Generalized Sampling: A Variational Approach

Jan Kybic, Thierry Blu, and Michael Unser

March 12, 2001

Biomedical Imaging Group, IOA, DMT
Swiss Federal Institute of Technology Lausanne
CH-1015 Lausanne, Switzerland
Jan.Kybic@epfl.ch

Abstract

We consider the problem of reconstructing a multidimensional and multivariate function $\mathring{\boldsymbol{f}} : \mathbb{R}^m \to \mathbb{R}^n$ from the discretely and irregularly sampled responses of q linear shift-invariant filters. Unlike traditional approaches which reconstruct the function in some signal space V, our reconstruction is optimal in the sense of a plausibility criterion J. The reconstruction is either consistent with the measures, or minimizes the consistence error. There is no band-limiting restriction for the input signals. We show that important characteristics of the reconstruction process are induced by the properties of the criterion J. We give the reconstruction formula and apply it to several practical cases.

1 Introduction

We will deal with the problem of finding a reconstruction $\boldsymbol{f} \in F$ of a multidimensional function $\mathring{\boldsymbol{f}} : \mathbb{R}^m \to \mathbb{R}^n$, using a set of samples $y_{ij} = \sum_k \big(h_{jk} * \mathring{f}_k\big)(\boldsymbol{x}_i) = \big(\boldsymbol{h}_j * \mathring{\boldsymbol{f}}\big)(\boldsymbol{x}_i)$ from a filter bank $\boldsymbol{H} = [\boldsymbol{h}_1 \dots \boldsymbol{h}_q]$ sampled at N locations $\boldsymbol{x}_i$.

The Shannon theory states that a band-limited signal f_o can be reconstructed exactly from its regularly spaced ideal samples ($h = \delta$). Papoulis [1] has shown that f_o may also be recovered from the output of q linear shift-invariant systems sampled at $(1/q)$-th the Nyquist rate. This theory has been further extended to multivariate [2] ($m > 1$) and multidimensional [3] ($n > 1$) functions. Unser and Zerubia [4] generalized this framework by dropping the band-limiting constraint. They sought an approximation f in the more general space $V(\varphi)$, generated by integer translates of a function φ. Their approximation f is consistent in the sense of producing the same measurements y_{ij} as $\mathring{f}$. For $\varphi = \mathrm{sinc}$, their reconstruction formulas are equivalent to those of Papoulis.

We will take a slightly different approach in the present paper. We keep the consistency constraint; we require that $\boldsymbol{f}$ and $\mathring{\boldsymbol{f}}$ be indistinguishable through our measurement system, i.e., $\big(\boldsymbol{h}_j * \boldsymbol{f}\big)(\boldsymbol{x}_i) = y_{ij} = \big(\boldsymbol{h}_j * \mathring{\boldsymbol{f}}\big)(\boldsymbol{x}_i)$. However, instead of prescribing a reconstruction space V, we seek a solution optimal in the sense of a plausibility criterion (penalty function) $J(\boldsymbol{f})$. In other words, we replace the sub-space constraint $\boldsymbol{f} \in V$ by a variational formulation. The criterion $J(\boldsymbol{f})$ provides the regularization needed to overcome the ambiguity of the reconstruction problem. It may also represent the *a priori* knowledge in the Bayesian framework, quantifying our confidence that a particular function $\boldsymbol{f}$ is close to the input $\mathring{\boldsymbol{f}}$.

2 Variational criterion

We define the solution to the reconstruction problem to be a function $\boldsymbol{f}$ minimizing $J(\boldsymbol{f})$ under the consistency constraints. Thus, the behavior of the reconstruction algorithm is completely described by the criterion J. We will work in a space F of functions for which J is defined. We will assume that J is a seminorm and can be written as $J(\boldsymbol{f}) = B(\boldsymbol{f}, \boldsymbol{f})^{1/2}$, where B is a bilinear form on F. This not only simplifies the subsequent analysis, but also insures the convexity and continuity of the criterion, which implies that all local minima are also global minima. As J is a semi-norm, not a norm, there is a kernel $K \subseteq F$ for which $J(\boldsymbol{f}) = 0$. It can be shown that if two functions $\boldsymbol{f}_1$, $\boldsymbol{f}_2$ solve the reconstruction problem, then $\boldsymbol{f}_1 - \boldsymbol{f}_2 \in K$. The bilinearity of B also makes the superposition principle applicable on the reconstruction process: a solution corresponding to a linear combination of sampling values corresponds to the same linear combination of solutions. In the multidimensional case ($n > 1$), it is usually desirable that all components be treated equally. Then the superposition principle implies invariance with respect to the rotation of the sampled values, and any other linear operation on them.

It can be shown that the invariance properties of the semi-norm J and the filter bank $\boldsymbol{H}$ translate directly to the invariance properties of the reconstruction problem solution. For example, if the value of $J(\boldsymbol{f})$ and remains unchanged when $\boldsymbol{f}$ is subject to translation and rotation, then translating or rotating the sampling points results in a solution which is a translation or rotation of the original solution, provided that the sampling is also translation and rotation invariant. In fact, instead of demanding complete invariance, it is enough to have a pseudo-invariance, where $J(\boldsymbol{f})$ is allowed to be transformed by an arbitrary increasing function independent of $\boldsymbol{f}$, such as multiplied by a constant. This greatly simplifies the task of having a scale change invariant reconstruction problem, because creating a scale pseudo-invariant J is straightforward, while truly scale-invariant J does not exist, except in the trivial cases. Finally, in many applications, we do not want to penalize linear polynomials, as they correspond to the purest form of the solution.

2.1 Proposed criterion

Is there a criterion corresponding to all the above mentioned requirements? The simplest one in the univariate/unidimensional case ($m = n = 1$) is the criterion proposed by Duchon [5]: $J(f) = \left(\int \left(\partial^2 f/\partial x^2\right)^2 \mathrm{d}x\right)^{1/2}$ with a corresponding bilinear form

$$B(f,g) = \int (\partial^2 f/\partial x^2)\,(\partial^2 g/\partial x^2)\,\mathrm{d}x$$

For arbitrary higher m and n this generalizes as

$$B(\boldsymbol{f},\boldsymbol{g}) = \int \sum_{i=1}^{n} \sum_{\sum j_k = 2} \frac{\partial^2 f_i}{\partial x_1^{j_1} \cdots \partial x_m^{j_m}} \frac{\partial^2 g_i}{\partial x_1^{j_1} \cdots \partial x_m^{j_m}}\,\mathrm{d}\boldsymbol{x}$$

$$J(\boldsymbol{f}) = \sqrt{B(\boldsymbol{f},\boldsymbol{f})} \tag{1}$$

For $m = 2$ this criterion leads to the often used *thin-plate splines*. It is not difficult to verify that J is invariant by rotation and translation, and pseudo-invariant by scaling. The kernel K corresponding to this semi-norm consists of linear polynomials $a_0 + \sum a_i x_i$.

If $\boldsymbol{g}$ is a *test function* (indefinitely differentiable and compactly supported), the bilinear form from (1) can be, by integration *per partes*, rewritten using the m-dimensional Laplacian as:

$$\begin{aligned} B(\boldsymbol{f},\boldsymbol{g}) &= \int \sum_{i=1}^{n} \sum_{\sum j_k = 4} \frac{\partial^4 f_i}{\partial x_1^{j_1} \cdots \partial x_m^{j_m}}\, g_i\,\mathrm{d}\boldsymbol{x} \\ &= \int \boldsymbol{g}^T \Delta^2 \boldsymbol{f}\,\mathrm{d}\boldsymbol{x} \end{aligned} \tag{2}$$

A large class of translation invariant bilinear forms can be (under suitable restrictions on $\boldsymbol{f}$) expressed as $B = \langle \mathcal{U} * \boldsymbol{f}, \boldsymbol{g} \rangle$, where $\mathcal{U}$ is an $n \times n$-matrix of m-variate distributions, a *convolutional kernel* of the bilinear form [6]. In our case $\mathcal{U} = \Delta * \Delta \overset{\text{def}}{=} \Delta^2$, with Fourier transform $\hat{\mathcal{U}} = \|\boldsymbol{\omega}\|^4 \boldsymbol{I}$. For example for $n = m = 1$ and $n = 1$, $m = 2$, we have $\mathcal{U} = \delta^{(\mathrm{IV})}$ and $\mathcal{U} = \delta^{(\mathrm{IV})}(x_1)\delta(x_2) + 2\delta''(x_1)\delta''(x_2) + \delta^{(\mathrm{IV})}(x_2)\delta(x_1)$, respectively, where δ is Dirac's mass distribution centered at 0 and δ'', $\delta^{(\mathrm{IV})}$ are its second and fourth derivatives.

3 Explicit solution

A solution to the reconstruction problem is given by a remarkably simple formula:

$$\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{p}(\boldsymbol{x}) + \sum_{i=1}^{N} \sum_{j=1}^{q} \lambda_{ij} \boldsymbol{\varphi}_j(\boldsymbol{x} - \boldsymbol{x}_i) \tag{3}$$

It consists of two parts. The first part ($\boldsymbol{p}$) belongs to the kernel K. If it has a countable basis, we can write it as $\boldsymbol{p}(\boldsymbol{x}) = \sum_k a_k \boldsymbol{p}_k(\boldsymbol{x})$. It does not contribute to the criterion ($J(\boldsymbol{p}) = 0$), so we intuitively see that it is useful to accommodate as much as possible of $\boldsymbol{f}$ in this part. The second part consists of a linear combination of generating functions $\boldsymbol{\varphi}_j$ shifted to all the sampling points. There are q generating functions, where q is the number of sampling filters $\boldsymbol{h}$. In the case of regularly spaced $\boldsymbol{x}_i$ and $q = 1$, we recover the solution of Unser and Zerubia [4], provided that we use a criterion corresponding to the adequate generating function φ. For $q > 1$, we have a multi-wavelet-like representation with several basis functions.

As a consequence of the minimization process, the solution $\boldsymbol{f}$ must satisfy

$$B(\boldsymbol{f}, \boldsymbol{g}) = \sum_{ij} \lambda_{ij} \big(\boldsymbol{h}_j * \boldsymbol{g}\big)(\boldsymbol{x}_i) \quad \text{for any} \quad \boldsymbol{g} \in F \quad (4)$$

where $\boldsymbol{g}$ is an arbitrary variation around the optimal $\boldsymbol{f}$ and λ_{ij} a Lagrange multiplier corresponding to a sample y_{ij}. It is enough to consider $\boldsymbol{g}$ only from among test functions, assuming that the test functions are dense in F, which is normally the case. This justifies our earlier restriction.

Provided that $\boldsymbol{\varphi}_j$ are *fundamental solutions* as detailed in the next section, the condition (4) translates into two sets of constraints. The first set makes $\boldsymbol{f}$ satisfy the interpolation conditions $y_{ij} = \big(\boldsymbol{h}_j * \boldsymbol{f}\big)(\boldsymbol{x}_i)$. This leads to qN equations. The second set must ensure that the criterion $J(\boldsymbol{f})$ is defined, that is, $\boldsymbol{f}$ must belong to F. This implies an orthogonality condition $B(\boldsymbol{f}, \boldsymbol{k}) = 0$, for all $\boldsymbol{k}$ from the kernel K. If we know its basis, then also $B(\boldsymbol{f}, \boldsymbol{p}_k) = 0$ for all $\boldsymbol{p}_k$. Therefore, the second set contains $\dim K$ equations, which makes $qN + \dim K$ equations for as many unknown coefficients λ_{ij} and a_k.

For the criterion J from (1), the second set of constraints implies that the second derivatives of $\boldsymbol{f}$ are square integrable. Therefore, they tend to zero (on the average, not necessarily pointwise) towards infinity and thus f tends to a linear polynomial.

3.1 Generating functions

The generating functions $\boldsymbol{\varphi}_j$ are fundamental solutions satisfying $\forall \boldsymbol{g} \in F; B(\boldsymbol{\varphi}_j, \boldsymbol{g}) = \big(\boldsymbol{h}_j * \boldsymbol{g}\big)(0)$. In the distributional setting, this condition translates to $\mathcal{U} * \boldsymbol{\varphi}_j = \boldsymbol{h}_j$. This corresponds to $\hat{\mathcal{U}}\hat{\boldsymbol{\varphi}}_j = \hat{\boldsymbol{h}}_j$, provided that the Fourier transforms $\hat{\mathcal{U}}$, $\hat{\boldsymbol{\varphi}}_j$ and $\hat{\boldsymbol{h}}_j$ exist. More concisely, in the matrix form, we get $\mathcal{U} * \boldsymbol{\Phi} = \boldsymbol{H}$ and $\hat{\mathcal{U}}\hat{\boldsymbol{\Phi}} = \hat{\boldsymbol{H}}$, where $\boldsymbol{\Phi} = [\boldsymbol{\varphi}_1 \dots \boldsymbol{\varphi}_q]$.

It can be inferred that rotationally invariant seminorms correspond to radial kernels $\mathcal{U}$ which in turn lead to radial generating functions $\boldsymbol{\varphi}$, provided that the filters $\boldsymbol{h}_j$ are also radial. We can then write $\boldsymbol{\varphi}(\boldsymbol{x}) = \boldsymbol{\rho}(\|\boldsymbol{x}\|) = \boldsymbol{\rho}(r)$.

Fundamental solutions for the iterated Laplacian kernel Δ^2 and $h = \delta$ are well-known and can be found taking the inverse Fourier transform of $\|\boldsymbol{\omega}\|^4$ (we are omitting some technical details here). For $n = 1$ and $m = 1, 2, 3$ we get $\rho = r^3$, $\rho = r^2 \log r$, and $\rho = r$, respectively, neglecting the multiplicative constant. In the multidimensional case, as the components are treated equally, we simply use the same function for all components, i.e. $\boldsymbol{\Phi} = \rho(r)\boldsymbol{I}$.

We have found the fundamental solution $\boldsymbol{\varphi}_\delta$ for ideal (zero-order) sampling $\boldsymbol{h} = \delta$. For other samplers, we have simply $\boldsymbol{\varphi} = \boldsymbol{\varphi}_\delta * \boldsymbol{h}$.

3.2 Approximation problem

When the measures are not exact (for example corrupted by noise), it might be more appropriate to drop the consistency constraints and minimize instead a weighted sum of a plausibility criterion J_p and some measure of the consistency error, that is, the difference between the desired and actual sampled values. The approximation problem then consists of minimizing

$$J_a(\boldsymbol{f}) = J_p(\boldsymbol{f}) + \sum_{ij} d_{ij}\big(y_{ij}, z_{ij}\big); \quad z_{ij} = \big(\boldsymbol{h}_j * \boldsymbol{f}\big)(\boldsymbol{x}_i)$$

where d is a suitable distance measure. As the consistency error depends only on z_{ij}, the minimization $\min_f J(\boldsymbol{f})$ can be decomposed as $\min_{\{z_{ij}\}} \min_f J(\boldsymbol{f})$. The inner minimization is equivalent to the reconstruction problem we have solved already. Therefore the solution to the approximation problem has also the form (3).

Let us now consider an approximation problem created by adding a least-squares consistency error measure to a criterion $J^2(\boldsymbol{f})$ from (1):

$$J_a(\boldsymbol{f}) = J^2(\boldsymbol{f}) + \gamma \sum_{ij} (z_{ij} - y_{ij})^2$$

By standard variational technique we find that the solution $\boldsymbol{f}$ must verify

$$B(\boldsymbol{f}, \boldsymbol{g}) = \gamma \sum_{ij} (z_{ij} - y_{ij}) \big(\boldsymbol{h}_j * \boldsymbol{g}\big)(\boldsymbol{x}_i) \quad \forall \boldsymbol{g} \in F$$

Comparing this equation with (4), we find a linear set of equations

$$\lambda_{ij} = \gamma(z_{ij} - y_{ij}) = \gamma \sum_k a_k \big(\boldsymbol{h}_j * \boldsymbol{p}_k\big)(\boldsymbol{x}_i) + \\ + \gamma \sum_{kl} \lambda_{kl} \big(\boldsymbol{h}_j * \boldsymbol{\varphi}_l\big)(\boldsymbol{x}_i - \boldsymbol{x}_k) - \gamma y_{ij}$$

which permits us, together with the orthogonality constraints $B(\boldsymbol{f}, \boldsymbol{p}_k) = 0$, to determine λ_{ij} and a_i.

4 Examples

4.1 Reconstruction from irregular samples

Let us consider the problem of finding a function $f : \mathbb{R} \to \mathbb{R}$, passing through a finite number of points (x_i, y_i) and minimizing a criterion $J(f) = \int (f'')^2 \mathrm{d}x$. We have seen that the corresponding bilinear form is $B(f,g) = \int f'' g'' \mathrm{d}x$ with kernel $\mathcal{U} = \delta^{(\mathrm{IV})}$. The fundamental solution is proportional to $|x|^3$, which can be localized (convolved with a discrete filter) to obtain a cubic B-spline. The reconstruction is thus

$$f(x) = a_0 + a_1 x + \sum_{i=1}^{N} \lambda_i |x - x_i|^3$$

which has $N+2$ unknown parameters. The second derivative is $f''(x) = 6 \sum_i \lambda_i (x - x_i)$. This leads to orthogonality conditions $\sum \lambda_i = 0$ and $\sum \lambda_i x_i = 0$,

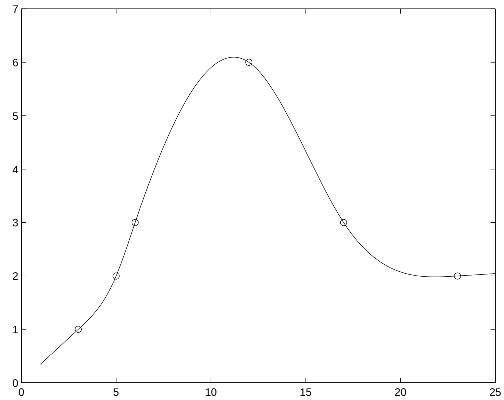

Figure 1: Interpolation from function values.

because if either of them were not satisfied we would have $\lim_{x \to \pm\infty} f'' \neq 0$ and consequently $f \notin F$. The remaining N equations come from the consistency conditions $f(x_i) = y_i$. The results by Micchelli [7] imply that for distinct points, there is always a unique solution. An example of a reconstruction result is shown in Figure 1.

4.2 Derivative sampling

Let us add derivative constraints $y_i' = f'(x_i)$ to the preceding example. The sampling filters will then be $\boldsymbol{H} = \begin{bmatrix} \delta & \delta' \end{bmatrix}$. The first fundamental solution corresponding to $h_1 = \delta$ remains $\varphi_1 = |x|^3$. The second one, corresponding to $h_2 = \delta'$, is obtain by convolving φ_1 with h_2 which gives $\varphi_2 = |x|x$. The reconstruction formula is thus

$$f(x) = a_0 + a_1 x + \\ + \sum_{i=1}^{N} \lambda_{i,1} |x - x_i|^3 + \lambda_{i,2} |x - x_i| (x - x_i)$$

The $2N+2$ unknown parameters can be determined from $2N$ consistency equations $f(x_i) = y_i$ and $y_i' = f'(x_i)$ and two orthogonality conditions $\sum \lambda_{i,1} = 0$ and $\sum \lambda_{i,2} - 3\lambda_{i,1} x_i = 0$. An example of reconstruction from derivative sampling is shown in Figure 2.

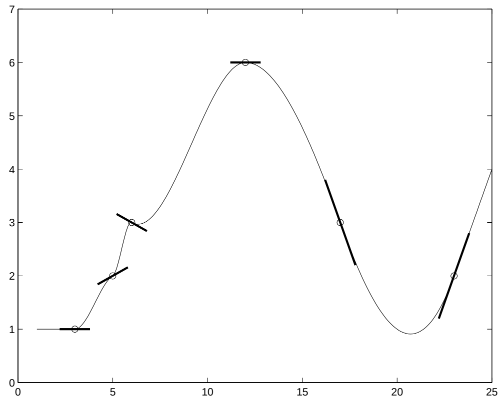

Figure 2: Interpolation from function values and derivatives.

4.3 Reconstruction consistent with Laplace equation

The problem treated in [8] by numerical integration—which we shall be able to solve explicitly here—consists of reconstructing a function from $\mathbb{R}^3 \to \mathbb{R}$ while minimizing the norm of the 3D Laplacian operator $J'(\boldsymbol{f})^2 = \int \|\Delta \boldsymbol{f}\|^2 \mathrm{d}\boldsymbol{x}$. The problem is ill-posed without additional constraints, because the kernel K' is too big, permitting an infinity of solutions with zero cost. To avoid the ambiguity, we impose $f \in F$. We then minimize the criterion (1), because for $f \in F$, the two criteria are equivalent. As expected, the solution will have the form

$$f(\boldsymbol{x}) = a_0 + \sum_{j=1}^{3} a_i x_i + \sum_{i=1}^{N} \lambda_i \|\boldsymbol{x} - \boldsymbol{x}_i\| \qquad (5)$$

where $\boldsymbol{x} = (x_1, x_2, x_3)$, with auxiliary conditions $\sum \lambda_i = 0$, $\sum \lambda_i x_{i,1} = 0$, and $\sum \lambda_i x_{i,2} = 0$.

5 Conclusions

We can reconstruct arbitrary multidimensional and multivariate functions from sampled outputs of an arbitrary filter bank. Unlike previous methods ([1–3]), our approach handles irregular sampling, does not impose band-limiting constraints, the solution is optimal in the sense of a variational criterion, can be made invariant to translations, rotations and scale changes, implicitly specifies the reconstruction space, and is usable also for noisy measurements. This comes at the cost of slightly more involved computation and less numerical stability.

References

[1] A. Papoulis, "Generalized sampling expansion," *IEEE Transactions Circ. Syst.*, vol. 24, pp. 652–654, 1977.

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

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

[4] M. Unser and J. Zerubia, "A generalized sampling theory without band-limiting constraints," *IEEE Transactions on Circuits and Systems—II: Analog and digital signal processing*, vol. 45, pp. 959–969, Aug. 1998.

[5] 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), pp. 85–100, Springer-Verlag, 1977.

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

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

[8] J. Maltz, R. De Mello Koch, and A. Willis, "Reproducing kernel Hilbert space method for optimal interpolation of potential field data," *IEEE Transactions on Image Processing*, vol. 7, Dec. 1998.