ and $M = 2$, we get the most often used Duchon's semi-norm (16). This semi-norm leads to the well known thin-plate splines [45].

Interestingly, the kernel of $\| \cdot \|_{D_2}$ contains only functions whose second partial derivatives are zero; i.e., linear polynomials $a_0 + a_1 x + a_2 y$. In fact, the kernel $\mathcal{K}_{D_M}$ of Duchon's semi-norm of order M contains only the polynomials of degree $M - 1$.

All the Duchon's semi-norms can be associated with a bilinear form so that $\|f\|^2 = J(f) = B(f, f)$. The norm (42) gives

$$B_M(f, g) = \int_{\mathbb{R}^m} (D^M f)^T (D^M g) \, \mathrm{d}\mathbf{x} \tag{44}$$

or equivalently

$$B_M(f, g) = \int_{\mathbb{R}^m} \sum_{|\mathbf{k}|=M} \frac{M!}{\mathbf{k}!} \frac{\partial^M f}{\partial \mathbf{x}^{\mathbf{k}}} \frac{\partial^M g}{\partial \mathbf{x}^{\mathbf{k}}} \, \mathrm{d}\mathbf{x}. \tag{45}$$

H. Semi-Norms for Fractional Derivatives

In many applications, the choice of discrete-order Duchon's semi-norms does not permit sufficiently fine tuning. However, Duchon has combined the time and Fourier domain definitions to also obtain semi-norms corresponding to fractional derivatives.

$$J(f) = \|f\|_{D_{M,s}}^2 = \int_{\mathbb{R}^m} \|\boldsymbol{\omega}\|^{2s} \left\| \mathcal{F} D^M f \right\|_{\ell_2}^2 \, \mathrm{d}\boldsymbol{\omega} \tag{46}$$

where $\mathcal{F}$ is the Fourier transform operator as defined in Sections II-A and B. When $s = 0$, this definition is completely equivalent to (42), that is, $\|f\|_{D_M} = \|f\|_{D_{M,0}}$. When, on the other hand, $M = 0$, this definition is equivalent to (39), i.e., $\|f\|_{\Delta_\alpha} = \|f\|_{D_{0,\alpha}}$, for $f \in \mathcal{D}$. Note that the kernel of $\|f\|_{D_M}$ is the kernel of $\|f\|_{D_{M,s}}$.

The associated bilinear form is

$$B_{M,s}(f, g) = \int_{\mathbb{R}^m} \|\boldsymbol{\omega}\|^{2s} \left(\mathcal{F} D^M f(\mathbf{x}) \right)^H \left(\mathcal{F} D^M g(\mathbf{x}) \right) \, \mathrm{d}\boldsymbol{\omega}. \tag{47}$$

V. Solution to the Variational Problem

In this section, we reconsider our variational problem $\mathcal{P}(J, \mathbf{R}, \mathbf{s})$ defined in Section II-C, derive some properties of its solution $\mathbf{f}_{\mathrm{out}}$, and use them to obtain the explicit form of the solution.

A. Lagrange Multipliers

First, we construct an augmented criterion according to the Lagrange multipliers' method

$$J_*(\mathbf{f}, \boldsymbol{\lambda}) = J(\mathbf{f}) - 2\boldsymbol{\lambda}^T(\langle \mathbf{R}, \mathbf{f}\rangle - \mathbf{s}) \tag{48}$$

where $\boldsymbol{\lambda} \in \mathbb{R}^Q$ is the vector of Lagrange multipliers. If $\mathbf{f}_{\boldsymbol{\lambda}}$ minimizes $J_*(\mathbf{f}, \boldsymbol{\lambda})$, then choosing $\boldsymbol{\lambda} = \boldsymbol{\lambda}_{\text{opt}}$ such that $\langle \mathbf{R}, \mathbf{f}\rangle = \mathbf{s}$ implies that $\mathbf{f}_{\text{out}} = \mathbf{f}_{\boldsymbol{\lambda}_{\text{opt}}}$ minimizes $J(\mathbf{f})$ under constraints (2).

We carry on using a standard variational argument. We take a small perturbation $\alpha\mathbf{g}$, where $\mathbf{g} \in F$ and $\alpha \in \mathbb{R}$, add it to $\mathbf{f}_{\text{out}}$, and study the new criterion value $J_*(\mathbf{f}_{\text{out}} + \alpha\mathbf{g})$. We consider its derivative

$$\frac{\partial}{\partial\alpha} J_*(\mathbf{f}_{\text{out}} + \alpha\mathbf{g}) = 2B(\mathbf{f}_{\text{out}}, \mathbf{g}) - 2\boldsymbol{\lambda}^T\langle \mathbf{R}, \mathbf{g}\rangle \tag{49}$$

which vanishes iff $(\mathbf{f}_{\text{out}}, \boldsymbol{\lambda})$ is a saddle point. In addition, $\langle \mathbf{R}, \mathbf{f}_{\text{out}}\rangle = \mathbf{s}$ must hold. This directly leads to the following lemma.

Lemma 1: A function $\mathbf{f}_{\text{out}}$ from F solves the variational problem $\mathcal{P}$ if and only if there is a real vector $\boldsymbol{\lambda}_{\text{opt}}$ such that

$$B(\mathbf{f}_{\text{out}}, \mathbf{g}) = \boldsymbol{\lambda}_{\text{opt}}^T\langle \mathbf{R}, \mathbf{g}\rangle, \qquad \text{for all } \mathbf{g} \in \mathcal{D}$$

and

$$\langle \mathbf{R}, \mathbf{f}_{\text{out}}\rangle = \mathbf{s}. \tag{50}$$

Note that because of the Property 5 on density, we can initially consider only $\mathbf{g}$ from $\mathcal{D}$ and then extend to F while the Lemma remains valid.

For $\mathbf{g}$ from the kernel $\mathcal{K}$, we have $B(\mathbf{f}_{\text{out}}, \mathbf{g}) = 0$ [because $B(\mathbf{g}, \mathbf{g}) = 0$], and thus

$$\boldsymbol{\lambda}^T\langle \mathbf{R}, \mathbf{g}\rangle = 0, \qquad \text{for each } \mathbf{g} \in \mathcal{K}. \tag{51}$$

B. Introducing Fundamental Solutions

We now suppose that we have found a set of functions $\boldsymbol{\varphi}_i$ such that

$$B(\boldsymbol{\varphi}_i, \mathbf{g}) = \langle \mathbf{r}_i, \mathbf{g}\rangle, \qquad \text{for all } \mathbf{g} \in \mathcal{D}. \tag{52}$$

We call $\boldsymbol{\varphi}_i$ a *fundamental solution* corresponding to a filter $\mathbf{r}_i$. (See also Section V-D.) There is often no fundamental solution $\boldsymbol{\varphi}_i$ in F. Then, we search $\boldsymbol{\varphi}_i$ in $\overline{F}$, which is why we had to restrict $\mathbf{g}$ to $\mathcal{D}$ (Section III-E).

We want $\langle \mathbf{R}, \boldsymbol{\varphi}_i\rangle$ to be finite for all $i = 1 \cdots Q$. If this is not the case, we can suspect that our minimization problem does not have a solution in F, which can hopefully be proven using another method.

C. Explicit Solution of the Variational Problem

In order to obtain a more useful result than Lemma 1, we will use the linearity of B. Take a function

$$\mathbf{f}(\mathbf{x}) = \sum_{i=1}^{Q} \lambda_i \boldsymbol{\varphi}_i. \tag{53}$$

Because of (52), the function (53) can be made to satisfy (50) for some λ_i. We might be tempted to conclude that it therefore solves $\mathcal{P}$. However, this will not necessarily work because $\boldsymbol{\varphi}_i$, and therefore $\mathbf{f}$ in (53) do not, in general, belong to the admissible solution space F.

With (53), we have exactly as many λ_is as there are consistency constraints (2). This means that there are not enough degrees of freedom in (53) to ensure the condition $\mathbf{f} \in F$.

Note that if $\mathbf{p}$ belongs to $\mathcal{K}$, then $B(\mathbf{f} + \mathbf{p}, \mathbf{g}) = B(\mathbf{f}, \mathbf{g})$. We can therefore add to $\mathbf{f}$ a function $\mathbf{p}$ from $\mathcal{K}$, obtaining $\mathbf{f}_{\text{out}} = \mathbf{f} + \mathbf{p}$, which gives us the possibility to make $\mathbf{f}_{\text{out}} \in F$ while conserving the validity of (50).

Equation (51) will allow us to find the $\mathbf{p}$. If $\mathcal{K}$ has a finite basis, we can express $\mathbf{p}(\mathbf{x})$ as

$$\mathbf{p}(\mathbf{x}) = \sum_{k=0}^{P-1} a_k \mathbf{p}_k(\mathbf{x}). \tag{54}$$

Through linearity, (51) is equivalent to the orthogonality constraints

$$\boldsymbol{\lambda}^T\langle \mathbf{R}, \mathbf{p}_k\rangle = 0, \qquad \text{for each } k. \tag{55}$$

This gives the same number of constraints as there are additional unknowns in (54). Combining (55) and (2) gives us a set of linear equations for exactly as many unknowns λ_i and a_k, which is a necessary condition for the unicity of the solution. (More on unicity in Section V-F.) Adding the kernel term gives us in general sufficient freedom to find an $\mathbf{f}_{\text{out}}$ in F; however, this needs to be verified *a posteriori* on a case-by-case basis. We summarize our findings in the form of a theorem. See Appendix C for a proof.

Theorem 2 (Variational Problem Solution): Let λ_i and a_k be real numbers and $\{\mathbf{p}_k\}$ a basis of the kernel $\mathcal{K}$ of J. Further, let $\{\boldsymbol{\varphi}_i\}$ be a set of fundamental solutions corresponding to filters $\mathbf{R}$ in the sense of (52). Then, the function

$$\mathbf{f}_{\text{out}}(\mathbf{x}) = \sum_{k=0}^{P-1} a_k \mathbf{p}_k + \sum_{i=1}^{Q} \lambda_i \boldsymbol{\varphi}_i \tag{56}$$

solves the interpolation problem $\mathcal{P}(J, \mathbf{R}, \mathbf{s})$ [where $J(\mathbf{f}) = B(\mathbf{f}, \mathbf{f})$] if and only if the following three conditions are satisfied.

i) The solution $\mathbf{f}_{\text{out}}$ belongs to F as defined by (3), i.e., $J(\mathbf{f}_{\text{out}}) < \infty$.
ii) The solution $\mathbf{f}_{\text{out}}$ is consistent with the constraints (2), i.e., $\langle \mathbf{R}, \mathbf{f}_{\text{out}}\rangle = \mathbf{s}$.
iii) The coefficients λ_i are orthogonal in the sense of (55), i.e., $\boldsymbol{\lambda}^T\langle \mathbf{R}, \mathbf{p}_k\rangle = \mathbf{0}$, $f_{\text{out}}\ \forall\ k$.

This provides a linear system of $P + Q$ equations with $P + Q$ unknowns, which can be solved exactly. The practical aspects will be dealt with in our companion paper [1]. □

Symbolically, we can combine the pair of equations (50) by substituting $\mathbf{g} = \mathbf{f}_{\text{out}}$, yielding a very simple expression for the optimal value of the criterion $J(\mathbf{f}_{\text{out}})$

$$J(\mathbf{f}_{\text{out}}) = \boldsymbol{\lambda}^T\mathbf{s} \tag{57}$$

where $\mathbf{s}$ is the measurement vector.

D. Finding the Fundamental Solutions

To find the fundamental solutions $\boldsymbol{\varphi}_i$ as defined by (52), it is useful to start from the convolutional formulation of the bilinear form (12). Equation (52) becomes

$$\langle \mathbf{U} * \boldsymbol{\varphi}_i, \mathbf{g} \rangle = \langle \mathbf{r}_i, \mathbf{g} \rangle. \tag{58}$$

The fundamental solutions are defined through the distributional equations

$$\mathbf{U} * \underbrace{[\boldsymbol{\varphi}_1 \cdots \boldsymbol{\varphi}_Q]}_{\boldsymbol{\Phi}} = \mathbf{R} \tag{59}$$

where we have used the matrix form (2) to describe the sampling. The task can be broken in two parts. We first solve for Green's functions $\boldsymbol{\psi}_i$ [46]–[48], which are defined by

$$\mathbf{U} * \underbrace{[\boldsymbol{\psi}_1 \cdots \boldsymbol{\psi}_n]}_{\boldsymbol{\Psi}} = \delta(\mathbf{x}) \mathbf{I}_{n \times n}, \qquad \mathbf{x} \in \mathbb{R}^m. \tag{60}$$

Once we have the Green's functions $\boldsymbol{\psi}_i$, we get the fundamental solutions $\boldsymbol{\varphi}_i$ by convolution with the measurement operators $\mathbf{R}$:

$$[\boldsymbol{\varphi}_1 \cdots \boldsymbol{\varphi}_Q] = \boldsymbol{\Phi} = \boldsymbol{\Psi} * \mathbf{R}. \tag{61}$$

We see from (60) that since $\mathbf{U}$ is symmetric, $\boldsymbol{\Psi}$ is symmetric as well. When further $\mathbf{R}$ is symmetric, then the same holds true for $\boldsymbol{\Phi}$ [from (61)]. This is often the case as $\mathbf{R}$ is mostly diagonal.

E. Green's Functions

As an example, let us first study a simple scalar case ($n = 1$, $m = 1$). As criterion, we choose Duchon's semi-norm $\|f\|_{D_2} = \|f''\|_{L_2}$, which corresponds to $l = \delta''$ in (15), and thus, $u = l * l = (\mathrm{d}^4/\mathrm{d}x^4)\,\delta$. The corresponding Green's function must satisfy

$$u * \psi = \frac{\mathrm{d}^4 \delta}{\mathrm{d}x^4} * \psi = \frac{\mathrm{d}^4 \psi}{\mathrm{d}x^4} = \delta. \tag{62}$$

Integrating four times, we get a family of possible Green's functions $\psi(x) = x_+^3/12 + a_3 x^3 + a_2 x^2 + a_1 x + a_0$, where x_+^3 is the one-sided power function. For convenience, we choose the symmetric solution $\psi(x) = |x|^3/12$.

The Green's functions corresponding to general Duchon's semi-norms (46) are best analyzed in the Fourier domain using (36) with $\alpha = M + s$. Then, the following must hold:

$$\hat{u}\hat{\psi} = \|\boldsymbol{\omega}\|^{2\alpha} \hat{\psi} = 1 \tag{63}$$

in the distributional sense as well. Because both u and δ are radial distributions and a convolution of two radial distributions is also radial, we can find a radial solution of (63). That is why the resulting functions ψ (and φ, if the sampling r preserves radiality) are called *radial basis functions*.

The problem of finding $\hat{\psi}$ from (63) is well studied (cf. [40, p. 258]). For $2\alpha - m$ not an even integer

$$\psi(\mathbf{x}) = \mathcal{F}^{-1} \mathrm{Pf} \|\boldsymbol{\omega}\|^{-2\alpha} = c \rho^{2\alpha - m} \tag{64}$$

where $\rho = \|\mathbf{x}\|$, and c is a constant that can be calculated but which is irrelevant for our purposes. The "Pf" ("partie finie" [40]—finite part) symbol means that we are considering a distribution that coincides with the function $\|\boldsymbol{\omega}\|^{-2\alpha}$ for $\boldsymbol{\omega} \neq \mathbf{0}$, which does not hinder the validity of (63) in the distributional sense.

If $2\alpha - m$ is an even positive integer, the above formula has to be modified as

$$\psi(\mathbf{x}) = \mathcal{F}^{-1} \mathrm{Pf} \|\boldsymbol{\omega}\|^{-2\alpha} = c_0 \rho^{2\alpha - m} \log \rho + c_1 \rho^{2\alpha - m}. \tag{65}$$

For our task, we do not have to consider the $c_1 \rho^{2\alpha - m}$ part of (65) because their linear combination is a quadratic polynomial that necessarily belongs to $\mathcal{K}$ so that the solution belongs to F.

For $2\alpha - m$ even, it is actually easier to work directly in the space domain. If we have a radial function $g(\rho)$ that satisfies $g' = \rho^{1-m}$, then $\Delta g = S_m \delta$. (See Appendix D for a proof.) The constant S_m is the surface of the m-dimensional unit hypersphere.[1] For example, for $m = 2$, we get $\Delta \log \rho = 2\pi\delta$. Iteratively applying the formula for the Laplacian of a radial function

$$\Delta \psi(\mathbf{x}) = \psi''(\rho) + \frac{m-1}{\rho} \psi'(\rho) \tag{66}$$

yields $\Delta^2 \rho^2 \log \rho = \Delta(4 \log \rho + 4) = 8\pi\delta$, $\Delta^3 \rho^4 \log \rho = 64\pi\delta$, and $\Delta^4 \rho^6 \log \rho = 2304\pi\delta$. For $m = 3$, we have $\Delta \rho^{-1} = 4\pi\delta$, $\Delta^2 \rho = 8\pi\delta$, $\Delta^3 \rho^3 = 96\pi\delta$, etc.

Generally, Duchon's semi-norm $\|f\|_{D_{M,s}}$ leads to a fundamental solution $\varphi(\rho) = c\rho^{2(M+s)-m}$ if the exponent is not even or $\varphi(\rho) = c\rho^{2(M+s)-m} \log \rho$ otherwise. This permits us to choose from the continuum of Duchon's semi-norms the one that suits us best.

In the multidimensional ($n > 1$) case, where $\mathbf{U} = u\mathbf{I}$, we get simply $\boldsymbol{\Psi} = \psi \mathbf{I}$.

F. Unicity of the Solution

Let us suppose that the set of fundamental solutions $\{\boldsymbol{\varphi}_i\}$ and a finite basis $\{\mathbf{p}_k\}$ exist. Then, there is a set of linear equations to determine the unknowns λ_i and a_k. If this set has a unique solution, the interpolation problem will also have a unique solution, provided, of course, that (55) implies $\mathbf{f} \in F$.

In the scalar case, Micchelli [42] proved that the matrix $(\mathbf{A})_{ij} = \varphi(\mathbf{x}_i - \mathbf{x}_j)$, corresponding to the fundamental part of the solution (56), is nonsingular, provided that ϕ' [defined by $\varphi(\mathbf{x}) = \phi(\rho^2)$] is completely monotonic[2] but not constant on $]0, \infty[$, ϕ is continuous on $[0, \infty[$ and positive on $]0, \infty[$, and $\mathbf{x}_j$ are distinct. Powell [23] has additionally shown that if $\mathcal{K}$ is the space of polynomials of order $M - 1$ and if either $\phi^{(M)}$ or $-\phi^{(M)}$ is strictly completely monotonic on $]0, \infty[$, then the complete system of equations is nonsingular, provided that the $\mathbf{x}_j$ are distinct and that there is no nonzero polynomial Q of order $M - 1$ such that $Q(\mathbf{x}_j) = 0$ for all j. This is closely related to our observation in Section III-A about the unicity of the solution. The radial functions $\varphi(\rho) = \rho^\beta$ or $\rho^\beta \log \rho$, stemming from Duchon's semi-norms $\|f\|_{D_{M,s}}$ with $\beta = 2(M + s) - m$, are completely monotonic [22], [23].

[1] $S_m = 2\pi^{m/2}/\Gamma(m/2)$, which for $m = 2$, $m = 3$ yields the familiar values 2π and 4π [48].

[2] A function ϕ is completely monotonic if it is in C^∞ and $(-1)^l \phi^{(l)} \geq 0$ for $l \in \{0, 1, \ldots\}$.

VI. Approximation Problem

In some applications, it might be interesting to replace the "hard" constraints (2) by "soft" ones by adding a data term penalizing solution far from the constraints. To define a variational *approximation problem*, we introduce a combined criterion J_a. We consider the following general form:

$$J_a(\mathbf{f}) = J(\mathbf{f}) + \mathcal{L}\left(\langle \mathbf{R}, \mathbf{f}\rangle, \mathbf{s}\right) \tag{67}$$

where $\mathcal{L}$: $\mathbb{R}^Q \times \mathbb{R}^Q \to \mathbb{R}$ is an arbitrary distance function. We use it to measure the distance between the measurements and the sampled solution $\langle \mathbf{R}, \mathbf{f}\rangle$. We then say that $\mathbf{f}_{\text{out}} \in F$ is a solution to an approximation problem $\mathcal{A}(J_a, \mathbf{R}, \mathbf{s})$ iff, for all functions $\mathbf{f} \in F$, we have $J_a(\mathbf{f}) \geq J_a(\mathbf{f}_{\text{out}})$.

The problem of solving the approximation problem $\mathcal{A}$ is closely related to the constrained problem $\mathcal{P}$, as demonstrated by the following theorem.

Theorem 3 (AP Solution): Let us denote $J_{\min}(\mathbf{z})$ as the criterion value $J(\mathbf{f})$ of a function $\mathbf{f}$ solving an interpolation problem $\mathcal{P}(J, \mathbf{R}, \mathbf{z})$. Let us further define $\mathbf{z}_*$ as

$$\mathbf{z}_* = \arg\min_{\mathbf{z}} \left(J_{\min}(\mathbf{z}) + \mathcal{L}(\mathbf{z}, \mathbf{s})\right). \tag{68}$$

Then, $\mathbf{f}_{\text{out}}$ solves the problem $\mathcal{A}(J_a, \mathbf{R}, \mathbf{s})$ iff it solves the interpolation problem $\mathcal{P}(J, \mathbf{R}, \mathbf{z}_*)$.

The proof follows from the observation that the data term in J_a depends only on the measurements $z_i = \langle \mathbf{r}_i, \mathbf{f}\rangle$ of the solution $\mathbf{f}$. Thus, the minimization

$$\mathbf{f}_{\text{out}} = \arg\min_{\mathbf{f}\in F} J_a(\mathbf{f}) \tag{69}$$

required to solve $\mathcal{A}$ can be broken into two parts: a) external minimization with respect to the $\mathbf{z}$ and b) internal minimization trying to find the proper $\mathbf{f}_{\text{out}}$ minimizing J given $\mathbf{z}$. We see that the internal optimization is exactly the constrained variational problem described previously. Once it is solved, the external minimization becomes a standard multidimensional optimization problem that can be solved by existing numerical methods [49] or, in some special cases, analytically (see the next section for an example).

A. Least-Squares Approximation

Often, the general criterion (67) can be replaced by a simple least-squares form

$$J_a(\mathbf{f}) = J(\mathbf{f}) + \gamma \sum_{i=1}^{Q} \left(\langle \mathbf{r}_i, \mathbf{f}\rangle - s_i\right)^2. \tag{70}$$

We first realize that according to Theorem 3, the solution has the form (56). We then use the method of small perturbations by evaluating $J_a(\mathbf{f}_{\text{out}} + \alpha\mathbf{g})$, similarly to Section V-A. Its derivative with respect to α needs to be zero for all $\mathbf{g} \in F$ in order for $\mathbf{f}_{\text{out}}$ to be a minimum. This implies

$$2B(\mathbf{f}_{\text{out}}, \mathbf{g}) + 2\gamma \sum_i \langle \mathbf{r}_i, \mathbf{g}\rangle \left(\langle \mathbf{r}_i, \mathbf{f}_{\text{out}}\rangle - s_i\right) = 0. \tag{71}$$

Identifying with (49) gives

$$-\boldsymbol{\lambda} = \gamma\left(\langle \mathbf{R}, \mathbf{f}_{\text{out}}\rangle - \mathbf{s}\right). \tag{72}$$

Substituting the solution (56) in (71) yields

$$\gamma^{-1}\lambda_i = s_i - \sum_{j=1}^{Q} \lambda_j \langle \mathbf{r}_i, \boldsymbol{\varphi}_j\rangle - \sum_{k=0}^{P-1} a_k \langle \mathbf{r}_i, \mathbf{p}_k\rangle. \tag{73}$$

By taking $\mathbf{g} \in \mathcal{K}$ in (71), we get the same orthogonality constraints (55) as in the interpolation case. Here too, the computational procedure boils down to the solution of a linear system of equations that is functionally identical to the one encountered for the interpolation problem $\mathcal{P}$, except for a diagonal regularizing term $\gamma^{-1}\lambda_i$, which stabilizes the system, cf. [1].

VII. Conclusions

We have presented a systematic way of solving variational problems minimizing quadratic regularization criteria under general linear constraints. We have also considered replacing the constraints by a corresponding penalty function, and we have shown that it leads to a solution with the same form, regardless of the penalty function. The solution of such problems lies in a vector space uniquely corresponding to the problem at hand, generated by a system of fundamental solutions, related to Green's functions. We have shown how the requirements we impose on the variational problem solution determine the choice of the criterion, leading to the family of semi-norms introduced by Duchon.

We are now ready to proceed with the application part of this research in the companion paper [1].

Appendix

A. Linearity With Respect to Measurements

Let us have a function $\mathbf{v}$ from F_0, where $F_0 = \{\mathbf{v} \in F; \forall\, i;\ \langle \mathbf{r}_i, \mathbf{v}\rangle = 0\}$. By linearity, $\gamma\mathbf{v}$ also belongs to F_0 for $\gamma \in \mathbb{R}$. We then have $J(\mathbf{f} + \gamma\mathbf{v}) \geq J(\mathbf{f})$ because $\mathbf{f}$ solves $\mathcal{P}$. Consequently, $2\gamma B(\mathbf{f}, \mathbf{v}) + \gamma^2 B(\mathbf{v}, \mathbf{v}) \geq 0$, and thus, $\gamma B(\mathbf{f}, \mathbf{v}) \geq 0$ for sufficiently small (positive or negative) γ, which implies

$$B(\mathbf{f}, \mathbf{v}) = 0 \qquad \text{for any } \mathbf{v} \text{ from } F_0. \tag{74}$$

This leads to $J(\alpha\mathbf{f} + \beta\mathbf{g} + \mathbf{v}) - J(\alpha\mathbf{f} + \beta\mathbf{g}) = 2B(\alpha\mathbf{f} + \beta\mathbf{g}, \mathbf{v}) + J(\mathbf{v}) = J(\mathbf{v}) \geq 0$, which proves that $\alpha\mathbf{f} + \beta\mathbf{g}$ solves the problem with measurements $\alpha\mathbf{r} + \beta\mathbf{s}$ when $\mathbf{f}$ and $\mathbf{g}$ solve problems with measurements $\mathbf{r}$ and $\mathbf{s}$, respectively. ■

B. Difference Between Two Solutions

We prove that if two functions $\mathbf{f}_1$ and $\mathbf{f}_2$ both minimize $J(\mathbf{f})$ under some constraints (2), then $J(\mathbf{f}_1 - \mathbf{f}_2) = \mathbf{0}$. Using (74), we deduce $B(\mathbf{f}_1, \mathbf{f}_1 - \mathbf{f}_2) = B(\mathbf{f}_2, \mathbf{f}_1 - \mathbf{f}_2) = \mathbf{0}$. This directly yields $J(\mathbf{f}_1 - \mathbf{f}_2) = B(\mathbf{f}_1, \mathbf{f}_1 - \mathbf{f}_2) + B(\mathbf{f}_2, \mathbf{f}_2 - \mathbf{f}_1) = \mathbf{0}$. ■

C. Interpolation Problem Solution

First, suppose that $\mathbf{f}_{\text{out}}$ solves $\mathcal{P}$. Then, by definition, $\mathbf{f}_{\text{out}} \in F$, and $\langle \mathbf{R}, \mathbf{f}_{\text{out}}\rangle = \mathbf{s}$. Equations (51) and (2) are valid by construction. As $\mathbf{p}_k \in \mathcal{K}$, we have (55). Conversely, suppose that $\langle \mathbf{R}, \mathbf{f}_{\text{out}}\rangle = \mathbf{s}$ and (55) holds. Formula (56) for $\mathbf{f}_{\text{out}}$ gives $B(\mathbf{f}_{\text{out}}, \mathbf{g}) = \sum_i \lambda_i B(\boldsymbol{\varphi}_i, \mathbf{g})$. Substituting (52) leads to $B(\mathbf{f}, \mathbf{g}) = \sum_i \lambda_i \langle \mathbf{r}_i, \mathbf{g}\rangle$ for all $\mathbf{g} \in \mathcal{D}$. As $\mathbf{f} \in F$ and $\mathcal{D}$ is dense in F,

the preceding formula holds also for all $\mathbf{f} \in F$, which permits us to apply Lemma 1. ∎

D. Dirac Laplacian

Consider $\langle \Delta g, v \rangle$, where v is a test function, and $g(\mathbf{x})$ is a radial function satisfying $\partial g/\partial \rho = \rho^{1-m}$. This scalar product equals $-\int \nabla g \nabla v \, \mathrm{d}\mathbf{x}$. We change to spherical coordinates $\mathbf{x} \rightarrow (\rho, \phi_1, \ldots, \phi_{m-1})$. The integral becomes $-\int g' (\partial v/\partial \rho) \rho^{m-1} \, \mathrm{d}r \, \mathrm{d}\mathbf{\Omega}$, where $\mathrm{d}\mathbf{\Omega} = \mathrm{d}\phi_1 \cdots \mathrm{d}\phi_{m-1}$ and $\int \mathrm{d}\mathbf{\Omega} = S_m$. We use the fact that $\rho^{m-1} g' = 1$. Then, by integration over ρ, we get $-\int [v(\rho)]_{\rho=0}^{\rho=\infty} \mathrm{d}\mathbf{\Omega}$. As v is a test function $v(\infty) = 0$, and the integral simplifies to $-S_m [v]_0^\infty = S_m v(0)$. Consequently, $\Delta g = S_m \delta$. ∎

REFERENCES

[1] J. Kybic, T. Blu, and M. Unser, "Generalized sampling: A variational approach. Part II—Applications," *IEEE Trans. Signal Processing*, pp. 1977–1985.

[2] C. E. Shannon, "Communication in the presence of noise," *Proc. IRE*, vol. 37, pp. 10–21, Jan. 1949.

[3] A. Papoulis, "Generalized sampling expansion," *IEEE Trans. Circuits Syst.*, vol. CAS-24, pp. 652–654, 1977.

[4] J. L. Brown, Jr. and K. Sangsari, "Sampling reconstruction of n-dimensional band-limited images after multi-linear filtering," *IEEE Trans. Circuits Syst.*, pp. 1035–1038, July 1989.

[5] D. Seidner, M. Feder, D. Cubanski, and S. Blackstock, "Introducing to vector sampling expansions," *IEEE Signal Processing Lett.*, vol. 5, pp. 115–117, May 1998.

[6] G. Thomas, "A comparison of motion-compensated interlace-to-progressive conversion methods," *Signal Process.: Image Commun.*, vol. 12, no. 3, pp. 209–229, 1998.

[7] E. Bellers and G. de Haan, "New algorithm for motion estimation on interlaced video," *Proc. SPIE, Visual Commun. Image Process.*, vol. 3309, pp. 111–121, 1998.

[8] H. Ur and D. Gross, "Improved resolution from subpixel shifted pictures," *Comput. Vision, Graph., Image Process.*, vol. 54, no. 2, pp. 181–186, 1991.

[9] H. Shekarforoush, M. Berthod, and J. Zerubia, "3D super-resolution using generalized sampling expansion," in *Proc. Int. Conf. Image Process.*, vol. II, Washington, DC, 1995, pp. 300–303.

[10] M. Unser and A. Aldroubi, "A general sampling theory for nonideal acquisition devices," *IEEE Trans. Signal Processing*, vol. 42, pp. 2915–2925, Nov. 1994.

[11] M. Unser and J. Zerubia, "A generalized sampling theory without band-limiting constraints," *IEEE Trans. Circuits Syst. II*, vol. 45, pp. 959–969, Aug. 1998.

[12] J. C. Carr, W. R. Fright, and R. K. Beatson, "Surface interpolation with radial basis functions for medical imaging," *IEEE Trans. Med. Imag.*, vol. 16, pp. 96–107, Feb. 1997.

[13] F. Bookstein, *Morphometric Tools for Landmark Data: Geometry and Biology*. Cambridge, U.K.: Cambridge Univ. Press, 1997.

[14] A. W. Toga, Ed., *Brain Warping*. San Diego, CA: Academic, 1999.

[15] B. Kim, J. Boes, K. Frey, and C. Meyer, "Mutual information for automated multimodal image warping," in *Visualization in Biomedical Computing*, K. H. Höhne and R. Kikinis, Eds. New York: Springer-Verlag, 1996.

[16] H. Lester and S. R. Arridge, "Summarizing fluid registration by thin-plate spline warps with many landmarks," *Med. Imag. Understanding Anal.*, July 1997. [Online]. Available: http://www.robots.ox.ac.uk/~mvl/miua97/.

[17] H. J. Landau, "Necessary density conditions for sampling and interpolation of certain entire functions," *Acta Math.*, vol. 117, pp. 37–52, 1967.

[18] A. I. Zayed and A. G. García, "Nonuniform sampling of bandlimited signals with polynomial growth on the real axis," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1717–1721, Sept. 1997.

[19] A. Aldroubi and H. Feichtinger, "Exact iterative reconstruction algorithm for multivariate irregularly sampled functions in spline-like spaces: The l_p theory," in *Proc. Amer. Math. Soc.*, vol. 126, 1998, pp. 2677–2686.

[20] A. J. Jerri, "The Shannon sampling theorem—Its various extensions and applications: A tutorial review," *Proc. IEEE*, vol. 65, pp. 1565–1596, Nov. 1977.

[21] M. Unser, "Sampling—50 years after Shannon," *Proc. IEEE*, vol. 88, pp. 569–587, Apr. 2000.

[22] M. D. Buhmann, "Radial basis functions," *Acta Numer.*, vol. 9, pp. 1–38, 2000.

[23] M. J. D. Powell, "The theory of radial basis function approximation in 1990," in *Advances in Numerical Analysis II: Wavelets, Subdivision Algorithms and Radial Functions*, W. A. Light, Ed. Oxford, U.K.: Oxford Univ. Press, 1992, pp. 105–210.

[24] J. Duchon, "Splines minimizing rotation-invariant semi-norms in Sobolev spaces," in *Constructive Theory of Functions of Several Variables*, W. Schempp and K. Zeller, Eds. Berlin, Germany: Springer-Verlag, 1977, pp. 85–100.

[25] ——, "Interpolation des fonctions de deux variables suivant le principe de la flexion des plaques minces," *Rev. Française d'Autom., Inform. Recherche Oper.*, vol. 10, no. 12, pp. 5–12, Dec. 1976.

[26] W. R. Madych and S. A. Nelson, "Multivariate interpolation and conditionally positive definite functions," *Approx. Theory Appl.*, vol. 4, no. 4, pp. 77–89, 1988.

[27] ——, "Multivariate interpolation and conditionally positive definite functions. II," *Math. Comput.*, vol. 54, no. 189, pp. 211–230, 1990.

[28] J. H. Ahlberg, E. N. Nilson, and J. L. Walsh, *The Theory of Splines and Their Applications*. New York: Academic, 1967.

[29] I. J. Schoenberg, "Spline functions and the problem of graduation," *Proc. Nat. Acad. Sci.*, vol. 52, pp. 947–950, 1964.

[30] I. Djokovic and P. P. Vaidyanathan, "Generalized sampling theorems in multiresolution subspaces," *IEEE Trans. Signal Processing*, vol. 45, pp. 583–599, Mar. 1997.

[31] Y. Lyubarskii and W. R. Madych, "The recovery of irregularly sampled band limited functions via tempered splines," *J. Funct. Anal.*, vol. 125, no. 1, pp. 201–222, Oct. 1994.

[32] G. G. Walter, "A sampling theorem for wavelet subspaces," *IEEE Trans. Inform. Theory*, vol. 38, pp. 881–884, Mar. 1992.

[33] J. Kovaĉević and M. Vetterli, "Nonseparable two- and three-dimensional wavelets," *IEEE Trans. Signal Processing*, vol. 43, pp. 1269–1273, May 1995.

[34] X. G. Xia and B. W. Suter, "Vector-valued wavelets and vector filter banks," *IEEE Trans. Signal Processing*, vol. 44, pp. 508–518, Mar. 1996.

[35] E. Arge, M. Daehlen, and A. Tveito, "Box spline interpolation—A computational study," *J. Comput. Appl. Math.*, vol. 44, no. 3, pp. 303–329, Dec. 1992.

[36] T. Poggio and F. Girosi, "Networks for approximation and learning," *Proc. IEEE*, vol. 78, pp. 1481–1497, Sept. 1990.

[37] R. Szeliski and J. Coughlan, "'Spline-based image registration," *Int. J. Comput. Vision*, vol. 22, pp. 199–218, 1997.

[38] P. Moulin, R. Krishnamurthy, and J. Woods, "Multiscale modeling and estimation of motion fields for video coding," *IEEE Trans. Image Processing*, vol. 6, pp. 1606–1620, Dec. 1997.

[39] M. Unser, A. Aldroubi, and M. Eden, "Recursive regularization filters: Design, properties, and applications," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, pp. 272–277, Mar. 1991.

[40] L. Schwartz, *Théorie des Distributions* (in French). Paris, France: Hermann, 1966.

[41] I. Richards and H. Youn, *Theory of Distributions: A Non-Technical Introduction*. Cambridge, U.K.: Cambridge Univ. Press, 1990.

[42] C. A. Micchelli, "Interpolation of scattered data: Distance matrices and conditionally positive definite functions," *Constr. Approx.*, no. 2, pp. 11–22, 1986.

[43] J. Maltz, R. De Mello Koch, and A. Willis, "Reproducing kernel Hilbert space method for optimal interpolation of potential field data," *IEEE Trans. Image Processing*, vol. 7, pp. 1725–1730, Dec. 1998.

[44] G. H. Golub and C. F. Van Loan, *Matrix Computations*, third ed. Baltimore, MD: Johns Hopkins Univ. Press, 1996.

[45] G. Wahba, *Spline Models for Observational Data*. Philadelphia, PA: SIAM, 1990.

[46] D. G. Duffy, "Green's functions with applications," in *Studies in Advances Mathematics*. Boca Raton, FL: Chapman & Hall/CRC, 2001.

[47] G. Barton, *Elements of Green's Functions and Propagation*. New York: Oxford Univ. Press, 1989.

[48] E. W. Weisstein, *CRC Concise Encyclopedia of Mathematics*. Boca Raton, FL: Chapman & Hall/CRC, 1998.

[49] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C*, second ed. Cambridge, U.K.: Cambridge Univ. Press, 1992.

Jan Kybic (M'00) was born in Prague, Czech Republic, in 1974. He received the Mgr. (B.Sc.) and Ing. (M.Sc.) degrees with honors from the Czech Technical University, Prague, in 1996 and 1998, respectively. In 2001, he received the Ph.D. degree in biomedical image processing from Ecole Polytechnique Federale de Lausanne (EPFL), Lausanne, Switzerland, for his thesis on elastic image registration using parametric deformation models.

He is now with INRIA, Sophia-Antipolis, France. His research interests include signal and image processing in general, image registration, splines and wavelets, speech processing and enhancement, computer vision, numerical methods, algorithm theory, and control theory.

Thierry Blu (M'96) was born in Orléans, France, in 1964. He received the Diplôme d'ingénieur" from École Polytechnique, Paris, France, in 1986 and the M.Sc. degree from Télécom Paris (ENST), in 1988. He received the Ph.D. in electrical engineering from ENST in 1996 for a study on iterated rational filterbanks applied to wideband audio coding.

He is currently with the Biomedical Imaging Group, Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland, on leave from the France Télécom National Center for Telecommunications Studies (CNET), Issy-les-Moulineaux, France. His research interests include (multi)wavelets, multiresolution analysis, multirate filterbanks, approximation and sampling theory, psychoacoustics, etc.

Michael Unser (M'89–SM'94–F'99) received the M.S. (summa cum laude) and Ph.D. degrees in electrical engineering in 1981 and 1984, respectively, from the Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland.

From 1985 to 1997, he was with the Biomedical Engineering and Instrumentation Program, National Institutes of Health, Bethesda, MD. He is now Professor and Head of the Biomedical Imaging Group at EPFL. His main research area is biomedical image processing. He has a strong interest in sampling theories, multiresolution algorithms, wavelets, and the use of splines for image processing. He is the author of 90 published journal papers in these areas. He was on the editorial board of *Signal Processing*.

Dr. Unser is an Associate Editor for the IEEE TRANSACTIONS ON MEDICAL IMAGING. He has been on the editorial boards the IEEE TRANSACTIONS ON IMAGE PROCESSING from 1992 to 1995 and the IEEE SIGNAL PROCESSING LETTERS from 1994 to 1998. He serves as regular chair for the SPIE Conference on Wavelets, held annually since 1993. He received the 1995 Best Paper Award and the 2000 Magazine Award from the IEEE Signal Processing Society.