Image Grid Invariant Regularization for Iterative Reconstruction

Zhou Yu, Charles A. Bouman, Jean-Baptiste Thibault, and Ken D. Sauer

***Abstract*— Regularization methods have been successfully applied to various reconstruction and denoising problems. In these problems, one needs to choose regularization parameters that properly balance resolution and noise. These parameters need to be adjusted when the image grid pattern changes. In this paper, we present a theory on regularization design derived from a continuous image model so that the regularization method is invariant to image grid pattern. We can use this theory to compute regularization parameters for various image grid patterns. To illustrate the idea, we applied this theory to regularization design on a rectangular image grid.**

***Index Terms*— Iterative reconstruction, regularization method, image model, multi-grid**

I. Introduction

Iterative reconstruction (IR) methods have been widely applied to solve various reconstruction and denoising problems in medical imaging, security CT, electron-microscopy, etc [1]–[3]. IR methods typically solve the reconstruction problem by formulating mathematical models of the physics and statistics of the imaging process and the image itself.

One important aspect of the modeling is to find a discrete model of the image and physical processes which are generally speaking continuous in nature. In a p-dimensional reconstruction problem, we typically model the image object as a continuous function $f(r) : \Re^p \mapsto \Re$, where r is the vector representing spatial location. The input to a reconstruction problem is a set of discrete measurements denoted by a vector y. The output of the reconstruction is a discrete image array denoted by vector x. To define a discrete representation of the image, one can define x to be the samples of f,

$$x_i = f(r_i) \tag{1}$$

where i is the pixel index and r_i are typically chosen to fall on a periodic grid.

The imaging process can be modeled as a mapping from f to y, that is, $y = F(f)$. For example, in 2D parallel beam CT reconstruction, F is the Radon transform. Once we define a discrete representation of f, we can also derive a discrete forward model, $y \approx \tilde{F}(x)$, mapping from x to y. In the CT reconstruction problem, for instances, various discrete forward models have been proposed [4]–[7]. Once the models are built, one can formulate a cost function to find the solution that best fits the model. In general, the image is reconstructed by minimizing a cost function such as,

$$\hat{x} = \arg\min \left\{ G(\tilde{F}(x), y) + U(x) \right\}, \tag{2}$$

in which $G(\tilde{F}(x), y)$ is the data mismatch term that penalizes the differences between the image x and the measurement y according to the forward model $\tilde{F}$ and a statistical model, and $U(x)$ is the regularization function that penalizes the roughness in the image [8]–[10]. $U(x)$ is typically derived from a prior model of the image, such as the Markov random field (MRF) model [11]. The regularization term $U(x)$ plays a very important role in defining the image quality [12].

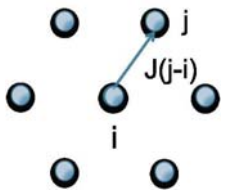


Fig. 1. This figure shows an example neighborhood on a non-rectangular periodic 2D image grid.

In many imaging problems, the image grid might vary from case to case. For example, when reconstructing images on rectangular grids, the sampling rate along each direction might vary depending on the desired resolution. In other cases, the images might not necessarily fall on rectangular grids. Therefore we would like to design a cost function that is invariant to the image grid. In equation (2), the discrete forward model $\tilde{F}$ naturally takes the pixel size and location into consideration. In equation (28) of [13], Oh et.al proposed a scale invariant design of $U(x)$, assuming the sampling pattern (in this case, the pixel's aspect ratio) remains the same. In this paper, we propose a general theory to design $U(x)$ to be image grid invariant. To do this, we first approximate $U(x)$ to be an integral of f. In order to be image grid invariant, this integral must not depend on the pixel location or the choice of the neighborhood. Based on this condition, we can derive a set of sufficient conditions to guide the design of $U(x)$. We found that these conditions can be satisfied in the case of quadratic regularization or when $U(x)$ has a special form. To provide an example of the theory, we describe a design for a commonly used 3D rectangular grid in section III. Finally, we apply the design to the 3D CT reconstruction problem as an example.

II. Theory

Let us first introduce notation to describe a periodic image grid. We use $i = (i_1, i_2, \ldots, i_p)$ and $j = (j_1, j_2, \ldots, j_p)$ to denote p-dimensional discrete indexes; and s to denote the coordinates of a pixel. Let S be a lattice of pixels at locations

Zhou Yu and Jean-Baptiste Thibault are with GE Healthcare, 3000 N Grandview Bvd, W-1180, Waukesha, WI 53188.

Charles Bouman is with the School of Electrical Engineering, Purdue University, West Lafayette, IN 47907-0501.

Ken Sauer is with the Department of Electrical Engineering, 275 Fitzpatrick, University of Notre Dame, Notre Dame, IN 46556-5637.

$s_i \in \Re^p$ taking on the values $f(s_i)$. Furthermore, let $\mathcal{P} \subset S^2$ be a neighborhood system on S, where $\mathcal{P}$ consists of all voxel pairs $\{i, j\}$ where i is a neighbor of j. We assume that the lattice S has a periodic structure, so the neighbors of $i \in S$ are $j = i+k$ where $k \in W$, and W is a set of neighboring pixels' index displacements. We also assume that W is symmetric, so that if $k \in W$ then $-k \in W$. We define J as a p by p transform matrix that computes the displacement of two neighboring pixels from k,

$$s_j - s_i = Jk. \tag{3}$$

In this notation, the neighborhood is defined by the set W describing the selection of the neighboring pixels and the matrix J describing the displacement between each neighboring pixel pair. Fig. 1 shows an exemplary neighborhood on a non-rectangular periodic 2D image grid, in which each pixel has 6 neighbors.

Let us also assume the prior distribution of x as a Markov random field (MRF), and we will discuss a general form later. In the MRF case, $U(x)$ is of the form,

$$U(x) = \frac{1}{\alpha} \sum_{\{i,j\} \in \mathcal{P}} b_{j-i} \rho(x_j - x_i) \tag{4}$$

where $\rho(\cdot)$ is the potential function. Our objective is to design α and b_{j-i} so that $U(x)$ is invariant to the image grid.

We first derive $U(x)$ as a discrete approximation to an integral of f. To do this, we use the finite difference in the image to approximate the local directional gradient of f, that is,

$$x_j - x_i \approx \nabla f(s_i)^t J(j-i). \tag{5}$$

Using this approximation, we can rewrite $U(x)$ as,

$$\begin{aligned} U(x) &= \frac{1}{\alpha} \sum_{\{i,j\} \in \mathcal{P}} b_{j-i} \rho(x_j - x_i) \\ &= \frac{1}{2\alpha} \sum_{i \in S} \sum_{k \in W} b_k \rho(x_{i+k} - x_i) \\ &\approx \frac{1}{2\alpha} \sum_{i \in S} \sum_{k \in W} b_k \rho(\nabla f(s_i)^t Jk) \\ &= \frac{1}{2\alpha|J|} \sum_{k \in W} b_k \sum_{i \in S} \rho(\nabla f(s_i)^t Jk)|J| \\ &\approx \frac{1}{2\alpha|J|} \sum_{k \in W} b_k \int_{\Re^p} \rho(\nabla f(s)^t Jk) ds, \end{aligned} \tag{6}$$

where $|J|$ is the determinant of matrix J.

Notice that, in equation (6), the right hand side is a summation of integrals, in which each integral is based on the gradient along the direction of a specific neighbor, k. Since the right hand side approximation still depends on the choice of J and W, it is not image grid invariant in general. However, in some special cases, it is possible to use b_k to compensate for the directional change in the neighborhood. In the following, we derive a sufficient condition that b_k needs to satisfy to yield an image grid invariant regularization in two cases. In the first case, ρ must be a quadratic function. In the second case, ρ can be of a general form, however we need to introduce a minor modification to the form of $U(x)$.

A. Quadratic Regularization

Assume $\rho(\Delta) = \Delta^2$, then

$$\begin{aligned} U(x) &\approx \frac{1}{2\alpha|J|} \int_{\Re^p} \sum_{k \in W} b_k \nabla f(s)^t Jkk^t J^t \nabla f(s) ds \\ &= \int_{\Re^p} ||\nabla f(s)||_H^2 ds \end{aligned} \tag{7}$$

where,

$$H = \frac{1}{2\alpha|J|} \sum_{k \in W} b_k J(kk^t) J^t \tag{8}$$

To use equation (8) to design $U(x)$, we first need to choose a desired H matrix, such as an identity matrix. We then choose the image grid and neighborhood, i.e. J and W. Finally, we find b_k so that the equation (8) holds. Therefore, equation (8) gives a sufficient condition for image grid invariant regularization design.

B. General Potential Function

In many imaging applications, $\rho(x)$ is designed to suppress noise while preserving spatial resolution. To apply the theory to general potential functions, we would like to propose a different form of $U(x)$, where

$$U(x) = \frac{1}{\alpha} \sum_{\{i,j\} \in \mathcal{P}} \rho(b_{i-j}^2 (x_j - x_i)^2) \tag{9}$$

In this form, we sum over the squared difference between each neighboring voxel pairs first, then apply the penalty function $\rho(\cdot)$. Similar to the derivation of equation (6) and (7) , $U(x)$ can be shown to approximate the following integral,

$$U(x) \approx \frac{1}{2\alpha|J|} \int_{\Re^p} \rho(||\nabla f(s)||_H^2) ds, \tag{10}$$

where H is given by equation (8)

III. Application to 3D Rectangular Grid

In this section, we would like to provide an example design for a 3D rectangular image grid, with voxels spacing $\Delta_x = \Delta_y$, and Δ_z, where Δ_x and Δ_y are voxel sizes along x, y and z axis respectively.

We assume W to be a 3 by 3 by 3 cubic neighborhood, therefore, $W = \{-1, 0, 1\}^3$. In this case,

$$J = \begin{bmatrix} \Delta_x & 0 & 0 \\ 0 & \Delta_y & 0 \\ 0 & 0 & \Delta_z \end{bmatrix}$$

The geometry of the voxel neighborhood is illustrated in Fig. 2, in which θ is the angle between $k = (1, 0, 1)^t$ and the x axis as shown in (a), and γ is the angle between $k = (1, 1, 1)$ and the $x - y$ plane as shown in (b). Thus, $\cos\theta = \frac{1}{\sqrt{1+\beta^2}}$ and $\cos\gamma = \frac{\sqrt{2}}{\sqrt{2+\beta^2}}$, where $\beta = \frac{\Delta_z}{\Delta_x}$ is the aspect ratio of the voxel in this particular case where $\Delta_x = \Delta_y$. When the aspect ratio changes, the direction between two neighbor voxels, Jk, also changes. Without proper compensation using b_k, the regularization might be stronger in one direction relative to

the other. Let us assume we would like the regularization to be isotropic. Therefore, we choose $H = I$. In the following we will derive a solution for b_k so that equation (8) holds in this case.

Let $R_k = \frac{(Jk)^t Jk}{||Jk||^2}$, so that we can rewrite equation (7) as

$$\frac{1}{2\alpha|J|}\sum_{k\in W} b_k||Jk||^2 R_k = H \tag{11}$$

A sufficient condition for equation (11) to hold is $\alpha = \frac{1}{2|J|}$, $b_k = \frac{w_k}{||J_k||^2}$, where w_k satisfy,

$$\sum_{k\in W} w_k R_k = H \tag{12}$$

Notice that, $||J_k||$ is the distance between the two-voxel pair $(i, i+k)$. Therefore, in this case, b_k is inversely proportional to the squared distance, and w_k is computed to compensate for the different effective regularization strength along the direction of the neighbors defined by the image grid. We can derive the R_k matrices for each direction as follows:

$$R_{\pm1,0,0} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, R_{0,\pm1,0} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix};$$

$$R_{0,0,\pm1} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}; R_{\pm1,\pm1,0} = \begin{bmatrix} \frac{1}{2} & \pm\frac{1}{2} & 0 \\ \pm\frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 \end{bmatrix};$$

$$R_{\pm1,0,\pm1} = \begin{bmatrix} \cos^2\theta & 0 & \pm\sin\theta\cos\theta \\ 0 & 0 & 0 \\ \pm\sin\theta\cos\theta & 0 & \sin^2\theta \end{bmatrix};$$

$$R_{0,\pm1,\pm1} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \cos^2\theta & \pm\sin\theta\cos\theta \\ 0 & \pm\sin\theta\cos\theta & \sin^2\theta \end{bmatrix};$$

$$R_{\pm1,\pm1,\pm1} = \begin{bmatrix} \frac{1}{2}\cos^2\gamma & \pm\frac{1}{2}\cos^2\gamma & \pm\frac{1}{\sqrt{2}}\cos\gamma\sin\gamma \\ \pm\frac{1}{2}\cos^2\gamma & \frac{1}{2}\cos^2\gamma & \pm\frac{1}{\sqrt{2}}\cos\gamma\sin\gamma \\ \pm\frac{1}{\sqrt{2}}\cos\gamma\sin\gamma & \pm\frac{1}{\sqrt{2}}\sin\gamma\cos\gamma & \sin^2\gamma \end{bmatrix} \tag{13}$$

In general the solution to equation (12) might not be unique. In the following we will derive a solution for w_k when $H = I$. Instead of solving for 26 unknown w_k coefficients, we propose to apply a constraint on the solution such that neighbor voxel pairs of symmetric directions will have the same w_k values. This allows us to reduce the problem to 7 unknown variables. We assign the weight w_x to the direction $k = (\pm1, 0, 0)$ and w_y to $k = (0, \pm1, 0)$. Next, we assign w_{xy} to the set of directions $k = (\pm1, \pm1, 0)$, so that

$$\sum_{k_1=\pm1,k_2=\pm1,k_3=0} w_{xy} R_{k_1,k_2,k_3} = 4w_{xy}\begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Notice that the matrices sum up to be a diagonal matrix. Similarly, we assign w_z to directions $k = (0, 0, \pm1)$, w_{xz} to directions $k = (\pm1, 0, \pm1)$, w_{yz} to directions $k = (0, \pm1, \pm1)$, and w_{xyz} to the directions $k = (\pm1, \pm1, \pm1)$

Substituting (13) into (12), we can verify that by setting the coefficients in the symmetric way, the non-diagonal entries will all cancel out. Therefore, we only need to consider the constraints of the three diagonal entries, which give us the following equations

$$\begin{cases} 2w_x + 4\times\frac{1}{2}w_{xy} + 4w_{xz}\cos^2\theta + 8\times\frac{1}{2}w_{xyz}\cos^2\gamma = 1 \\ 2w_y + 4\times\frac{1}{2}w_{xy} + 4w_{yz}\cos^2\theta + 8\times\frac{1}{2}w_{xyz}\cos^2\gamma = 1 \\ 2w_z + 4w_{xz}\sin^2\theta + 4w_{yz}\sin^2\theta + 8w_{xyz}\sin^2\gamma = 1 \end{cases} \tag{14}$$

The above equations still do not have a unique solution. Here we propose to formulate a cost function that minimizes the total energy of the w_k coefficients subject to (14) and a non-negativity constraint.

$$\begin{aligned} w^* &= \arg\min_{w_k\geq0}\sum_{k\in W} w_k^2 \\ &= \arg\min_{w_k\geq0}\{2w_x^2 + 2w_y^2 + 2w_z^2 \\ &\quad 4w_{xy}^2 + 4w_{xz}^2 + 4w_{yz}^2 + 8w_{xyz}^2\} \end{aligned} \tag{15}$$

The above constrained optimization problem yields the following analytical solution

$$\begin{bmatrix} w_x \\ w_y \\ w_{xy} \\ w_z \\ w_{xz} \\ w_{yz} \\ w_{xyz} \end{bmatrix} = \begin{bmatrix} 1/12\,\frac{17\beta^4+4+24\beta^6+9\beta^8}{14\beta^4+4+8\beta^2+10\beta^6+3\beta^8} \\ 1/12\,\frac{17\beta^4+4+24\beta^6+9\beta^8}{14\beta^4+4+8\beta^2+10\beta^6+3\beta^8} \\ 1/12\,\frac{17\beta^4+4+24\beta^6+9\beta^8}{14\beta^4+4+8\beta^2+10\beta^6+3\beta^8} \\ 1/6\,\frac{12\beta^2+2\beta^4+12+\beta^8}{14\beta^4+4+8\beta^2+10\beta^6+3\beta^8} \\ 1/12\,\frac{(1+\beta^2)(5\beta^4+4+16\beta^2+2\beta^6)}{14\beta^4+4+8\beta^2+10\beta^6+3\beta^8} \\ 1/6\,\frac{(2+\beta^2)(5\beta^4+1+2\beta^2+\beta^6)}{14\beta^4+4+8\beta^2+10\beta^6+3\beta^8} \end{bmatrix}, \tag{16}$$

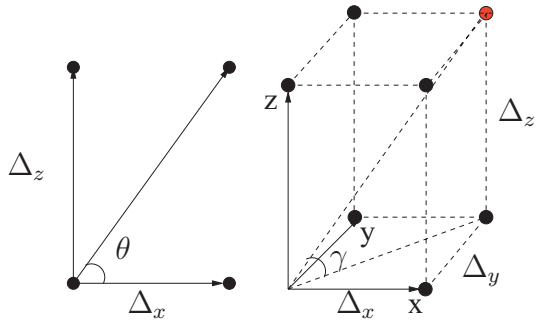
where $\beta = \frac{\Delta_z}{\Delta_x}$. Fig. 3 shows w_k as a function of β.

IV. Experimental Results

In this section, we apply the proposed regularization model to 3D CT reconstruction problems. In computed tomography, the resolution and noise properties are mainly determined by system geometry and scan techniques. Therefore, this paper does not focus on achieving specific image quality properties such as uniform or isotropic resolution in the general case for any scanning geometry. Instead, we would like to demonstrate that with the proposed model, image quality is less sensitive to voxel size changes compared to the baseline methods.

The methods in comparison are the proposed regularization model, and two baseline methods. In the first baseline method, we choose $\alpha = \frac{1}{2|J|}$ and $b_k = \frac{1}{||Jk||^2}$, that is, we set w_k to be a constant in all directions. In the second baseline method, we choose $b_k = \frac{1}{||Jk||^2}$ and α to be a constant for all voxel sizes. In both baseline methods, we adjust the overall regularization strength to match the proposed model at $dx = dy = dz = 0.625mm$. At this point, all three models yield identical regularization parameters.

To measure image quality, we simulated a 3D digital phantom containing an array of high contrast point sources as shown in Fig. 4. We test the proposed and baseline algorithms with various voxel sizes $dx = dy$ at fixed $dz = 0.625$mm. In Fig. 5, we measure the 50% MTF (lp/cm) in $x-y$ plane, full width half maxium (FWHM) in mm along the z axis and noise standard deviation (HU) in the uniform portion of the phantom. Considering the variations of these image quality metrics when the voxel size changes, we notice that the proposed method is less sensitive to the voxel size change compared to the baseline methods. Furthermore, the baseline 1 is also less sensitive compared to the baseline 2, since it is closer to the proposed model.



(a) (b)

Fig. 2. This figure illustrates the geometry of the voxel neighborhood on a rectangular grid. In (a), we show the voxels in the $x-z$ plane. The angle between the diagonal voxel and the x axis is denoted as θ. In (b), we show the voxels in the first octant, in which γ is the angle between the corner voxel shown in red a

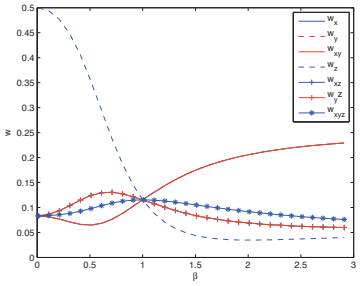


Fig. 3. This figure shows w_k as a function of voxel aspect ratio parameter β

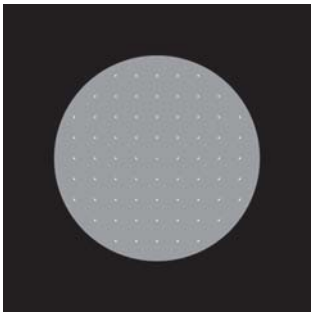


Fig. 4. The figure shows the digital phantom used in the experiment.

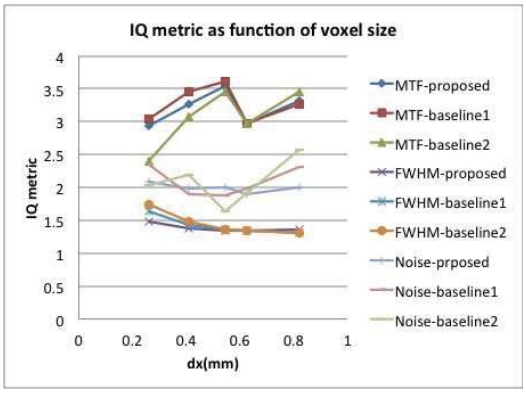


Fig. 5. The figure shows the IQ metrics when $dx = dy$ changes at fixed dz. The proposed method appears less sensitive to such changes compared to the baseline methods.

V. Discussion

The theory proposed in this paper can be useful in many reconstruction as well as denoising problems. In our example, we change the pixel size independently of slice thickness, which yields a non-isotropic voxel. Without modeling non-isotropic sampling in $U(x)$, one could end up with over-regularized results along certain directions while other directions are under-regularized. In 4D reconstructions (3D + temporal), the voxel size and temporal sampling rate can also be adjusted independently. The proposed theory can be used to yield a consistent design balancing spatial and temporal resolution. Although our derivation assumes a periodic lattice, the theory can also be extended to non-periodic lattices where b_k needs to be computed for each location.

References

[1] S. Geman and D. McClure, "Statistical methods for tomographic image reconstruction," *Bull. Int. Stat. Inst.*, vol. LII-4, pp. 5–21, 1987.

[2] E. Ü. Mumcuoğlu, R. Leahy, S. Cherry, and Z. Zhou, "Fast gradient-based methods for Bayesian reconstruction of transmission and emission pet images," *IEEE Trans. on Medical Imaging*, vol. 13, no. 4, pp. 687–701, December 1994.

[3] C. Bouman and K. Sauer, "A unified approach to statistical tomography using coordinate descent optimization," *IEEE Trans. on Image Processing*, vol. 5, no. 3, pp. 480–492, March 1996.

[4] P. Joesph, "An improved algorithm for reprojecting rays through pixel images," *IEEE Trans. on Medical Imaging*, vol. 1, pp. 192–196, 1982.

[5] R. Lewitt, "Multidimensional digital image representations using generalized kaiser-bessel window functions," *J. Opt. Soc. Am. A*, vol. 7(10), pp. 1834–1846, October 1990.

[6] B. DeMan and S. Basu, "Distance-driven projection and backprojection in three-dimensions," *Physics in Medicine and Biology*, vol. 49, pp. 2463–2475, 2004.

[7] A.Ziegler, Th.Köhler, and R.Proksa, "Efficient projection and backprojection scheme for spherically symmetric basis functions in divergent beam geometry," *Med. Phys.*, vol. 33, no. 12, pp. 4653–4663, 2006.

[8] D. Geman and G. Reynolds, "Constrained restoration and the recovery of discontinuities," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 14, no. 3, pp. 367–383, March 1992.

[9] C. Bouman and K. Sauer, "A generalized Gaussian image model for edge-preserving MAP estimation," *IEEE Trans. on Image Processing*, vol. 2, no. 3, pp. 296–310, July 1993.

[10] V. Panin, G. Zeng, and G. Gullberg, "Total variation regulated EM algorithm," *IEEE Trans. on Nuclear Science*, vol. 46, no. 6, pp. 2202–2210, December 1999.

[11] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. PAMI-4, pp. 721–741, November 1984.

[12] J.-B. Thibault, K. D. Sauer, C. A. Bouman, and J. Hsieh, "A three-dimensional statistical approach to improved image quality for multislice helical CT," *Med. Phys.*, vol. 34, no. 11, pp. 4526–4544, 2007.

[13] S. Oh, A. B. Milstein, C. A. Bouman, and K. J. Webb, "A general framework for nonlinear multigrid inversion," *IEEE Trans. on Image Processing*, vol. 14, no. 10, pp. 125–140, January 2005.