Regularized Interpolation for Noisy Images

Sathish Ramani*, *Member, IEEE*, Philippe Thévenaz and Michael Unser, *Fellow, IEEE*

***Abstract*—Interpolation is the means by which a continuously-defined model is fit to discrete data samples. When the data samples are exempt of noise, it seems desirable to build the model by fitting them exactly. In medical imaging, where quality is of paramount importance, this ideal situation unfortunately does not occur. In this paper, we propose a scheme that improves on the quality by specifying a tradeoff between fidelity to the data and robustness to the noise. We resort to variational principles which allow us to impose smoothness constraints on the model for tackling noisy data. Based on shift-, rotation-, and scale-invariant requirements on the model, we show that the L_p-norm of an appropriate vector derivative is the most suitable choice of regularization for this purpose. In addition to Tikhonov-like quadratic regularization, this includes edge-preserving total-variation-like (TV) regularization. We give algorithms to recover the continuously-defined model from noisy samples and also provide a data-driven scheme to determine the optimal amount of regularization. We validate our method with numerical examples where we demonstrate its superiority over an exact fit as well as the benefit of TV-like non-quadratic regularization over Tikhonov-like quadratic regularization.**

***Index Terms*—Interpolation, Regularization, Regularization parameter, Tikhonov functional, Total-variation functional, Splines.**

I. Introduction

INTERPOLATION is an integral part of many image-processing and biomedical algorithms [1]–[6]. It is employed in registration for performing geometric transformation (*e.g.*, sub-pixel translation and rotation) of discrete data [7]–[9]. In volumetric imaging, it is used for rescaling three-dimensional (3-D) volumes [10]. Fitting 3-D data on geometric shapes is also best done by taking the interpolation model into consideration [11]. Other applications where it plays a vital role include volume rendering for visualization of scalar fields [12]–[14], evaluation of image gradients [15], [16], and texture mapping where a two-dimensional (2-D) image is painted on a 3-D surface [17], [18]. Recently, it has also been used for modeling diffusion tensors in magnetic resonance imaging (MRI) [19].

Interpolation can be stated as the problem of fitting discrete data samples with a continuously-defined model that is usually represented as a weighted sum of shifted basis functions [4], [5]. The standard approach is to specify the model so as to fit the data exactly [6]. When the basis functions take a unit value at the origin and are zero at all other integer locations, the weights are given by the data themselves. Otherwise, they are determined by imposing an exact-fitting requirement [20]. In biomedical imaging applications, interpolation is most often carried out by an exact fit to the data. Although this is meaningful for a noise-free scenario, it is less appropriate when data samples are corrupted by noise since the model is forced to fit noise also.

To tackle noisy data, a desirable alternative is to enforce "smoothness" properties on the continuous-domain solution by means of regularization. Various authors have formulated interpolation as a variational problem to accommodate regularization constraints [21]–[29]. The resulting scheme is often termed *regularized interpolation* where the objective is to obtain the solution by minimizing a cost criterion that jointly measures the data-fitting error and the regularity of the solution. Regularized interpolation techniques can be broadly classified into digital-domain or analog-domain approaches. The former refers to the case where the solution is a discrete entity defined on a grid that is finer than that of the data—these methods specifically cater to the image-upsampling problem [21]–[23]. In the latter case, a continuously-defined solution—a smoothing-spline—is obtained by minimizing the L_2-norm of some scalar derivative of the solution—Tikhonov-like quadratic regularization—subject to certain data-fitting requirements [24]–[29].

In this paper, we concentrate on analog-domain regularized interpolation and propose to extend smoothing-spline-like approaches [24]–[29] by considering the use of non-quadratic regularization—the motivation is to overcome the shortcoming of Tikhonov-like quadratic regularization which tends to smear important signal features (*e.g.*, edges in images). We also want the solution to be invariant to translation, rotation and scaling of the coordinates. Our first contribution toward these ends is a theoretical result that states that the L_p-norm of an appropriate vector derivative is the most suitable choice of regularization with respect to these invariances. This includes edge-preserving total-variation-like (TV) regularization ($p = 1$), which we propose to use as an alternative to quadratic regularization. We consider a shift-invariant signal model and obtain the interpolation weights by minimizing the L_p-norm subject to a data-domain constraint that measures the statistical infidelity of the solution to the given data (in terms of a negative log-likelihood function). We propose to handle the problem in a numerical optimization framework. We design algorithms based on the majorize-minimize (MM) strategy [30] for performing the corresponding minimization, which constitutes our second contribution. Finally, we also provide a practical scheme for determining an appropriate balance between data-fitting and regularization. We validate our method by carrying out experiments on noisy medical

Sathish Ramani is with the Department of Electrical Engineering and Computer Science, University of Michigan, 1301 Beal Ave, Ann Arbor, MI-48109, U.S.A. E-mail: sramani@umich.edu.

Philippe Thévenaz and Michael Unser are with the Biomedical Imaging Group (BIG), École polytechnique fédérale de Lausanne (EPFL), CH-1015 Lausanne, VD, Switzerland. E-mail: philippe.thevenaz@epfl.ch, michael.unser@epfl.ch.

images and illustrate the superior performance of our method over standard interpolation. We also numerically justify that the use of TV-like non-quadratic regularization brings about further improvement over quadratic regularization.

The paper is organized as follows. In Section II, we briefly review the standard interpolation technique [20], following which we present a numerical example in Section II-A to demonstrate its poor performance when applied to noisy data. In Section III, we elaborate the proposed regularized interpolation scheme. We first specify the problem mathematically and provide hypotheses related to the data-fidelity and regularization terms. In Section IV, we present algorithms to carry out the corresponding optimization. We then focus our attention on spline-based interpolation in Section V. Section VI is dedicated to experimental results where we numerically verify the superiority of the proposed approach over other methods in the literature. In Section VII, we provide evidence that the proposed scheme achieves a reasonable tradeoff between computational cost and performance improvement and also discuss issues related to the selection of the regularization parameter. Finally, we draw our conclusions in Section VIII.

II. Standard Interpolation

Standard interpolation is the process of computing a continuously-defined function $f_{\rm int}$ which exactly fits an unknown analog signal $f_{\rm true}$ at the given sample points[1] $f_{\rm true}(\mathbf{k})$, $\mathbf{k} = (k_1, k_2, \ldots, k_d) \in \mathbb{Z}^d$. Typically, $f_{\rm int}$ is constructed for $\mathbf{x} = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$ as

$$f_{\rm int}(\mathbf{x}) = \sum_{\mathbf{k}\in\mathbb{Z}^d} f_{\rm true}(\mathbf{k})\, \varphi_{\rm int}(\mathbf{x}-\mathbf{k}), \tag{1}$$

where $\varphi_{\rm int}$ is an interpolating function such that $\varphi_{\rm int}(\mathbf{0}) = 1$ and $\varphi_{\rm int}(\mathbf{x})|_{\mathbf{x}=\mathbf{k}\in\mathbb{Z}^d\setminus\{\mathbf{0}\}} = 0$. Popular examples of $\varphi_{\rm int}$ are the linear B-spline [31] and the sinc function which perform linear and bandlimited interpolation, respectively.

An equivalent formulation of (1) has been provided in [20] for an arbitrary non-interpolating function φ by considering the integer-shift-invariant model

$$f_{\rm int}(\mathbf{x}) = \sum_{\mathbf{q}\in\mathbb{Z}^d} c[\mathbf{q}]\, \varphi(\mathbf{x}-\mathbf{q}), \tag{2}$$

where the coefficients c are determined by solving the set of linear equations: $\forall\ \mathbf{k} \in \mathbb{Z}^d$, $f_{\rm int}(\mathbf{x})|_{\mathbf{x}=\mathbf{k}} = \sum_{\mathbf{q}\in\mathbb{Z}^d} c[\mathbf{q}]\, \varphi(\mathbf{k}-\mathbf{q}) = f_{\rm true}(\mathbf{k})$, which ensures perfect fitting of the given samples. In this paper, we propose to use the shift-invariant model (2) for our regularized interpolation scheme, but instead of a perfect-fit requirement, we are going to specify c based on certain regularization requirements (see Section III) on the interpolation model in addition to the data-fitting constraint.

A. Standard Interpolation in the Presence of Noise

While standard interpolation (exact fitting) is desirable in the noise-free scenario, it can lead to unfavorable results when applied on noise-corrupted data. To demonstrate this, we consider the following experiment: First, we rotate some noise-free input image by a random angle using an interpolator that provides high-quality rotation in the noise-free scenario. Then, we add zero-mean white Gaussian noise of variance σ^2 to enforce a prescribed signal-to-noise ratio (SNR). Finally, we rotate back the noisy image with nearest-neighbor (using the rect function), linear, and cubic B-spline based interpolation. We repeat this for a fixed number of realizations and average the SNR of the output image over all realizations. We show in Figure 1 the plot of the (averaged) SNR of the output image for a range of input SNRs. We observe that piecewise linear interpolation outperforms cubic splines at high noise levels, which contradicts the noise-free behavior reported in [20].

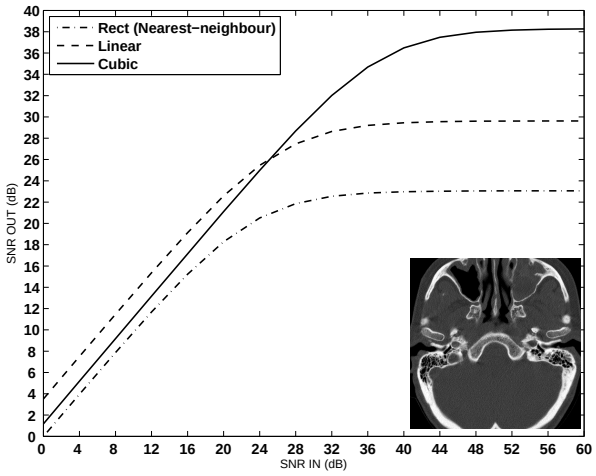


Fig. 1. Rotation experiment in the presence of noise: Piecewise-linear interpolation performs better than cubic splines at high noise levels (input SNR $<$ 24 dB), in contradiction with the expected behavior. We show in the inset the central region of the image which was used for computing SNR.

This can be qualitatively explained as follows: Let $g_{\rm int}$ be the interpolant constructed from the noise-corrupted samples g. Since standard interpolation is a linear operation and because noise is zero-mean and uncorrelated with the image, the mean-squared error (MSE) between $f_{\rm true}$ and $g_{\rm int}$ can be expressed as $\epsilon^2_{\rm mse} = \epsilon^2_{\rm int} + \epsilon^2_{\rm noise}$. Here, $\epsilon^2_{\rm int}$ is the squared-norm error between $f_{\rm true}$ and $f_{\rm int}$ which is completely characterized by the approximation order L [32], i.e., the ability of the model to reproduce polynomials of degree $n \in [0,\, L-1]$. It is known from approximation theory [32] that the higher the value of L, the lower the $\epsilon^2_{\rm int}$ error. Since $L_{\rm Cubic} > L_{\rm Linear} > L_{\rm Rect}$, we have

$$\epsilon^2_{\rm int_{Cubic}} < \epsilon^2_{\rm int_{Linear}} < \epsilon^2_{\rm int_{Rect}}. \tag{3}$$

The quantity $\epsilon^2_{\rm noise}$ is the energy of the continuous-domain signal that interpolates only the noise component in g (absence of signal); it can be shown to be proportional to $\nu^2 = \|\varphi_{\rm int}\|^2_{L_2}$. Computing ν^2 for the rect ($\nu^2_{\rm Rect} = 1$), linear ($\nu^2_{\rm Linear} = \frac{2}{3}$), and cubic B-spline ($\nu^2_{\rm Cubic} \approx 0.874$), we find that

$$\epsilon^2_{\rm noise_{Linear}} < \epsilon^2_{\rm noise_{Cubic}} < \epsilon^2_{\rm noise_{Rect}}. \tag{4}$$

Therefore, in Figure 1, at high noise levels (input SNRs in the range 0 to 24 dB) where the effect of $\epsilon^2_{\rm noise}$ is dominant, piecewise linear interpolation does better than cubic splines

[1]We concentrate on the case of uniformly-spaced samples. Moreover, we use unit-length sampling step for simplicity.

because of (4). At low noise levels (input SNR $>$ 24 dB), ϵ_{int}^2 becomes effective and the cubic splines take the lead due to (3). Nearest-neighbour interpolation has the poorest performance at all noise levels since both ϵ_{int}^2 and $\epsilon_{\text{noise}}^2$ are higher for the rect function than for linear and cubic B-splines.

III. Regularized Interpolation

A. Problem Formulation

When the input data samples are noisy, it is meaningful to adopt a variational approach to enforce regularity constraints on the interpolation model to counterbalance the effect of noise. We develop our method in a penalized-likelihood setting in the spirit of [33]–[35], where the solution is obtained by minimizing a cost functional composed of a negative log-likelihood term (also called the data-fidelity term) $\mathcal{L}\{g, f\}$ and a continuous-space regularization functional $\Psi\{f\}$. The log-likelihood measures in a statistical sense the goodness-of-fit between the samples $\{g[\mathbf{k}]\}_{\mathbf{k}\in\mathbb{Z}^d}$ and $\{f(\mathbf{k})\}_{\mathbf{k}\in\mathbb{Z}^d}$ while the regularization penalizes heavy oscillations in the solution. Mathematically, this is written as

$$f_\lambda = \arg\min_f E_\lambda\{g, f\}, \tag{5}$$

where E_λ is the cost functional given by $E_\lambda\{g, f\} = \mathcal{L}\{g, f\} + \lambda\,\Psi\{f\}$, and where $\lambda > 0$ is the regularization parameter that governs the tradeoff between goodness-of-fit and smoothness of f_λ. We shall address the problem of selecting an appropriate λ in Section VII-D and propose a practical scheme that minimizes the mean-squared error within the given class of solutions.

Before moving on, we argue that the present formulation, which addresses the signal reconstruction problem globally, is preferable conceptually to denoising the data first and then performing a standard interpolation, although the latter strategy may produce competitive results depending on the choice of the denoising algorithm. The primary argument is statistical: It can be proved that the minimization of a proper version of (5) will yield the minimum-mean-squared-error reconstruction of the signal (under the assumption that the signal is a stationary Gaussian process), provided that the basis functions are matched to the regularization operator [27], [29]. More generally, we may adopt a Bayesian point of view and use (5) to specify the maximum *a posteriori* estimator of the unknown signal, which is continuously-defined. The proposed framework is fairly general and readily extendable to more complicated situations where the data is non-uniformly sampled (and conventional denoising is not directly applicable). Finally, the proposed signal estimators can be designed to be invariant to scaling and rotation of the coordinate system, which is obviously only possible if we formulate the problem in the continuous domain.

1) Data-Fidelity Term: In the penalized-likelihood framework, we have that $\mathcal{L}\{g, f\} = -\log(q(g|f))$ where q is the probability density of $\{g[\mathbf{k}]\}_{\mathbf{k}\in\mathbb{Z}^d}$ given $\{f(\mathbf{k})\}_{\mathbf{k}\in\mathbb{Z}^d}$, or, equivalently, the probability density of the noise in the data. Here, $\mathcal{L}\{g, f\}$ is always a discrete-domain entity since it measures the statistical infidelity of the samples of f to the data. We assume that the noise is statistically independent at different sample locations. Then, the joint-probability density can be written as $q(g|f) = \prod_{\mathbf{k}} q_{\mathbf{k}}(g[\mathbf{k}]\,|f(\mathbf{k}))$, where $q_{\mathbf{k}}$ is the marginal density, so that $\mathcal{L}\{g, f\} = -\sum_{\mathbf{k}} \log(q_{\mathbf{k}}(g[\mathbf{k}]\,|f(\mathbf{k})))$.

In this paper, we deal with two specific instances of $\mathcal{L}\{g, f\}$. The first is i.i.d. zero-mean additive white Gaussian noise (AWGN), where $q_{\mathbf{k}}(g[\mathbf{k}]\,|f(\mathbf{k})) \propto e^{-\kappa(g[\mathbf{k}]-f(\mathbf{k}))^2}$. This leads to the negative log-likelihood $\mathcal{L}_{\text{Q}}\{g, f\} = \sum_{\mathbf{k}}(g[\mathbf{k}] - f(\mathbf{k}))^2$. From a signal-processing perspective, the AWGN model is often preferred for mathematical ease as the quadratic nature of $\mathcal{L}_{\text{Q}}$ simplifies the optimization process.

Next, we consider the signal-dependent Poisson model (as an instance of non-quadratic data-fidelity) that is appropriate for imaging applications such as fluorescence microscopy [36] and emission tomography [37]. Here, the sample $g[\mathbf{k}]$ represents the detector counts at the $\mathbf{k}$-th pixel and $q_{\mathbf{k}}(g[\mathbf{k}]\,|f(\mathbf{k})) = (g[\mathbf{k}]!)^{-1}\, e^{-f(\mathbf{k})}(f(\mathbf{k}))^{g[\mathbf{k}]}$. The corresponding negative log-likelihood is given by $\mathcal{L}_{\text{Poisson}}\{g, f\} = \sum_{\mathbf{k}\in\mathbb{Z}^d}(-g[\mathbf{k}]\log(f(\mathbf{k})) + f(\mathbf{k}))$, where we have neglected the additive constant log(g[**k**]!) which is irrelevant for optimization purposes.

2) Regularization: Most regularization functionals describe the continuous-space "smoothness" in terms of some derivative of the solution. For our purpose, we consider a class of multivariate regularization functionals that can be written in the form

$$\Psi\{f\} = \int_{\mathbb{R}^d} \Phi\left(\|\mathbf{L}\{f\}(\mathbf{x})\|\right)\,\mathrm{d}\mathbf{x}, \tag{6}$$

where $\mathbf{L}$ is a vector composed of s linear differential operators L_m, $m = 1, \ldots, s$, which measures the "smoothness" of f at $\mathbf{x}$ in terms of the vector-norm $\|\mathbf{L}\{f\}(\mathbf{x})\| = \sqrt{\sum_{m=1}^{s}\left(\mathrm{L}_m\{f\}(\mathbf{x})\right)^2}$, and where Φ is called the potential function that characterizes the penalty associated with $\|\mathbf{L}\{f\}\|$. Due to the non-negativity of $\|\mathbf{L}\{f\}\|$, Φ need only be specified on the set of non-negative real numbers.

Definition 1: The one-sided potential function Φ is said to be appropriate for the purpose of regularization if it is non-negative, strictly increasing and differentiable. ■

This definition is consistent with the minimization in (5) since we wish to increase the penalty whenever $\|\mathbf{L}\{f\}\|$ increases. In order for (6) to be beneficial for the interpolation problem, we additionally require that $\Psi\{f\}$ be invariant to translation, rotation, and dilation. Then, it is guaranteed that the solution is invariant to such transformations of the given data, thereby becoming independent of the data-grid. Mathematically, the invariance requirements are prescribed as follows: We want the value of $\Psi\{f\}$ to remain unchanged (up to a multiplicative constant) when f is

1) shifted by $\mathbf{x}_0 \in \mathbb{R}^d$ (translation-invariance)

$$\Psi\{f(\cdot - \mathbf{x}_0)\} = \Psi\{f\}, \tag{7}$$

2) rotated about the origin by an arbitrary angle θ (rotation-invariance)

$$\Psi\{f(\mathbf{R}_\theta \cdot)\} = \Psi\{f\}, \text{ and} \tag{8}$$

3) dilated by $\tau > 0$ (scale-invariance)

$$\Psi\left\{f\left(\frac{\cdot}{\tau}\right)\right\} = \zeta(\tau)\,\Psi\{f\}, \tag{9}$$

where $\zeta(\tau) > 0$ is an appropriate scalar that is differentiable with respect to τ; its role is to balance the regularization against a change of scale at which it is calculated.

Since $\Psi\{f\}$ is specified via the vector-norm $\|\mathbf{L}\{f\}\|$, (7)-(9) necessitates that $\|\mathbf{L}\{f\}\|$ be preserved under translation, rotation, and dilation of f, up to the Lebesgue measure in the integral (6). This curtails the choice of $\mathbf{L}$ to those that are shift-, rotation-, and scale-invariant in nature.

Definition 2: The vector-differential operator $\mathbf{L}$ is said to be a shift-, rotation-, and scale-invariant operator if $\forall\ \mathbf{x} \in \mathbb{R}^d$, $\|\mathbf{L}\{f\}(\mathbf{x})\|$ commutes with translation

$$||\mathbf{L}\{f(\cdot - \mathbf{x}_0)\}(\mathbf{x})|| = ||\mathbf{L}\{f\}(\mathbf{x} - \mathbf{x}_0)||,\ \forall\ \mathbf{x}_0 \in \mathbb{R}^d, \tag{10}$$

with rotation

$$||\mathbf{L}\{f(\mathbf{R}_\theta \cdot)\}(\mathbf{x})|| = ||\mathbf{L}\{f\}(\mathbf{R}_\theta \mathbf{x})||,\ \forall\ \theta \in [0,\, 2\pi), \tag{11}$$

and with dilation

$$\left\|\mathbf{L}\left\{f\left(\frac{\cdot}{\tau}\right)\right\}(\mathbf{x})\right\| = \rho(\tau)\ \left\|\mathbf{L}\left\{f\right\}\left(\frac{\mathbf{x}}{\tau}\right)\right\|,\ \forall\ \tau > 0, \tag{12}$$

where $\rho(\cdot) > 0$ is a differentiable function that captures the response of $\mathbf{L}$ to a scaling operation. ■

Interestingly, common multivariate differential operators such as the gradient ($\mathbf{L} = \boldsymbol{\nabla}$) and the Laplacian ($\mathbf{L} = \Delta$) turn out to be shift-, rotation-, and scale-invariant in nature. In the case of the gradient operator, we have $\mathrm{L}_m = \frac{\partial}{\partial x_m}$, $m = 1, 2, \ldots, d$. Obviously, the relation $\frac{\partial f(\mathbf{x}/\tau)}{\partial x_m} = \tau^{-1}\,\frac{\partial f(\mathbf{x})}{\partial x_m}\Big|_{\mathbf{x}=\mathbf{x}/\tau}$ implies that the response of $\boldsymbol{\nabla}$ to the dilation operation is $\rho(\tau) = \tau^{-1}$. Similarly, for the case of the Laplacian $\mathbf{L} = \Delta = \sum_{m=1}^{d} \frac{\partial^2}{\partial x_m^2}$ (scalar operator), we see that $\rho(\tau) = \tau^{-2}$ in (12).

Going back to (6), it may seem that Φ can be arbitrarily chosen. This is true with respect to translation and rotation invariance of $\Psi\{f\}$ since (10) and (11) ensure that without the need for specifying an explicit functional form for Φ. However, invariance of $\Psi\{f\}$ to dilation calls for special attention as it couples the scale invariance of $\mathbf{L}$ and the effect of dilation on the potential function Φ. In fact, this connection together with (9) narrows down the choice of Φ as shown below.

Theorem 1: Let $\mathbf{L}$ be a linear, scale-, rotation-, and shift-invariant differential operator and the potential function Φ be as defined in Definition 1. Then, $\Psi\{f\}$ is invariant to scaling of the coordinates if and only if $\Phi(x) = \gamma\, x^p$, $\forall\ x \geq 0$, where $p > 0$ and γ is an arbitrary constant.

Proof: Writing down (9) explicitly in terms of integrals yields

$$\int_{\mathbb{R}^d} \Phi\left(\left\|\mathbf{L}\left\{f\left(\frac{\cdot}{\tau}\right)\right\}(\mathbf{x})\right\|\right) \mathrm{d}\mathbf{x} = \zeta(\tau) \int_{\mathbb{R}^d} \Phi(\|\mathbf{L}\{f\}(\mathbf{x})\|) \mathrm{d}\mathbf{x}. \tag{13}$$

We start from the l.h.s. of (13) and use the fact that $\mathbf{L}$ is scale-invariant (12) to obtain

$$\begin{aligned}\int_{\mathbb{R}^d} \Phi\left(\left\|\mathbf{L}\left\{f\left(\frac{\cdot}{\tau}\right)\right\}(\mathbf{x})\right\|\right) \mathrm{d}\mathbf{x} &= \int_{\mathbb{R}^d} \Phi\left(\rho(\tau)\ \left\|\mathbf{L}\left\{f\right\}\left(\frac{\mathbf{x}}{\tau}\right)\right\|\right) \mathrm{d}\mathbf{x} \\ &= \tau^d \int_{\mathbb{R}^d} \Phi(\rho(\tau)\,\|\mathbf{L}\{f\}(\mathbf{x})\|)\,\mathrm{d}\mathbf{x}. \end{aligned} \tag{14}$$

Then, comparing the r.h.s. of (13) and (14), we infer that Φ must necessarily satisfy

$$\Phi(\rho(\tau)\,x) = \vartheta(\tau)\,\Phi(x) \quad \forall\ x \in \mathbb{R}, \tag{15}$$

where $\vartheta(\tau) = \tau^{-d}\,\zeta(\tau)$. Differentiating (15) with respect to τ and setting $\tau = 1$, we get

$$x\,\Phi'(x) = p\,\Phi(x), \tag{16}$$

where we have used the fact that $\rho(1) = 1$ (there is no scaling for $\tau = 1$ in (12)) and $p = \frac{\vartheta'(1)}{\rho'(1)}$ is a real number. The general solution to (16) is of the form[2] $\Phi(x) = \gamma\, x_+^p + \kappa\,(-x)_+^p$, where γ and κ are arbitrary constants, and

$$x_+^p = \begin{cases} x^p, & \text{if } x \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$

Using the hypothesis that Φ is defined only for $x \geq 0$ and is non-negative and strictly increasing, we see from (16) that $p > 0$, which leads to the desired result: $\Phi(x) = \gamma\, x^p$, $\forall\ x \geq 0$ with $p > 0$. Conversely, it is verified that $\Phi(x) = \gamma\, x^p$ ensures scale invariance of $\Psi\{f\}$. ■

As a direct consequence of Theorem 1, we see that the following L_p-norm is the only choice of regularization with respect to (9): Ignoring the multiplicative constant γ and substituting $\Phi(x) = x^p$ in (6), we get

$$\Psi\{f\} = \int_{\mathbb{R}^d} ||\mathbf{L}\{f\}(\mathbf{x})||^p\,\mathrm{d}\mathbf{x}. \tag{17}$$

In this work, we shall focus on the convex class of regularization functionals in (17) which precludes $p < 1$. Therefore, the practical range of interest of the p-values is $1 \leq p \leq 2$. Some popular instances of convex $\Psi\{f\}$ in (17) that can be found in regularization or spline literature are

1) Total-variation regularization [39]–[41] where $p = 1$, $\mathbf{L} = \boldsymbol{\nabla}$, and $\zeta(\tau) = \tau^{d-1}$,
2) Quadratic regularization: Set $p = 2$,
   a) Laplacian semi-norm [42] with $\mathbf{L} = \Delta$ and $\zeta(\tau) = \tau^{d-4}$,
   b) Duchon's semi-norm of order M [43] where $\mathbf{L}$ is a vector composed of every possible M-th order partial derivative operator and $\zeta(\tau) = \tau^{d-2M}$.

[2]The general solution may contain distributions for negative integer values of p [38]. However, in the present context, we would like the solution to be a true function of x, which leads to the given form for Φ.

TABLE I
CHARACTERISTICS OF DIFFERENT PENALIZED-LIKELIHOOD INTERPOLATION SCHEMES

Scheme	Cost	Likelihood	Regularization	Optimization	Solution	Type of Algorithm
GLQ (Section IV-A)	Quadratic	Gaussian	Quadratic	Analytical	Explicit, Closed-Form	Linear, One-Step (Algorithm 1)
GLNQ (Section IV-B2)	Non-quadratic	Gaussian	Non-quadratic	Numerical	Numerical	Non-Linear, Iterative (Algorithm 2)
PLQ, PLNQ (Section IV-B3)	Non-quadratic	Non-Gaussian (Poisson)	Quadratic, Non-quadratic	Numerical	Numerical	Non-Linear, Iterative (Algorithm 3)

B. Discretization of the Problem

As we are dealing with an interpolation problem, we seek a solution of the form

$$f_\lambda(\mathbf{x}) = \sum_{\mathbf{k}\in\mathbb{Z}^d} c_\lambda[\mathbf{k}]\,\varphi(\mathbf{x}-\mathbf{k}) \tag{18}$$

for our regularized scheme. Then, the original problem (5) can be posed as the discrete-domain optimization problem

$$c_\lambda = \arg\min_c J_\lambda\{g,c\}, \tag{19}$$

with

$$\begin{aligned} J_\lambda\{g,c\} &= E_\lambda\{g,f\} \\ &= \mathcal{L}\{g,(c*b)\} \\ &\quad + \lambda\int_{\mathbb{R}^d}\left(\sum_{m=1}^{s}\left(\sum_{\mathbf{k}\in\mathbb{Z}^d} c[\mathbf{k}]\mathrm{L}_m\{\varphi\}(\mathbf{x}-\mathbf{k})\right)^2\right)^{\frac{p}{2}}\mathrm{d}\mathbf{x}, \end{aligned} \tag{20}$$

where f is now given by (18) so that

$$f(\mathbf{k}) = \sum_{\mathbf{m}\in\mathbb{Z}^d} c[\mathbf{m}]\,b[\mathbf{k}-\mathbf{m}] = (c*b)[\mathbf{k}]$$

with $b[\mathbf{k}] = \varphi(\mathbf{k})|_{\mathbf{k}\in\mathbb{Z}^d}$ in the data-fidelity term. Thus, while we consider the same continuous-domain model in (2) and (18), standard and regularized schemes differ in the way the coefficients are obtained. However, when $\lambda = 0$ in (20), the regularized scheme (19) reduces to the standard case (2) since we only minimize $\mathcal{L}\{g,(c*b)\}$ which leads to close-fitting of data; at the other extreme, as $\lambda\to\infty$, (19) results in a maximum-likelihood estimate within the null-space of L_m, $m = 1,\ldots,s$.

1) Quadratic Regularization (QR): When $p = 2$, the integrand in the r.h.s. of (20) is a quadratic term. Up to technical details related to φ and L_m, $m = 1,\ldots,s$, [27]–[29], the second term in the r.h.s. of (20) can be explicitly evaluated by interchanging the summation (with respect to $\mathbf{k}$) and the integral to yield

$$\begin{aligned} \sum_{m=1}^{s}\int_{\mathbb{R}^d}\left(\sum_{\mathbf{k}\in\mathbb{Z}^d} c[\mathbf{k}]\,\mathrm{L}_m\{\varphi\}(\mathbf{x}-\mathbf{k})\right)^2\mathrm{d}\mathbf{x} &= \sum_{\mathbf{k}\in\mathbb{Z}^d}(c*r)[\mathbf{k}]\,c[\mathbf{k}] \\ &= \langle c*r, c\rangle_{\ell_2} \\ &= \Psi_{\mathrm{Q}}\{c\}, \end{aligned}$$

where $\langle\cdot,\cdot\rangle_{\ell_2}$ represents the ℓ_2 inner-product of two discrete sequences and where r is the discrete sequence whose $\mathbf{k}$-th component is given by

$$\begin{aligned} r[\mathbf{k}] &= \sum_{m=1}^{s} r_m[\mathbf{k}] \\ &= \sum_{m=1}^{s}\int_{\mathbb{R}^d}\mathrm{L}_m\{\varphi\}(\mathbf{x})\,\mathrm{L}_m\{\varphi\}(\mathbf{x}-\mathbf{k})\,\mathrm{d}\mathbf{x}. \end{aligned} \tag{21}$$

Thus, in the quadratic case ($p = 2$), the discretization of $\Psi\{f\}$ implicitly follows from (21) and leads to a quadratic function $\Psi_{\mathrm{Q}}\{c\}$ of the coefficients.

2) Non-Quadratic Regularization (NQR): In this case, the problem (19) can be handled only in a numerical optimization framework. For the purpose of numerical tractability, we replace the integral in (20) by a Riemann sum which leads to the discrete non-quadratic regularization

$$\Psi_{\mathrm{NQ}}\{c\} = \sum_{\mathbf{k}\in\mathbb{Z}^d}\left(\sum_{m=1}^{s}\left((c*\eta_m)[\mathbf{k}]\right)^2\right)^{\frac{p}{2}}, \tag{22}$$

where $\eta_m[\mathbf{k}] = \mathrm{L}_m\{\varphi\}(\mathbf{k})$ represents the discretized version of the differential operator L_m. The use of $\Psi_{\mathrm{NQ}}\{c\}$ for regularized interpolation distinguishes this work from those in the literature [24]–[29] which primarily deal with quadratic regularization. An important characteristic of Ψ_{NQ} is that the discretized derivatives η_m are obtained by sampling the corresponding continuous-domain derivatives $\mathrm{L}_m\{\varphi\}$, $m = 1,2,\ldots,s$. We illustrate this connection in Section V where we present η_m, $m = 1,\ldots,s$, for the case of the gradient operator and polynomial B-splines.

IV. OPTIMIZATION ALGORITHMS

Setting $p = 2$ under the Gaussian likelihood model leads to a J_λ that is quadratic in c in (20). The corresponding optimization can be performed analytically and leads to an explicit closed-form solution that is related to g in a linear fashion. However, J_λ becomes non-quadratic as soon as $p\neq 2$ in the regularization or when the data-fidelity is dictated by a non-Gaussian likelihood (irrespective of whether $p = 2$ or not). For non-quadratic J_λ, the optimization has to be performed numerically and the corresponding solution depends on g in a non-linear fashion. This discussion is summarized in Table I where we present the characteristics of different regularized interpolation schemes. We first describe the optimization of

quadratic cost (first row of Table I). Then, we give a detailed exposition of the minimization procedure that we adopt for some specific instances of non-quadratic costs involving Gaussian and Poisson likelihoods (second and third rows emphasized by bold-face font in Table I).

A. Quadratic Cost: Gaussian Likelihood with Quadratic Regularization (GLQ)

The cost to be minimized is composed of $\mathcal{L}_{\mathrm{Q}}$ and Ψ_{Q}, i.e.,

$$J_{\lambda,\mathrm{GLQ}}\{g,c\} = \sum_{\mathbf{k}\in\mathbb{Z}^d} (g[\mathbf{k}] - (c*b)[\mathbf{k}])^2 + \lambda \langle c*r, c\rangle_{\ell_2}.$$

As $J_{\lambda,\mathrm{GLQ}}$ is quadratic, setting its derivative with respect to $c[\mathbf{k}]$ to zero yields the system of linear equations $\forall\ \mathbf{k} \in \mathbb{Z}^d,\ (c*b*\bar{b}+\lambda\, r*c)[\mathbf{k}] = (g*\bar{b})[\mathbf{k}]$, where $\bar{b}[\mathbf{k}] = b[-\mathbf{k}]$. This system can be solved in the discrete-space Fourier domain to obtain

$$\begin{aligned} C_\lambda(\boldsymbol{z})|_{\boldsymbol{z}=e^{j\boldsymbol{\omega}}} &= H_\lambda(e^{j\boldsymbol{\omega}})\, G(e^{j\boldsymbol{\omega}}) \\ &= \left(\frac{B^*(e^{j\boldsymbol{\omega}})}{|B(e^{j\boldsymbol{\omega}})|^2 + \lambda\, R(e^{j\boldsymbol{\omega}})}\right) G(e^{j\boldsymbol{\omega}}), \quad (23) \end{aligned}$$

where C_λ, G, B, and R are the z-transforms of the solution c_λ, the data g, and the sequences b and r, respectively. Substituting for c_λ in (18) from (23) results in a smoothing-spline-type solution that has already been investigated in [26]–[29] where conditions necessary for the existence and stability of (23) are derived in a rigorous fashion.

Equation (23) amounts to writing $c_\lambda = h_\lambda * g$, where h_λ is the digital-correction filter whose z-transform is H_λ. The solution is therefore linear, in the sense that f_λ is related to g in a linear fashion. The implementation of (23) is straightforward and can be done via FFTs (see Algorithm 1 in Appendix A-A). We also note that $H_\lambda(e^{j\boldsymbol{\omega}}) \to \frac{1}{B(e^{j\boldsymbol{\omega}})}$ for $\lambda \to 0$, in which case f_λ performs the standard interpolation of g [20]. However, while B can be separable, this is never the case for H_λ when $\lambda > 0$.

B. Non-Quadratic Costs

When J_λ is a non-quadratic, non-convex function of the coefficients, conventional techniques such as the non-linear conjugate gradient and BFGS methods [44] may be used for determining its local minima. However, when J_λ is convex, several methodologies such as the majorize-minimize (MM) approach [30] (or, equivalently, bounded-optimization) and the half-quadratic method [45]–[47] can be adopted for developing elegant minimization algorithms that ensure a steady decrease of J_λ. Moreover, it is well known that the minimization of a convex J_λ always leads to a solution (which need not necessarily be unique) for which J_λ achieves its global-minimum value.

In the present work, we resort to the MM approach [30] as it is easy to comprehend: The idea, as described in [30], is to replace the original difficult task by several easy-to-optimize problems that will guarantee a monotonic decrease of the original cost. We briefly review from [30] the mathematical details underlying the MM philosophy in Section IV-B1 and then apply it to the following two instances of convex non-quadratic cost:[3]

1) $J_{\lambda,\mathrm{GLNQ}}$ that combines Ψ_{NQ} with the Gaussian likelihood $\mathcal{L}_{\mathrm{Q}}$ (second row of Table I):

$$J_{\lambda,\mathrm{GLNQ}}\{g,c\} = \sum_{\mathbf{k}\in\mathbb{Z}^d} (g[\mathbf{k}] - (c*b)[\mathbf{k}])^2 + \lambda\, \Psi_{\mathrm{NQ}}\{c\},$$

2) $J_{\lambda,\mathrm{PLNQ}}$ that combines Ψ_{NQ} (or $J_{\lambda,\mathrm{PLQ}}$ that uses Ψ_{Q}) with the Poisson likelihood $\mathcal{L}_{\mathrm{Poisson}}$:

$$\begin{aligned} J_{\lambda,\mathrm{PLNQ}}\{g,c\} = &- \sum_{\mathbf{k}\in\mathbb{Z}^d} g[\mathbf{k}] \log((c*b)[\mathbf{k}]) \\ &+ \sum_{\mathbf{k}\in\mathbb{Z}^d} (c*b)[\mathbf{k}] + \lambda\, \Psi_{\mathrm{NQ}}\{c\}, \end{aligned}$$

where we have considered $\mathcal{L}_{\mathrm{Poisson}}$ as a prototypical example of a non-Gaussian likelihood model (third row of Table I) that yields a convex non-quadratic data term.

1) MM Philosophy: In the MM setting, we construct an auxiliary cost function $J_{\mathrm{AUX}}\{g,c|c_t\}$ at the current estimate c_t such that

$$J_\lambda\{g,c_t\} = J_{\mathrm{AUX}}\{g,c_t|c_t\}, \quad (24)$$

$$J_\lambda\{g,c\} < J_{\mathrm{AUX}}\{g,c|c_t\},\ c \neq c_t. \quad (25)$$

Then, the idea is to find the estimate c_{t+1} at the next iteration such that

$$J_{\mathrm{AUX}}\{g,c_{t+1}|c_t\} < J_{\mathrm{AUX}}\{g,c_t|c_t\}, \quad (26)$$

which leads to a definite decrease of J_λ as shown below:

$$\begin{aligned} J_\lambda\{g,c_{t+1}\} &= J_{\mathrm{AUX}}\{g,c_{t+1}|c_t\} \\ &\quad + \underbrace{J_\lambda\{g,c_{t+1}\} - J_{\mathrm{AUX}}\{g,c_{t+1}|c_t\}}_{<0 \text{ using (25)}} \\ &< \underbrace{J_{\mathrm{AUX}}\{g,c_t|c_t\}}_{\text{using (26)}} = J_\lambda\{g,c_t\}. \end{aligned}$$

The whole philosophy relies on finding a suitable $J_{\mathrm{AUX}}\{g,c|c_t\}$ that satisfies (24) and (25) and which is also easy to minimize. In what follows, we make explicit the steps necessary for finding $J_{\mathrm{AUX}}\{g,c|c_t\}$ for the two instances $J_{\lambda,\mathrm{GLNQ}}$ and $J_{\lambda,\mathrm{PLNQ}}$ (or $J_{\lambda,\mathrm{PLQ}}$).

2) Gaussian Likelihood with Non-Quadratic Regularization (GLNQ): In this case, we need to apply the MM principle only on Ψ_{NQ} since the data term in $J_{\lambda,\mathrm{GLNQ}}$ is already quadratic. Our construction of the auxiliary cost J_{AUX} is based on the inequality

$$|x|^p \leq \frac{2-p}{2}|x_0|^p + \frac{p}{2}x^2|x_0|^{p-2}, \quad 1 \leq p < 2 \quad (27)$$

for some $x_0 \neq 0$ [30], where the equality holds only when $p = 2$ or $x = x_0$. The inequality is preserved under summation and multiplication by $\lambda > 0$ which leads to

$$\begin{aligned} \lambda\, \Psi_{\mathrm{NQ}}\{c\} \leq \lambda \frac{p}{2} \sum_{\mathbf{k}\in\mathbb{Z}^d} |\chi_t[\mathbf{k}]|^{p-2} \sum_{m=1}^{s} ((c*\eta_m)[\mathbf{k}])^2 \\ + \mathcal{A}_{\chi_t}, \quad (28) \end{aligned}$$

[3]For the p-values considered, Ψ_{Q} and Ψ_{NQ} are convex as they are derived from (17). Therefore, $J_{\lambda,\mathrm{GLNQ}}$ and $J_{\lambda,\mathrm{PLNQ}}$ (or $J_{\lambda,\mathrm{PLQ}}$) are convex as well because of the convexity of associated data-fidelity terms.

where $\chi_t[\mathbf{k}] = \sqrt{\sum_{m=1}^{s}\left((c_t * \eta_m)[\mathbf{k}]\right)^2}$ and $\mathcal{A}_{\chi_t} = \lambda\left(\frac{2-p}{2}\right)\sum_{\mathbf{k}\in\mathbb{Z}^d}|\chi_t[\mathbf{k}]|^p$ is a constant independent of c. Thus, we obtain

$$\begin{aligned} J_{\text{AUX}}\{g, c|c_t\} = &\sum_{\mathbf{k}\in\mathbb{Z}^d}(g[\mathbf{k}] - (c*b)[\mathbf{k}])^2 \\ &+ \lambda\frac{p}{2}\sum_{\mathbf{k}\in\mathbb{Z}^d}|\chi_t[\mathbf{k}]|^{p-2}\sum_{m=1}^{s}\left((c*\eta_m)[\mathbf{k}]\right)^2 \\ &+ \mathcal{A}_{\chi_t}, \end{aligned} \tag{29}$$

which satisfies (24) and (25). The details associated with the minimization of the above J_{AUX} and the algorithm (Algorithm 2) that results therefrom are provided in Appendix A-B.

3) Poisson Likelihood with Non-Quadratic Regularization (PLNQ): Since $J_{\lambda,\text{PLNQ}}$ contains a term of the form $\log((c*b)[\mathbf{k}])$, it is imperative that $(c*b)[\mathbf{k}] > 0$, $\forall\ \mathbf{k}\in\mathbb{Z}^d$. Therefore, whenever we use the Poisson likelihood[4] $\mathcal{L}_{\text{Poisson}}$, we restrict ourselves to basis functions that satisfy $b[\mathbf{k}] = \varphi(\mathbf{k}) \geq 0$ $\forall\ \mathbf{k}\in\mathbb{Z}^d$ (*e.g.*, integer-degree polynomial B-splines). Then, the above constraint simply boils down to ensuring the positivity of c.

To construct the auxiliary cost at the current estimate c_t, we use the convexity of $\log\left(\frac{1}{\cdot}\right)$ [30, Equation (10)] to write

$$\begin{aligned} &-\sum_{\mathbf{k}\in\mathbb{Z}^d} g[\mathbf{k}]\log((c*b)[\mathbf{k}]) \\ &\quad\leq -\sum_{\mathbf{k},\mathbf{q}\in\mathbb{Z}^d} g[\mathbf{k}]\frac{b[\mathbf{k}-\mathbf{q}]c_t[\mathbf{q}]}{(c_t*b)[\mathbf{k}]}\log\left((c_t*b)[\mathbf{k}]\frac{c[\mathbf{q}]}{c_t[\mathbf{q}]}\right). \end{aligned}$$

Similarly, Equation (27) and the convexity of $(\cdot)^2$ [30, Equation (9)] result in

$$\begin{aligned} &\lambda\,\Psi_{\text{NQ}}\{c\} \\ &\leq\ \lambda\frac{p}{2}\sum_{\substack{\mathbf{k},\mathbf{q}\in\mathbb{Z}^d \\ m=1\cdots s}}\left(\frac{\eta_m[\mathbf{k}-\mathbf{q}]}{\alpha[\mathbf{k}-\mathbf{q}]}(c[\mathbf{q}]-c_t[\mathbf{q}]) + (c_t*\eta_m)[\mathbf{k}]\right)^2 \\ &\qquad\qquad\times\ |\chi_t[\mathbf{k}]|^{p-2}\alpha[\mathbf{k}-\mathbf{q}] \\ &\quad + \mathcal{A}_{\chi_t}, \end{aligned}$$

where αs are constants such that $\alpha[\mathbf{k}] = 0$ if $\eta_m[\mathbf{k}] = 0$, $\alpha[\mathbf{k}] > 0$ if $\eta_m[\mathbf{k}] \neq 0$, and $\sum_{\mathbf{k}\in\mathbb{Z}^d}\alpha[\mathbf{k}] = 1$. After some algebraic manipulations, we obtain

$$J_{\lambda,\text{PLNQ}}\{g, c\} \quad\leq\quad \sum_{\mathbf{q}\in\mathbb{Z}^d} J_{\text{AUX}}\{g, c|c_t, \mathbf{q}\}, \tag{30}$$

where $J_{\text{AUX}}\{g, c|c_t, \mathbf{q}\}$ is the decoupled auxiliary cost that depends only on $c[\mathbf{q}]$:

$$\begin{aligned} &J_{\text{AUX}}\{g, c|c_t, \mathbf{q}\} \\ &\quad= -c_t[\mathbf{q}]\,\mathcal{A}_{1\mathbf{q}}\log(c[\mathbf{q}]) \\ &\qquad + p\lambda\,c[\mathbf{q}]\left(\frac{c[\mathbf{q}]\,\mathcal{A}_{2\mathbf{q}}}{2} - c_t[\mathbf{q}]\,\mathcal{A}_{2\mathbf{q}} + \mathcal{A}_{3\mathbf{q}} + \frac{\mathcal{A}_b}{p\lambda}\right) \\ &\qquad + \mathcal{A}_{4\mathbf{q}}. \end{aligned} \tag{31}$$

[4]The development for $J_{\lambda,\text{PLQ}}$ (Ψ_{Q} with $\mathcal{L}_{\text{Poisson}}$) follows from that of $J_{\lambda,\text{PLNQ}}$ as we shall see in Appendix A-C.

TABLE II
z-TRANSFORMS B AND $B^{(1)}$ CORRESPONDING TO VARIOUS SPLINE DEGREES

n	$B(z)$	$B^{(1)}(z)$
1	1	$\frac{1}{2}(z^{-1}-z)$
2	$\frac{3}{4}+\frac{1}{8}(z^{-1}+z)$	$\frac{1}{2}(z^{-1}-z)$
3	$\frac{4}{6}+\frac{1}{6}(z^{-1}+z)$	$\frac{1}{2}(z^{-1}-z)$
4	$\frac{115}{192}+\frac{19}{96}(z^{-1}+z)$ $+\frac{1}{384}(z^{-2}+z^2)$	$\frac{11}{24}(z^{-1}-z)$ $+\frac{1}{48}(z^{-2}-z^2)$
5	$\frac{11}{20}+\frac{13}{60}(z^{-1}+z)$ $+\frac{1}{120}(z^{-2}+z^2)$	$\frac{5}{6}(z^{-1}-z)$ $+\frac{1}{24}(z^{-2}-z^2)$

The quantity $\mathcal{A}_b = \sum_{\mathbf{k}\in\mathbb{Z}^d} b[\mathbf{k}] > 0$ is a constant independent of c_t and

$$\mathcal{A}_{1\mathbf{q}} = (\bar{b} * w_t)[\mathbf{q}], \tag{32}$$

$$\mathcal{A}_{2\mathbf{q}} = \sum_{m=1}^{s}(\bar{\mu}_m * |\chi_t|^{p-2})[\mathbf{q}], \tag{33}$$

$$\mathcal{A}_{3\mathbf{q}} = \sum_{m=1}^{s}\sum_{\mathbf{k}\in\mathbb{Z}^d}|\chi_t|^{p-2}[\mathbf{k}]\,(c_t*\eta_m)[\mathbf{k}]\,\eta_m[\mathbf{q}-\mathbf{k}], \tag{34}$$

where $w_t[\mathbf{k}] = \frac{g[\mathbf{k}]}{(c_t*b)[\mathbf{k}]}$, $\mu_m[\mathbf{k}] = \frac{\eta_m^2[\mathbf{k}]}{\alpha[\mathbf{k}]}$, and $\mathcal{A}_{4\mathbf{q}}$ in (31) is an additive constant which is irrelevant for the minimization of $J_{\text{AUX}}\{g, c|c_t, \mathbf{q}\}$. The corresponding optimization procedure and the algorithm (Algorithm 3) are described in Appendix A-C.

V. Spline-Based Regularized Interpolation

Here, we make explicit the link between the sequences η_m, r_m, $m = 1, \ldots, s$, and φ and $\mathbf{L}$ for spline-based interpolation which is well-suited for imaging problems [20], [31]. We consider a separable basis given by

$$\varphi(\mathbf{x}) = \varphi(x_1)\,\varphi(x_2)\cdots\varphi(x_d), \tag{35}$$

where $\varphi(x)$ is the univariate basis function. For $\mathbf{L}$, we select the gradient operator $\boldsymbol{\nabla}$. Then, the proposed regularization is related to the total-variation functional (for $p = 1$) which is of particular interest to us because of its edge-preserving characteristics.

The separable nature of φ in (35) leads to the separability of the corresponding discrete sequences b, η_m, and r_m, i.e., $b[\mathbf{k}] = b[k_1]\,b[k_2]\cdots b[k_d]$, where $b[k] = \varphi(x)|_{x=k}$. For $\mathbf{L} = \boldsymbol{\nabla}$, we have $s = d$, which yields for $m = 1, \ldots, d$,

$$\eta_m[\mathbf{k}] = \prod_{\substack{l=1 \\ l\neq m}}^{d} b[k_l]\ b^{(1)}[k_m],$$

where $b^{(1)}[k] = \varphi'(x)|_{x=k}$ and

$$r_m[\mathbf{k}] = \prod_{l=1 l\neq m}^{d} a[k_l]\ a^{(2)}[k_m],$$

TABLE III
z-TRANSFORM $A^{(2)}$ CORRESPONDING TO VARIOUS SPLINE DEGREES

n	$A^{(2)}(z)$
1	$2-(z^{-1}+z)$
2	$1-\frac{1}{3}(z^{-1}+z)-\frac{1}{6}(z^{-2}+z^{2})$
3	$\frac{2}{3}-\frac{1}{8}(z^{-1}+z)-\frac{1}{5}(z^{-2}+z^{2})-\frac{1}{120}(z^{-3}+z^{3})$
4	$\frac{35}{72}-\frac{11}{360}(z^{-1}+z)-\frac{17}{90}(z^{-2}+z^{2})-\frac{59}{2520}(z^{-3}+z^{3})-\frac{1}{5040}(z^{-4}+z^{4})$
5	$\frac{809}{2160}+\frac{1}{64}(z^{-1}+z)-\frac{31}{1890}(z^{-2}+z^{2})-\frac{907}{24192}(z^{-3}+z^{3})-\frac{25}{18144}(z^{-4}+z^{4})-\frac{1}{362880}(z^{-5}+z^{5})$

where $a[k]=(\bar{\varphi}*\varphi)(x)|_{x=k}$, $\bar{\varphi}(x)=\varphi(-x)$, and $a^{(2)}[k]=(\bar{\varphi}'*\varphi')(x)|_{x=k}$.

We list out the sequences b, $b^{(1)}$, and $a^{(2)}$ (in terms of their transfer functions B, $B^{(1)}$, and $A^{(2)}$, respectively) in Tables II and III, for $\varphi=\beta^n$, which is the symmetric polynomial B-spline of degree $n\geq 1$. We see that, for splines of sufficiently high degree, $b^{(1)}$ and $a^{(2)}$ are very different from the simple finite-difference filters (*e.g.*, first row of both Tables II and III) that are typically used in the literature for TV-based image restoration.

VI. EXPERIMENTS

We validate the proposed regularized interpolation scheme by carrying out 2-D rotation experiments in the presence of noise. We adopt the separable polynomial B-spline model in Section V for all implementations and use $\mathbf{L}=\boldsymbol{\nabla}$. Standard interpolation is performed as described in [20]. In the regularized case, we implement quadratic regularized (QR) interpolation—Algorithm 1 corresponding to GLQ in Table I—using FFTs. The various steps of non-quadratic-regularized (NQR) interpolation—Algorithms 2 and 3 corresponding to GLNQ and PLNQ, respectively, in Table I—are executed via discrete convolutions. Periodic boundary conditions were applied whenever necessary. In all experiments, we set $p=1$ in (22) for NQR interpolation. The stopping criterion for Algorithms 2 and 3 is $\|c_{t+1}-c_t\|^2<10^{-4}\frac{\|g\|}{N}$, where N is the size of g. The performance of all methods is quantified by

$$\text{SNR}(\lambda)=10\log_{10}\left(\frac{\sum_{\mathbf{k}\in\Omega}f_{\text{true}}^2(\mathbf{k})}{\sum_{\mathbf{k}\in\Omega}(f_{\text{true}}(\mathbf{k})-f_\lambda(\mathbf{k}))^2}\right), \quad (36)$$

which is a function of the regularization parameter λ and where $\{f_{\text{true}}(\mathbf{k})\}_{\mathbf{k}\in\mathbb{Z}^d}$ and $\{f_\lambda(\mathbf{k})\}_{\mathbf{k}\in\mathbb{Z}^d}$ are the values of the continuous-domain noise-free signal and regularized interpolant, respectively, sampled on the grid of data. The SNR is evaluated inside a circular region Ω concentric with the image so as to avoid boundary effects. Our main aim in this section is to characterize the best-possible performance of the proposed regularized interpolation methods. For this purpose, we conduct oracle-based experiments, i.e., we set λ so as to obtain f_λ that yields the highest SNR for a given realization of the noisy data g. In Section VII-D, we introduce a data-driven scheme for obtaining MSE-optimal λ directly from g.

A. *Regularized* versus *Standard Interpolation—Rotation Experiment*

We compare the proposed regularized scheme against standard interpolation by carrying out two sets of 2-D rotation experiments where the setup is exactly similar to that described in Section II-A. For the first experiment, we use a 512×512 image of a CT slice. For the second, we consider a stack of clean MRI images [48] where we use different slices picked randomly from the MRI stack for different realizations. We perform exact-fitting interpolation (non-regularized) using nearest-neighbour, linear B-spline, cubic B-spline and cubic Keys basis [20], [49], while, for regularized methods, we consider quadratic regularized (QR) cubic B-spline (Algorithm 1) and non-quadratic regularized (NQR) cubic B-spline (Algorithm 2), respectively.

1) SNR Comparison: We plot the output SNR for each of these methods in Figure 2a and 2b corresponding to the CT slice and the MRI stack, respectively. In both figures, we observe that the non-regularized methods exhibit the same trend as depicted in Figure 1. The observation that is most relevant to us is that both regularized interpolation methods (QR and NQR) perform far better than the non-regularized ones at high noise levels. The consistently superior trend exhibited by regularized interpolation in both figures clearly illustrate its robustness against noise. Finally, the output SNR of the two regularized methods (for cubic B-spline) converges to that of the non-regularized cubic B-spline for relatively high input SNRs since the effect of regularization becomes negligible under very low noise.

2) Visual Comparison: We present in Figure 3 output images for one realization of the experiment in Section VI-A (with MRI stack). Since standard cubic B-spline interpolation is not regularized, the corresponding output (Figure 3c) is still noisy. On the contrary, regularized methods lead to significant noise reduction as seen in Figures 3d and 3e corresponding to QR-cubic and NQR-cubic outputs, respectively.

B. *Regularized* versus *Standard Interpolation—Image Zooming*

We also compare standard and regularized (QR—Algorithm 1 and NQR—Algorithm 2) cubic B-spline-based interpolation in a scaling experiment: Here, the objective is to zoom into a noisy MRI image by an irrational factor $\tau=\pi$. For the proposed regularized interpolation schemes, we select $\lambda=\zeta(\pi)\,\lambda_{\text{opt}}$, where $\zeta(\cdot)$ is the scaling constant in (9) and λ_{opt} is that value which maximizes the SNR (36)

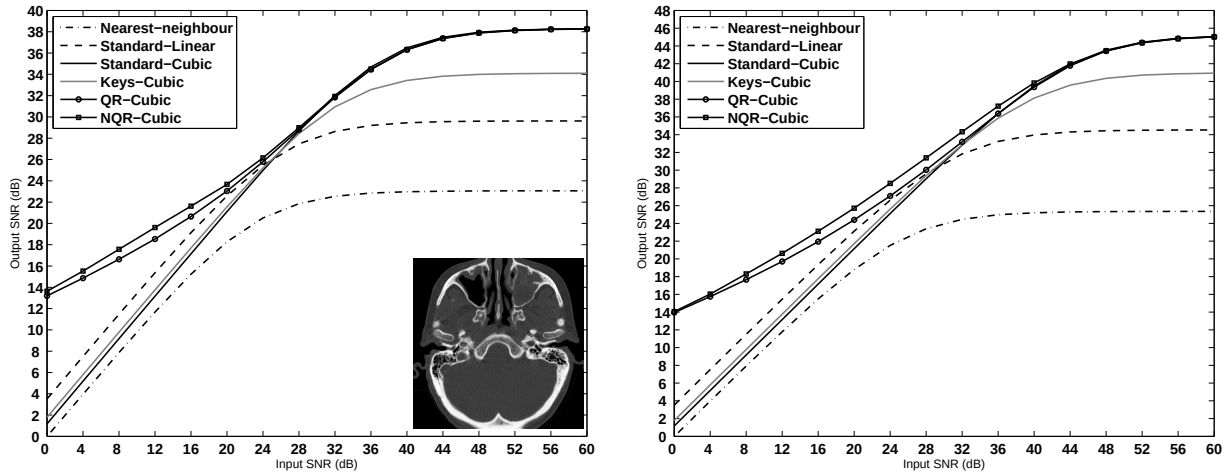


Fig. 2. Comparison of performance of non-regularized and regularized interpolation: Rotation experiments on noisy versions of (a) slice of a CT image and (b) MRI slices.

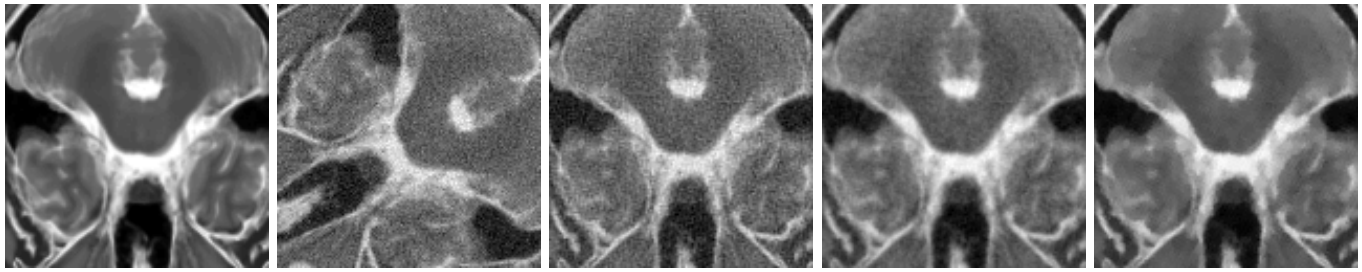


Fig. 3. Visual Comparison for the rotation experiment; a cubic B-spline interpolation model was used for all schemes: (a) Noise-free image; (b) Noisy data (rotated by -60°, $\sigma = 17.54$); (c) Non-regularized (standard) result (SNR = 17.12 dB); (d) Quadratic regularized (QR) result (Algorithm 1 with optimal λ, SNR = 19.49 dB); (e) Non-quadratic regularized (NQR) result (Algorithm 2 with optimal λ, SNR = 21.04 dB).

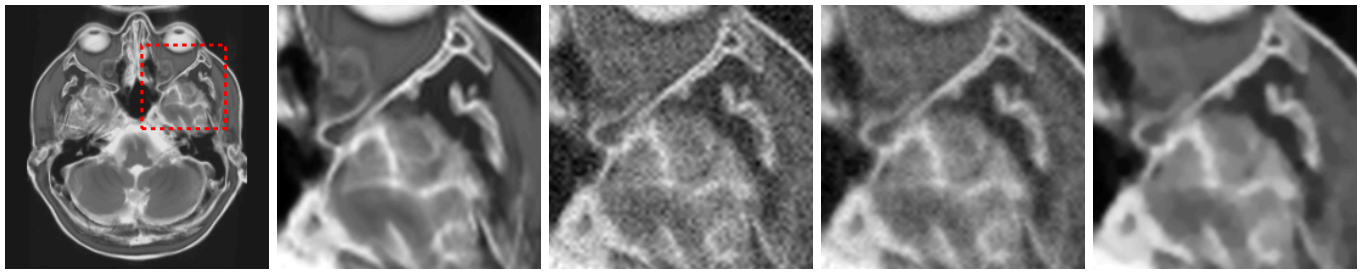


Fig. 4. Visual Comparison for the scaling experiment; a cubic B-spline interpolation model was used for all schemes: (a) Noise-free image indicating the portion that is zoomed; (b) Noise-free image zoomed-in by non-regularized (standard) method; (c) Noisy data zoomed-in by non-regularized method; (d) Noisy data zoomed-in by quadratic regularized (QR) method; (e) Noisy data zoomed-in by non-quadratic regularized (NQR) method.

for data given on the original (unzoomed) grid. The above choice of λ is justified by (5) and (9) which indicate that the regularization parameter must be suitably compensated whenever the continuous-domain output f_λ is dilated. This compensation is easily computed since Theorem 1 gives us an explicit expression for $\zeta(\cdot)$ for the chosen regularization operator $\mathbf{L} = \boldsymbol{\nabla}$. Substituting for $\mathbf{L}$ in (9) and (17) and manipulating the L_p-norm, it is seen that $\zeta_{\mathrm{QR}}(\tau) = 1$ and $\zeta_{\mathrm{NQR}}(\tau) = \tau$ for QR ($p = 2$) and NQR ($p = 1$) interpolation schemes, respectively.

We portray in Figure 4 the output images corresponding to non-regularized-cubic (Figure 4c), QR-cubic (Figure 4d) and NQR-cubic (Figure 4e) methods, respectively. It is clearly seen that both regularized methods exhibit superior performance compared to the standard scheme in terms of noise reduction. While they also seem to suppress some subtle components of the underlying noise-free image (which is an inevitable by-product of noise-filtering), the key point is that in performing noise reduction NQR (Figure 4e) preserves prominent image features, i.e., edges, better than QR (Figure 4d).

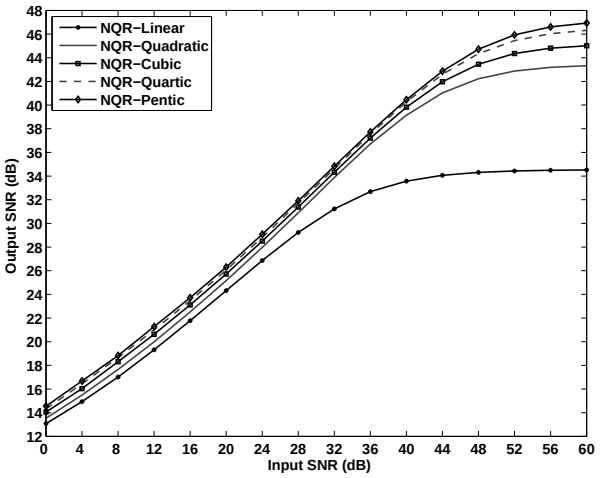


Fig. 5. Non-quadratic regularized (NQR) interpolation for varying spline-degree.

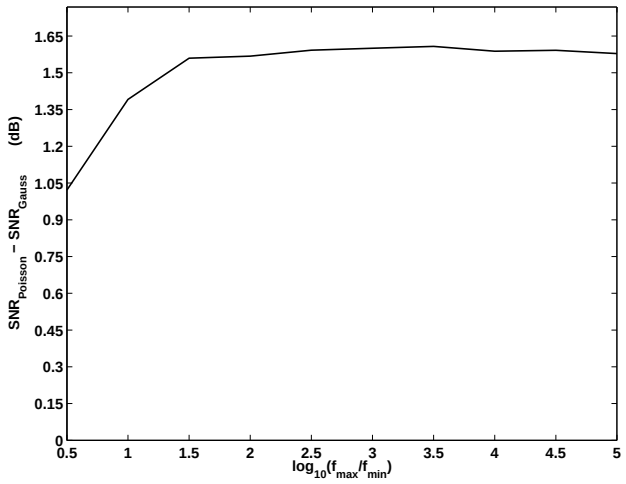


Fig. 6. Comparison of the effect of likelihood on non-quadratic regularized (NQR) interpolation: Plot of output SNR of $\mathcal{L}_{\mathrm{Poisson}}$-based method (PLNQ) relative to $\mathcal{L}_{\mathrm{Q}}$-based method (GLNQ) as a function of dynamic range $\log_{10}(\frac{f_{\max}}{f_{\min}})$. A positive value in the graph indicates that $\mathcal{L}_{\mathrm{Poisson}}$ provides better performance than $\mathcal{L}_{\mathrm{Q}}$ for data corrupted by Poisson noise.

C. Regularized Interpolation: Varying Spline Degree

To study the effect of the basis function φ on the discretization of the non-quadratic regularization (NQR) in (22), we repeat the experiment in Section VI-A (for MRI images) with B-splines of integer degree varying from 1 to 5 and concentrate on NQR interpolation (Algorithm 2). We show in Figure 5 the performance of the NQR interpolation based on linear ($n = 1$), quadratic ($n = 2$), cubic ($n = 3$), quartic ($n = 4$), and pentic ($n = 5$) B-splines, respectively.

In terms of quality, higher-degree B-splines yield better performance with NQR interpolation as the output SNR consistently increases with the degree of the B-spline over the entire range of input SNRs in Figure 5. Particularly, there is a notable improvement going from linear to a higher-degree B-spline. This is probably because, for the linear B-spline, the discretization does not adequately capture the features of the corresponding continuous-domain model, while the situation improves when $n \geq 2$. For higher-degree B-splines ($n \geq 4$), we only observe marginal increments in the output SNR that tends to saturate. This is to be expected since the cardinal splines (corresponding to these B-splines) rapidly converge to the sinc function with increasing n [31].

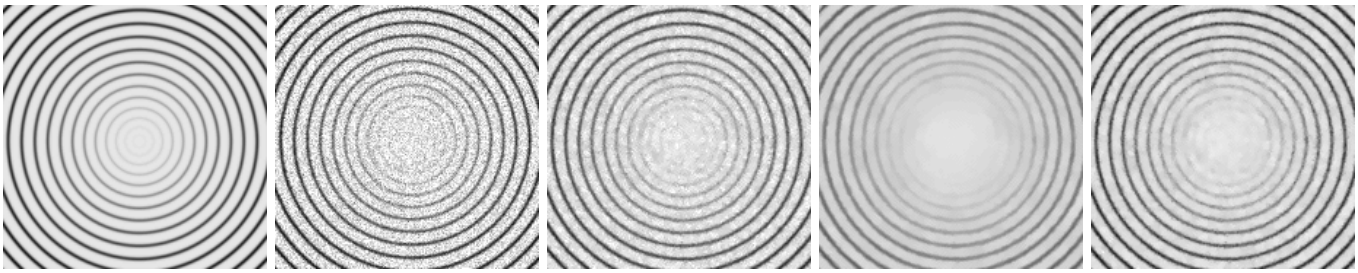
In terms of computation, since the number of filter-taps in b and $\{\eta_m\}$ (see Section V) increases linearly with n, the cost of implementing the convolutions in Algorithm 2 also increases linearly with n. Therefore, in practice, it is important to choose a B-spline (φ, in general) that strikes a good balance between computation load and quality. Based on our observations from Figures 2-5, we find cubic B-spline to provide satisfactory results.

D. Experiments with Poisson Noise

Until now, we have presented results that demonstrate the superior performance of regularized interpolation (using Algorithms 1 and 2) for data corrupted by additive Gaussian noise. In this section, we investigate the proposed approach using Algorithm 3 (in Appendix A-C) by performing rotations in the presence of signal-dependent Poisson noise. As the outcome was very similar to that exhibited in Figures 2a and 2b, in the interest of space we do not show the results here but summarize our findings by stating again that the proposed regularized interpolation scheme outperforms standard methods by a wide margin.

Our concern in this section is rather to investigate whether or not the choice of the data-fidelity term based on the likelihood model is crucial for regularized interpolation. Specifically, we propose to study the performance of quadratic data fidelity $\mathcal{L}_{\mathrm{Q}}$ (which corresponds to a Gaussian likelihood) when applied to signal-dependent Poisson noise and compare it with $\mathcal{L}_{\mathrm{Poisson}}$ which is statistically the most appropriate data-fidelity term for this type of noise. For this purpose, we repeat the rotation experiment described in Section VI-A with the noise-free phantom shown in Figure 7a, but we now consider data corrupted by signal-dependent Poisson noise. An important property of a Poisson random variable is that its variance is equal (or proportional, if there is a multiplicative gain factor) to its mean. Moreover, since the intensity value of the noise-free phantom at a given pixel characterizes the signal-dependent Poisson noise at that pixel (mean of the Poisson random variable), the overall variance of the Poisson noise is controlled by the mean of the phantom. Therefore, we keep the mean of the phantom constant but vary its dynamic range (ratio of the maximum to the minimum intensity value) which we believe provides a suitable handle to inspect how well the two likelihood models capture image details in a varied range of intensity levels for the same amount of input Poisson noise. We perform NQR-cubic interpolation using Algorithm 2 ($\mathcal{L}_{\mathrm{Q}}$ data-fidelity) and Algorithm 3 ($\mathcal{L}_{\mathrm{Poisson}}$ data-fidelity), respectively. We compute the output SNR of the latter relative to the former ($\mathrm{SNR}_{\mathrm{Poisson}} - \mathrm{SNR}_{\mathrm{Gauss}}$) and plot it in Figure 6 as a function of the dynamic range (in $\log_{10}$-scale).

For the entire extent of abscissa in Figure 6, we see that Algorithm 3 yields substantial improvement (≥ 1 dB) over Algorithm 2 pointing to the superiority of the Poisson likelihood



(a) (b) (c) (d) (e)

Fig. 7. Visual comparison of the effect of likelihood on non-quadratic regularized (NQR) interpolation: (a) Noise-free image, dynamic range of $10^{2.5}$; (b) Data corrupted by signal-dependent Poisson noise; (c) Result based on $\mathcal{L}_{\mathrm{Q}}$ data-fidelity (Algorithm 2 optimized for best output SNR); (d) Result based on $\mathcal{L}_{\mathrm{Q}}$ data-fidelity (Algorithm 2 with slight over-regularization); (e) Result based on $\mathcal{L}_{\mathrm{Poisson}}$ data-fidelity (Algorithm 3 optimized for best output SNR).

model in adapting to image details in a large range of intensity levels. This is also clearly illustrated in Figure 7 where we show the output images corresponding to one realization of this experiment: The output of Algorithm 2 (Figure 7c) retains the innermost circles, but is still noisy. Increasing the regularization strength reduces the noise significantly (Figure 7d) but at the cost of loosing the smaller circles. However, the output of Algorithm 3 (Figure 7e) is less noisy and preserves the smaller circles as well.

VII. Discussion

A. *Quadratic* versus *Non-Quadratic Regularization*

Among the regularized schemes investigated in this work, NQR interpolation (based on TV-like regularization) performs significantly better than QR interpolation both in terms of visual quality and SNR: In Figures 3 and 4, the NQR result (Figure 3e and 4e) is sharper and less noisy than the QR result (Figure 3d and 4d), while in Figures 2a and 2b we see that NQR brings about a consistent SNR improvement over QR interpolation for a range of input SNRs. We also observed a similar trend in SNR improvement of NQR over QR in a pure denoising scenario (where we repeated the experiments in Section VI-A without applying rotations). These results are a direct consequence of the fact that TV-like NQRs have a good ability to preserve edges while Tikhonov-like QRs tend to blur them thus compromising the quality. Therefore, from a performance point of view, it is better to employ TV-like NQR for regularized interpolation.

B. *Influence of the Likelihood Model*

We were able to present an example in Section VI-D to illustrate a case where the likelihood term plays a significant role in variational problems. There, the Poisson likelihood $\mathcal{L}_{\mathrm{Poisson}}$ outperforms the Gaussian likelihood $\mathcal{L}_{\mathrm{Q}}$ since $\mathcal{L}_{\mathrm{Poisson}}$ provides intensity-dependent regularization that adapts to the non-stationary nature of Poisson noise. This is unlike $\mathcal{L}_{\mathrm{Q}}$ which leads to a uniform regularization that either washes out low-intensity details (Figure 7d) or retains them at the cost of performing poorly in high-intensity regions (Figure 7c). However, it must be noted that we had to tailor the circular phantom in Figure 7a to demonstrate the advantage of using $\mathcal{L}_{\mathrm{Poisson}}$ for Poisson noise, by enforcing the presence of dominant peaks at high intensities in the histogram.

This is to be contrasted with the many experiments we performed with real-world autofluorescence images, where we observed that there was no significant difference in the performances of $\mathcal{L}_{\mathrm{Poisson}}$ and $\mathcal{L}_{\mathrm{Q}}$. Our understanding of this behavior is that, in those situations, the regularization had a stronger effect on the solution than the likelihood. Therefore, when the data-fidelity term is less important than the regularization, algorithms can be designed to reduce the amount of computations. For instance, since Algorithm 2 uses $\mathcal{L}_{\mathrm{Q}}$ for the data fidelity, it only requires the linearization of the regularization—the corresponding optimization is simple as it only amounts to solving a set of linear equations. This is difficult to accomplish under the Poisson-likelihood model—the logarithm in $\mathcal{L}_{\mathrm{Poisson}}$ requires careful handling of the problem as positivity of the solution often becomes a harsh constraint.

C. *Computational Cost*

In our analysis so far, we ranked the various regularized interpolation algorithms purely based on performance gain. However, we must also consider the computation cost associated with these algorithms. The authors of [20] performed a thorough cost-performance analysis for standard interpolation; for a given quality measure, their emphasis was on reducing the cost of evaluating $\varphi(\mathbf{x}-\mathbf{k})$ for many arguments $(\mathbf{x}-\mathbf{k})$. In the context of regularized interpolation, since we are concerned with obtaining the coefficients c, we only consider the cost of computing c for a given φ.

For standard interpolation, c is computed by linear filtering of the data [20] which can be achieved with $\mathcal{O}(N)$ complexity. In the context of regularized interpolation, Algorithm 1 is easily implemented in the Fourier domain (via FFTs[5])—it requires about $\mathcal{O}(N\log(N))$ operations. Thus, with only a slightly larger computation load (an extra $\log(N)$ factor), Algorithm 1 yields significant improvement in quality and is generally preferable to standard interpolation.

Algorithms 2 and 3 are non-linear and iterative in nature: The CG-solver forms the predominant step in Algorithm 2 (see Appendix A-B), while in Algorithm 3, the major workload

[5]In this work, the Radix-2 Cooley-Tukey method was used which required the data to be zero-padded to a size that is a power of 2. Correspondingly, the time taken for Algorithm 1 is the same for the last three rows in Table IV.

TABLE IV
TIME TAKEN (IN SECONDS) BY VARIOUS SCHEMES FOR COMPUTING COEFFICIENTS

	Algorithm			
	Standard	Regularized		
Image Size		Algorithm 1	Algorithm 2	Algorithm 3
256×256	0.071	0.04	10.57	21.90
340×340	0.073	0.17	20.22	39.79
420×420	0.077	0.17	32.97	62.30
512×512	0.096	0.17	51.04	95.32

is in the computation of the constants $\mathcal{A}_{1\mathbf{q}}$, $\mathcal{A}_{2\mathbf{q}}$, and $\mathcal{A}_{3\mathbf{q}}$ (see Appendix A-C). These operations require the evaluation of several convolutions with the estimate c_t which are in turn repeated for each iteration. It is therefore clear that these algorithms are computationally more demanding than Algorithm 1.

We present in Table IV the execution times of standard (that of [20]) and regularized schemes (Algorithms 1-3) on a 2.66 GHz Intel Macintosh for rotation experiments with the circular phantom (shown in Figure 7a) of varying sizes. The cubic B-spline model was used in all algorithms. Algorithms 2 and 3 were executed until they satisfied the convergence criterion specified at the beginning of Section VI. The durations reported in each row of Table IV have been averaged over 10 realizations.

As expected, Algorithm 1 is much faster than Algorithms 2 and 3, while Algorithm 2 seems to have a considerable lead over Algorithm 3. The latter observation is in tune with our discussion at the end of Section VII-B. However, from a quality point of view, Figures 2a and 2b indicate that Algorithm 2 is preferable to Algorithm 1 for strong Gaussian noise. Therefore, Algorithm 2 is best-suited to carry out regularized interpolation in a general setting. However, at low noise levels, the effect of regularization becomes negligible (the output SNR curves in Figures 2a and 2b eventually meet); then, it may be desirable to use Algorithm 1 as it has the lowest computational complexity.

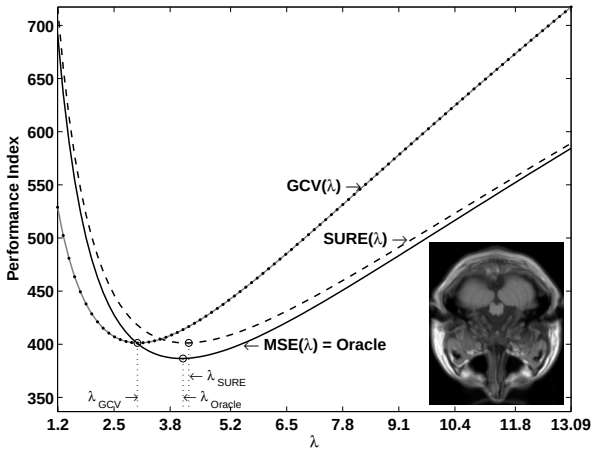


Fig. 8. Plot of GCV(λ), SURE(λ), and MSE(λ) *versus* λ for one realization (at 0 dB input SNR) of a rotation experiment in the presence of noise. We see that $\lambda_{\mathrm{SURE}} = 4.23$ is very near the optimum value $\lambda_{\mathrm{Oracle}} = 4.10$, unlike $\lambda_{\mathrm{GCV}} = 3.04$. Correspondingly, the index MSE(λ_{SURE}) $= 386.91$ is lower than MSE(λ_{GCV}) $= 400.72$ and is very close to the oracle minimum MSE($\lambda_{\mathrm{Oracle}}$) $= 386.65$, indicating that SURE-based selection leads to a better result.

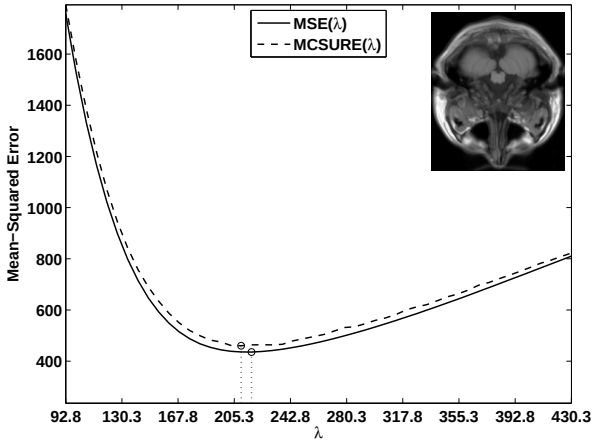


Fig. 9. MCSURE(λ) captures the trend of MSE(λ) for NQR-cubic interpolation scheme at 0 dB input SNR.

D. Selection of the Regularization Parameter

Since the regularization parameter λ balances regularization against fidelity to the data, choosing an appropriate λ is crucial for obtaining meaningful results. While our experiments used oracle-based tuning of λ, we now propose a data-driven means of selecting an appropriate value for this parameter to minimize the mean-squared error (MSE) given by the averaged-value of the denominator of (36)

$$\mathrm{MSE}(\lambda) = \frac{1}{N_\Omega} \sum_{\mathbf{k} \in \Omega} (f_{\mathrm{true}}(\mathbf{k}) - f_\lambda(\mathbf{k}))^2, \tag{37}$$

where N_Ω is the cardinality of the set Ω.

Generalized cross-validation (GCV), which does not require the knowledge of the noise variance, is often advocated for determining λ as it yields asymptotically optimal performance [26] (under certain hypotheses [50, Proposition 3.1]), i.e., λ_{GCV} minimizes (37) as $N_\Omega \to \infty$. However, its applicability is limited to linear algorithms (*e.g.*, Algorithm 1 in this paper).

Our goal here is to minimize (37) directly. To circumvent the dependence of MSE(λ) on the unknown noise-free samples $\{f_{\mathrm{true}}(\mathbf{k})\}_{\mathbf{k} \in \mathbb{Z}^d}$, we propose the use of Stein's unbiased risk estimate (SURE) [51], which provides a means for estimating MSE(λ) unbiasedly from the data g. Unlike GCV, SURE requires the knowledge of the noise variance. But, as it directly relates to MSE(λ), SURE is more robust than GCV and yields optimal performance even in the non-asymptotic case.

In the context of variational problems, SURE has been put to use under one form or another for tuning λ in smoothing-spline-like linear algorithms [26], [52]. For the non-linear case, we recently developed a Monte-Carlo scheme in [53]—Monte-Carlo SURE (MCSURE)—whose numerical evaluation requires twice the cost of the original algorithm but is applicable to iterative as well as non-iterative denoising methods with arbitrary nonlinearities. Since MSE(λ) measures the error in the data domain, MCSURE is applicable in the present scenario. So, we adopt the analytical SURE formula proposed

TABLE V
OUTPUT SNR OBTAINED BY MSE-BASED (ORACLE) AND SURE-BASED TUNING OF λ

Input SNR (dB)	Output SNR (dB)				
	QR-Cubic			NQR-Cubic	
	MSE-based	SURE-based	GCV-based	MSE-based	MCSURE-based
0	13.95	13.95	13.82	14.13	14.11
4	15.75	15.75	15.43	16.10	16.10
8	17.66	17.66	16.94	18.22	18.22
12	19.71	19.71	18.28	20.45	20.44
16	21.95	21.95	19.67	22.85	22.85
20	24.40	24.40	21.39	25.44	25.43
24	27.09	27.09	25.20	28.25	28.24
28	30.03	30.03	29.09	31.19	31.19
32	33.19	33.19	32.86	34.23	34.22
36	36.39	36.39	36.39	37.20	37.19
40	39.35	39.35	39.23	39.86	39.85
44	41.78	41.78	41.88	42.02	42.02
48	43.43	43.42	43.40	43.51	43.51
52	44.38	44.38	44.37	44.38	44.38
56	44.83	44.82	44.80	44.83	44.83
60	45.04	45.04	44.99	45.02	45.02

in [26, Theorem 1.1] for regularized interpolation based on a linear algorithm and propose the use of MCSURE for those based on a non-linear algorithm, respectively. In the sequel, we demonstrate that SURE(λ) and MCSURE(λ) not only mimic the behavior of MSE(λ) very closely, but also yield a good estimate of the optimal λ that minimizes MSE(λ).

We repeat now the experiment described in Section VI-A (with MRI stack) where, for each noisy realization, we compute GCV(λ) [26, Equation (1.9)] and SURE(λ) [26, Equation (1.8)] for QR-cubic and MCSURE(λ) [53, Equation (6)] for NQR-cubic regularized interpolation schemes along with MSE(λ) corresponding to each of these methods. The noise variance was assumed to be known in all the experiments for computing SURE(λ) and MCSURE(λ) (in practice, an estimate can be used). We plot GCV(λ) and SURE(λ) in Figure 8 and MCSURE(λ) in Figure 9, respectively, in comparison with the individual MSE(λ) for one realization of this experiment at 0 dB input SNR (test image shown in the inset). It is clearly seen that the SURE curves capture the trend of the corresponding MSE curves. Moreover, they also yield very close estimates of the optimal λ in both cases, while GCV fails to achieve this. This indicates that a SURE-based tuning procedure is more reliable than GCV even at such low levels of input SNR.

We further quantify the performance of GCV, SURE, and MCSURE in Table V where we list out the output SNR (averaged over all the realizations corresponding to an input SNR) obtained by minimizing GCV and SURE for the QR-cubic and MCSURE for the NQR-cubic interpolation algorithm, respectively. Also provided are the oracle values of the output SNR (obtained by minimizing the corresponding MSEs). We see that the GCV-based result is far from the oracle for low input SNRs (< 28 dB). This may be because the hypotheses [50, Proposition 3.1] required for the optimality of GCV are probably not fulfilled for the QR-cubic method. However, SURE-based tuning consistently yields SNR values very close to the oracle for both QR and NQR methods indicating that it can be reliably employed for data-driven adjustment of λ.

VIII. SUMMARY & CONCLUSIONS

Standard interpolation performs exact fitting of the given data. In the presence of noise, we have shown that this can have a detrimental influence on the interpolation quality. To interpolate noisy data, we have developed a regularized scheme that counterbalances the effect of noise by imposing smoothness constraints on the resulting continuous-domain solution. We have adopted an integer shift-invariant signal model for interpolation where the model parameters (coefficients of the integer-shift-invariant expansion) are obtained by minimizing the statistical infidelity of the solution to the given data (negative log-likelihood data term) subject to a regularization constraint prescribed as the L_p-norm of a general vector derivative of the solution. In addition to Tikhonov-like quadratic functional (which leads to smoothing-spline-like interpolants), the formulation includes non-quadratic ones in the spirit of edge-preserving total-variation regularization. We have presented algorithms based on the MM (majorize-minimize) strategy to perform the optimization of the coefficients for non-quadratic cost criteria.

We have carried out 2-D rotation experiments in the presence of noise and numerically quantified the performance of standard (non-regularized) and regularized interpolation. We observe that regularized interpolation always yields a significant improvement in quality over standard interpolation. Furthermore, among the regularized methods, we could note that, in the presence of noise, non-quadratic regularized (NQR) interpolation consistently outperforms the quadratic regularized (QR) one; we associate this phenomenon to the fact that NQR interpolation preserves edges and achieves better noise reduction than QR interpolation. We also illustrated

that Stein's unbiased risk estimate (SURE) can be reliably employed for data-driven selection of the regularization parameter for performing optimally regularized (QR and NQR) interpolation. We conclude from these observations that regularized interpolation, specifically, NQR interpolation, can be of potential interest in medical-imaging applications.

APPENDIX A
ALGORITHMS FOR REGULARIZED INTERPOLATION

A. Minimization of $J_{\lambda,\mathrm{GLQ}}$

Algorithm 1: Algorithm for minimizing $J_{\lambda,\mathrm{GLQ}}$ for a given λ

- Step 1: Precompute and store $B(e^{j\boldsymbol{\omega}})$, $R(e^{j\boldsymbol{\omega}})$, and $G(e^{j\boldsymbol{\omega}})$
- Step 2: Construct $H_\lambda(e^{j\boldsymbol{\omega}})$ using (23) for the given λ and compute $C_\lambda(e^{j\boldsymbol{\omega}})$
- Step 3: Perform inverse Fourier transform of C_λ to obtain c_λ

B. Minimization of $J_{\lambda,\mathrm{GLNQ}}$

From (29), we note that $J_{\mathrm{AUX}}\{g, c|c_t\}$ is quadratic in c. We minimize it by setting its derivative to zero which results in the following system of linear equations:

$$(c * b * \bar{b})[\mathbf{k}] + \lambda\frac{p}{2}\sum_{\substack{\mathbf{q}\in\mathbb{Z}^d \\ m=1\cdots s}} |\chi_t[\mathbf{q}]|^{p-2}\,(c*\eta_m)[\mathbf{q}]\,\eta_m[\mathbf{q}-\mathbf{k}] = (g*\bar{b})[\mathbf{k}],\ \forall\ \mathbf{k}\in\mathbb{Z}^d, \tag{38}$$

where $\bar{b}[\mathbf{k}] = b[-\mathbf{k}]$. The presence of χ_t in (38) prohibits the use of Fourier-domain techniques such as those used in Section IV-A. Nevertheless, the MM technique requires only a decrease of J_{AUX}; there is no need to minimize it completely. Therefore, efficient numerical schemes such as the conjugate-gradient (CG) method can be utilized to solve (38) partially. For $p < 2$, the factor $|\chi_t[\cdot]|^{p-2}$ can pose a problem whenever $\chi_t = 0$. An ad hoc modification is to saturate χ_t to a small positive value ϵ (i.e., $\chi_t \leftarrow \max(\epsilon, \chi_t)$) to avoid instabilities in (38). In our implementation, we chose the initial estimate c_0 to be the given data g slightly perturbed by additive noise. This ensured that $\chi_t \neq 0$ at all iterations.

Algorithm 2: MM algorithm for minimizing $J_{\lambda,\mathrm{GLNQ}}$ for a given λ and p

- Step 1: Precompute the sequences $\bar{b} * b$ and $g * \bar{b}$
- Step 2: Initial estimate = c_0; $t = 0$
  Repeat Step 3 to 5 until **Stop Criterion** is met
- Step 3: Compute the sequence χ_t using c_t
- Step 4: Apply CG iterations to partially solve for c_{t+1} in (38)
- Step 5: Set $t = t + 1$

While we used Algorithm 2 in all our experiments, it may be beneficial to consider the following alternative which provides a mathematically rigorous means of tackling the instability of (38). This procedure applies for minimizing J_{GLNQ} and requires the assumption that the kernel b has a stable convolution inverse b^{-1} such that $(b^{-1} * b)[\mathbf{k}] = \delta[\mathbf{k}]$. Specifically, let $\mathbf{B}$, $\mathbf{B}^{-1}$, and $\mathbf{R}_m$ denote toeplitz matrices corresponding to the sequences $\{b[\mathbf{k}]\}_{\mathbf{k}\in\mathbb{Z}^d}$, $\{b^{-1}[\mathbf{k}]\}_{\mathbf{k}\in\mathbb{Z}^d}$, and $\{\eta_m[\mathbf{k}]\}_{\mathbf{k}\in\mathbb{Z}^d}$, respectively. Then, rewriting (38) in terms of matrices, we obtain

$$(\mathbf{B}^{\mathrm{T}}\,\mathbf{B} + \mathbf{R}^{\mathrm{T}}\,\boldsymbol{\Upsilon}_t\,\mathbf{R})\,\mathbf{c} = \mathbf{B}^{\mathrm{T}}\,\mathbf{g}, \tag{39}$$

where $\mathbf{c}$ and $\mathbf{g}$ are column vectors containing sequences $\{c[\mathbf{k}]\}_{\mathbf{k}\in\mathbb{Z}^d}$, and $\{g[\mathbf{k}]\}_{\mathbf{k}\in\mathbb{Z}^d}$, respectively, and $\mathbf{R} = [\mathbf{R}_1^{\mathrm{T}}\,\mathbf{R}_2^{\mathrm{T}}\cdots\mathbf{R}_s^{\mathrm{T}}]^{\mathrm{T}}$, and $\boldsymbol{\Upsilon}_t = \lambda\frac{p}{2}\,\mathrm{diag}\{\underbrace{\Upsilon_t, \ldots, \Upsilon_t}_{s \text{ times}}\}$, with Υ_t a diagonal matrix containing entries $\{|\chi_t[\mathbf{k}]|^{p-2}\}_{\mathbf{k}\in\mathbb{Z}^d}$. Performing manipulations on (39) using matrix inversion lemma, we get

$$\mathbf{c} = \mathbf{B}^{-1}\,\mathbf{g} - (\mathbf{B}^{\mathrm{T}}\,\mathbf{B})^{-1}\mathbf{R}^{\mathrm{T}}\,\mathbf{z}, \tag{40}$$

$$(\boldsymbol{\Upsilon}_t^{-1} + \mathbf{R}\,(\mathbf{B}^{\mathrm{T}}\,\mathbf{B})^{-1}\,\mathbf{R}^{\mathrm{T}})\,\mathbf{z} = \mathbf{R}\,\mathbf{B}^{-1}\,\mathbf{g}. \tag{41}$$

Rewriting (40) and (41) in terms of convolutions, we obtain

$$c[\mathbf{k}] = (b^{-1} * g)[\mathbf{k}] - \sum_{m=1}^{s}(b^{-1} * \bar{b}^{-1} * \bar{\eta}_m * z_m)[\mathbf{k}], \tag{42}$$

where $\bar{b}^{-1}[\mathbf{k}] = b^{-1}[-\mathbf{k}]$, $\bar{\eta}_m[\mathbf{k}] = \eta[-\mathbf{k}]$, and $\{z_m[\mathbf{k}]\}_{m=1,\,\mathbf{k}\in\mathbb{Z}^d}^{s}$ is the solution of the following linear systems of equations: $\forall\ \mathbf{k}\in\mathbb{Z}^d$, and $m = 1, \ldots, s$,

$$\frac{2}{\lambda p}|\chi_t[\mathbf{k}]|^{2-p}\,z_m[\mathbf{k}] + \sum_{l=1}^{s}(\eta_m * b^{-1} * \bar{b}^{-1} * \bar{\eta}_l * z_l)[\mathbf{k}] = (\eta_m * b^{-1} * g)[\mathbf{k}]. \tag{43}$$

Thus, at iteration t, we employ the CG method to partially solve for a set of temporary variables $\{z_m\}_{m=1}^{s}$ using (43) which is then used to update c in (42). The linear system corresponding to $\{z_m\}_{m=1}^{s}$ now involves $|\chi_t[\cdot]|^{2-p}$ rather than $|\chi_t[\cdot]|^{p-2}$ which does not explode when $\chi_t = 0$ for $p < 2$. Based on (40)-(43), we provide the following alternative to Algorithm 2 for minimizing J_{GLNQ}:

Algorithm 2a: MM algorithm for minimizing $J_{\lambda,\mathrm{GLNQ}}$ for a given λ and p

- Step 1: Precompute $b^{-1} * g$, $\eta_m * b^{-1} * g$, $b^{-1} * \bar{b}^{-1} * \bar{\eta}_m$, and $\eta_m * b^{-1} * \bar{b}^{-1} * \bar{\eta}_l$, $l, m = 1, 2, \ldots, s$
- Step 2: Initial estimate = c_0; $t = 0$
  Repeat Steps 3 to 6 until **Stop Criterion** is met
- Step 3: Compute the sequence χ_t using c_t
- Step 4: Apply CG iterations to partially solve for $\{z_{m,t}\}_{m=1}^{s}$ in (43)
- Step 5: Compute c_{t+1} in (42) using $\{z_{m,t}\}_{m=1}^{s}$
- Step 6: Set $t = t + 1$

C. Minimization of $J_{\lambda,\mathrm{PLNQ}}$

From (30) and (31), we see that minimizing $J_{\lambda,\mathrm{PLNQ}}\{g, c\}$ simply amounts to minimizing each of the individual $J_{\mathrm{AUX}}\{g, c|c_t, \mathbf{q}\}$ which is accomplished by setting their derivative with respect to $c[\mathbf{q}]$ to zero. The solution is given by

$$c[\mathbf{q}] = \frac{1}{2\lambda\,p\,\mathcal{A}_{2\mathbf{q}}}\left(-\gamma + \sqrt{\gamma^2 + 4\lambda\,p\,c_t[\mathbf{q}]\,\mathcal{A}_{1\mathbf{q}}\,\mathcal{A}_{2\mathbf{q}}}\right), \tag{44}$$

where $\gamma = \lambda\, p\,(\mathcal{A}_{3\mathbf{q}} - c_t[\mathbf{q}]\mathcal{A}_{2\mathbf{q}}) + \mathcal{A}_b$. Since, the sequences b and χ_t are non-negative and because g represents the data counts under the Poisson noise model, we have that $\mathcal{A}_{1\mathbf{q}} > 0$ and $\mathcal{A}_{2\mathbf{q}} > 0$ whenever $c_t[\mathbf{q}] > 0\ \forall\ \mathbf{q} \in \mathbb{Z}^d$. Thus, the solution given by (44) is always positive. Based on (30)-(34) and (44), we present Algorithm 3 for the minimization of $J_{\lambda,\mathrm{PLNQ}}$.

In the case of quadratic regularization, we have that

$$\Psi_{\mathrm{Q}}\{c\} = \langle c, r * c\rangle_{\ell_2} = \sum_{\mathbf{k}\in\mathbb{Z}^d}\sum_{m=1}^{s}((c * r_{m\frac{1}{2}})[\mathbf{k}])^2, \qquad (45)$$

where $r_{m\frac{1}{2}}$ is specified via its z-transform $R_{m\frac{1}{2}}(\boldsymbol{z})|_{\boldsymbol{z}=e^{j\boldsymbol{\omega}}} = \sqrt{R_m(e^{j\boldsymbol{\omega}})}$ wherein R_m is the z-transform of the sequence r_m in (21). Comparing (22) and (45), we see that Algorithm 3 can be applied to minimize $J_{\lambda,\mathrm{PLQ}}$ by setting $p = 2$ and replacing η_m by $r_{m\frac{1}{2}}$ in (33) and (34).

Algorithm 3: MM algorithm for minimizing $J_{\lambda,\mathrm{PLNQ}}$ for a given λ and p

Step 1: Precompute $\mathcal{A}_b$ and the sequence μ_m in (33)
Step 2: Initial estimate = c_0; $t = 0$
Repeat Steps 3 to 5 until **Stop Criterion** is met
Step 3: Update constants $\mathcal{A}_{1\mathbf{q}}$, $\mathcal{A}_{2\mathbf{q}}$, $\mathcal{A}_{3\mathbf{q}}$ using (32)-(34) $\forall\ \mathbf{q} \in \mathbb{Z}^d$
Step 4: Compute c_{t+1} using (44)
Step 5: Set $t = t + 1$

ACKNOWLEDGEMENT

This work was supported by the Swiss National Science Foundation under grant 200020-109415 and in part under fellowship PBELP2-125446. We thank the reviewers for pointing out enhancements on theoretical and practical aspects which have improved the quality and the presentation of the paper.

REFERENCES

[1] T. M. Lehmann, C. Gönner, and K. Spitzer, "Survey: Interpolation Methods in Medical Image Processing," *IEEE Trans. Medical Imaging*, vol. 18, no. 11, pp. 1049–1075, November 1999.

[2] T. M. Lehmann, C. Gönner, and K. Spitzer, "Addendum: B-Spline Interpolation in Medical Image Processing," *IEEE Trans. Medical Imaging*, vol. 20, no. 7, pp. 660–665, July 2001.

[3] R. W. Parrot, M. R. Stytz, P. Amburn, and D. Robinson, "Towards Statistically Optimal Interpolation for 3-D Medical Imaging," *IEEE Eng. Med. Biol.*, vol. 12, no. 5, pp. 49–59, September 1993.

[4] E. Meijering, W. J. Niessen, and M. A. Viergever, "Quantitative Evaluation of Convolution-Based Methods for Medical Image Interpolation," *Med. Img. Anal.*, vol. 5, no. 2, pp. 111–126, June 2001.

[5] P. Thévenaz, T. Blu, and M. Unser, "Image Interpolation and Resampling," in *Handbook of Medical Imaging, Processing and Analysis*, I. N. Bankman, Ed., pp. 393–420. Academic Press, San Diego, CA, USA, 2000.

[6] E. Meijering, "A Chronology of Interpolation: From Ancient Astronomy to Modern Signal and Image Processing," *Proceedings of IEEE*, vol. 90, no. 3, pp. 319–342, 2002.

[7] E. Meijering, K. J. Zuiderveld, and M. A. Viergever, "Image Registration for Digital Subtraction Angiography," *Int. J. Comput. Vision*, vol. 32, no. 2-3, pp. 227–246, 1999.

[8] V. Walimbe and R. Shekhar, "Automatic Elastic Image Registration by Interpolation of 3-D Rotations and Translations from Discrete Rigid-Body Transformations," *Med. Img. Anal.*, vol. 10, no. 6, pp. 899–914, 2006.

[9] M. Holden, "A Review of Geometric Transformations for Nonrigid Body Registration," *IEEE Trans. Medical Imaging*, vol. 27, no. 1, pp. 111–128, January 2008.

[10] M. R. Smith and S. T. Nichols, "Efficient Algorithms for Generating Interpolated (Zoomed) MR Images," *Magn. Reson. Med.*, vol. 7, no. 2, pp. 156–171, June 1988.

[11] E. B. Ameur, D. Sbibih, A. Almhdie, and C. Léger, "New Spline Quasi-Interpolant for Fitting 3-D Data on the Sphere: Applications to Medical Imaging," *IEEE Sig. Proc. Lett.*, vol. 14, no. 5, pp. 333–336, May 2007.

[12] A. Kadosh, D. Cohen-Or, and R. Yagel, "Tricubic Interpolation of Discrete Surfaces for Binary Volumes," *IEEE Trans. Visual. Comput. Graph.*, vol. 9, no. 4, pp. 580–586, October-December 2003.

[13] O. Sommer, A. Dietz, R. Westermann, and T. Ertl, "An Interactive Visualization and Navigation Tool for Medical Volume Data," *Computers & Graphics*, vol. 23, no. 2, pp. 233–244, 1999.

[14] B. Csébfalvi, "An Evaluation of Prefiltered Reconstruction Schemes for Volume Rendering," *IEEE Trans. Visual. Comput. Graph.*, vol. 14, no. 2, pp. 289–301, March-April 2008.

[15] M. Unser, A. Aldroubi, and M. Eden, "B-Spline Signal Processing: Part I—Theory," *IEEE Trans. Signal Processing*, vol. 41, no. 2, pp. 821–832, February 1993.

[16] T. Möller, R. Machiraju, K. Mueller, and R. Yagel, "Evaluation and Design of Filters Using a Taylor Series Expansion," *IEEE Trans. Visual. Comput. Graph.*, vol. 3, no. 2, pp. 184–199, April-June 1997.

[17] F. M. Weinhaus and V. Devarajan, "Texture Mapping 3D Models of Real-World Scenes," *ACM Comput. Surv.*, vol. 29, no. 4, pp. 325–365, December 1997.

[18] Y. Tang, J. Wang, H. Bao, and Q. Peng, "RBF-Based Constrained Texture Mapping," *Computers & Graphics*, vol. 27, no. 3, pp. 415–422, 2003.

[19] S. Heim, L. Fahrmeir, P. H. C. Eilers, and B. D. Marx, "3D Space-Varying Coefficient Models with Application to Diffusion Tensor Imaging," *Computation. Stat. Data Anal.*, vol. 51, no. 12, pp. 6212–6228, 2007.

[20] P. Thévenaz, T. Blu, and M. Unser, "Interpolation Revisited," *IEEE Trans. Medical Imaging*, vol. 19, no. 7, pp. 739–758, July 2000.

[21] R. R. Schultz and R. L. Stevenson, "A Bayesian Approach to Image Expansion for Improved Definition," *IEEE Trans. Image Processing*, vol. 3, no. 3, pp. 233–242, May 1994.

[22] A. Raj and K. Thakur, "Fast and Stable Bayesian Image Expansion Using Sparse Edge Priors," *IEEE Trans. Image Processing*, vol. 16, no. 4, pp. 1073–1084, April 2007.

[23] J. H. Shin, J. H. Jung, and J. K. Paik, "Regularized Iterative Image Interpolation and Its Application to Spatially Scalable Coding," *IEEE Trans. Cons. Elec.*, vol. 44, no. 3, pp. 1042–1047, August 1998.

[24] C. M. Reinsch, "Smoothing by Spline Functions," *Numer. Math.*, vol. 10, no. 3, pp. 177–183, 1967.

[25] I. J. Schoenberg, "Spline Functions and the Problem of Graduation," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 52, no. 4, pp. 947–950, 1964.

[26] P. Craven and G. Wahba, "Smoothing Noisy Data with Spline Functions—Estimating the Correct Degree of Smoothing by the Method of Generalized Cross-Validation," *Numer. Math.*, vol. 31, no. 4, pp. 377–403, 1979.

[27] M. Unser and T. Blu, "Generalized Smoothing Splines and the Optimal Discretization of the Wiener Filter," *IEEE Trans. Signal Processing*, vol. 53, no. 6, pp. 2146–2159, 2005.

[28] Y. C. Eldar and M. Unser, "Nonideal Sampling and Interpolation From Noisy Observations in Shift-Invariant Spaces," *IEEE Trans. Signal Processing*, vol. 54, no. 7, pp. 2636–2651, July 2006.

[29] S. Ramani, D. Van De Ville, T. Blu, and M. Unser, "Nonideal Sampling and Regularization Theory," *IEEE Trans. Signal Processing*, vol. 56, no. 3, pp. 1055–1070, March 2008.

[30] D. R. Hunter and K. Lange, "A Tutorial on MM Algorithms," *The American Statistician*, vol. 59, no. 1, pp. 30–37, February 2004.

[31] M. Unser, "Splines: A Perfect Fit for Signal and Image Processing," *IEEE Sig. Proc. Mag.*, vol. 16, no. 6, pp. 22–38, November 1999.

[32] T. Blu and M. Unser, "Quantitative Fourier Analysis of Approximation Techniques: Part I—Interpolators and Projectors," *IEEE Trans. Signal Processing*, vol. 47, no. 10, pp. 2783–2795, October 1999.

[33] P. J. Green, "On Use of the EM for Penalized Likelihood Estimation," *J. Royal Stat. Soc. B*, vol. 52, no. 3, pp. 443–452, 1990.

[34] A. R. De Pierro, "A Modified Expectation Maximization Algorithm for Penalized Likelihood Estimation in Emission Tomography," *IEEE Trans. Medical Imaging*, vol. 14, no. 1, pp. 132–137, March 1995.

[35] M. Yavuz and J. A. Fessler, "Penalized-Likelihood Estimators and Noise Analysis for Randoms-Precorrected PET Transmission Scans," *IEEE Trans. Medical Imaging*, vol. 18, no. 8, pp. 665–674, August 1999.

[36] C. Vonesch, F. Aguet, J. -L. Vonesch, and M. Unser, "The Colored Revolution of Bioimaging," *IEEE Sig. Proc. Mag.*, vol. 23, no. 3, pp. 20–31, May 2006.

[37] L. A. Shepp and Y. Vardi, "Maximum Likelihood Reconstruction for Emission Tomography," *IEEE Trans. Medical Imaging*, vol. 1, no. 2, pp. 113–122, October 1982.

[38] I. M. Gelfand and G. E. Shilov, *Generalized Functions*, vol. 1, Academic Press, New York, 1964.

[39] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear Total Variation Based Noise Removal Algorithms," *Physica D*, vol. 60, no. 1-4, pp. 259–268, November 1992.

[40] T. Chan, S. Esedoglu, F. Park, and A. Yip, "Recent Developments in Total Variation Image Restoration," in *Mathematical Models in Computer Vision*, N. Paragios, Y. Chen, and O. Faugeras, Eds. 2005, pp. 17–30, Springer-Verlag.

[41] T. Le, R. Chartrand, and T. J. Asaki, "A Variational Approach to Reconstructing Images Corrupted by Poisson Noise," *J. Math. Img. Vis.*, vol. 27, no. 3, pp. 257–263, 2007.

[42] J. Kybic, T. Blu, and M. Unser, "Generalized Sampling: A Variational Approach—Part I: Theory," *IEEE Trans. Signal Processing*, vol. 50, no. 8, pp. 1965–1976, August 2002.

[43] 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., pp. 85–100. Springer-Verlag, Berlin, Germany, 1977.

[44] C. T. Kelley, *Iterative Methods for Optimization*, SIAM, Philadelphia, USA., 1999.

[45] D. Geman and G. Reynolds, "Constrained Restoration and Recovery of Discontinuities," *IEEE Trans. Pat. Anal. Machine Intell.*, vol. 14, no. 3, pp. 367–383, 1992.

[46] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, "Deterministic Edge-Preserving Regularization in Computed Imaging," *IEEE Trans. Image Processing*, vol. 2, no. 6, pp. 298–311, February 1997.

[47] M. Nikolova and M. K. Ng, "Analysis of Half-Quadratic Minimization Methods for Signal and Image Recovery," *SIAM Journal on Scientific Computing*, vol. 27, no. 3, pp. 937–966, 2005.

[48] "Brainweb: Simulated MRI Volumes for Normal Brain," McConnell Brain Imaging Center, www.bic.mni.mcgill.ca/brainweb/selection_normal.html.

[49] R. G. Keys, "Cubic Convolution Interpolation for Digital Image Processing," *IEEE Trans. Acoust. Speech, and Signal Processing*, vol. ASSP-29, no. 6, pp. 1153–1160, December 1981.

[50] K. C. Li, "Asymptotic Optimality of C_L and Generalized Cross-validation in Ridge Regression with Application to Spline Smoothing," *Ann. Stat.*, vol. 14, no. 3, pp. 1101–1112, 1986.

[51] C. Stein, "Estimation of the Mean of a Multivariate Normal Distribution," *Ann. Stat.*, vol. 9, no. 6, pp. 1135–1151, 1981.

[52] C. L. Mallows, "Some Comments on C_P," *Technometrics*, vol. 15, no. 4, pp. 661–675, 1973.

[53] S. Ramani, T. Blu, and M. Unser, "Monte-Carlo SURE: A Black-Box Optimization of Regularization Parameters for General Denoising Algorithms," *IEEE Trans. Image Processing*, vol. 17, no. 9, pp. 1540–1554, September 2008.