5-10-2011

Non-negative Quadratic Programming Total Variation Regularization for Poisson Vector-Valued Image Restoration

Paul A. Rodriguez

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

SCHOOL OF ENGINEERING
UNIVERSITY OF NEW MEXICO

Non-negative Quadratic Programming Total Variation Regularization for Poisson Vector-Valued Image Restoration

Paul A. Rodriguez*

Department of Electrical and Computer Engineering
The University of New Mexico
Albuquerque, NM 87131

*Dual affiliation with Department of Electrical Engineering
Pontificia Universidad Católica del Perú
Lima, Peru
e-mail: `prodrig@pucp.edu.pe`

UNM Technical Report: EECE-TR-11-0003

Report Date: May 10, 2011

Abstract

We propose a flexible and computationally efficient method to solve the non-homogeneous Poisson (NHP) model for grayscale and color images within the TV framework. The NHP model is relevant to image restoration in several applications, such as PET, CT, MRI, etc. The proposed algorithm uses a quadratic approximation of the negative log-likelihood function to pose the original problem as a non-negative quadratic programming problem.

The reconstruction quality of the proposed algorithm outperforms state of the art algorithms for grayscale image restoration corrupted with Poisson noise. Furthermore, it places no prohibitive restriction on the forward operator, and to best of our knowledge, the proposed algorithm is the only one that explicitly includes the NHP model for color images within the TV framework.

Keywords

Total Variation regularization, non-homogeneous Poisson model, Non-negative Quadratic Programming

1 Introduction

Fundamentally, almost every device for image acquisition is a photon counter: Positron Emission Tomography (PET), computed tomography (CT), magnetic resonance imaging (MRI), radiography, CCD cameras, etc. When the count level is low, the resulting image is noisy; such noise is usually modeled via the non-homogeneous Poisson (NHP) model, with the resulting noise being non-additive and pixel-intensity dependent.

There are several methods (e.g. Richardson-Lucy algorithm and its regularized variants, Wavelet methods, Bayesian methods, etc.) to restore (denoise/deconvolve) non-negative grayscale images corrupted with Poisson noise. For the scope of this paper we focus our attention to methods based on Total variation (TV).

The minimization of the TV regularized Poisson log-likelihood cost functional [10, 6, 5, 1, 15, 18, 8, 11]

$$\begin{aligned} T(\mathbf{u}) &= \sum_k (A\mathbf{u})_k - \mathbf{b}_k \cdot \log((A\mathbf{u})_k) + \\ & \frac{\lambda}{q} \left\| \sqrt{\sum_{n \in C} (D_x \mathbf{u}_n)^2 + (D_y \mathbf{u}_n)^2} \right\|_q^q \end{aligned} \tag{1}$$

where $n \in C = \{gray\}$, $q = 1$, $\mathbf{b}$ is the acquired image data (corrupted with Poisson noise), A is a forward linear operator, and $(A\mathbf{u})_m$ is the m-th element of $(A\mathbf{u})$, has been successfully employed to restore non-negative grayscale images corrupted with Poisson noise from a number of applications, such as PET images [10], confocal microscopy images [6], and others [5, 1, 15, 18, 8, 11]. Currently for color images, to best of our knowledge, there is no algorithm that explicitly includes the NHP model within the TV framework.

In this paper we present a computationally efficient algorithm that uses a second order Taylor approximation of the data fidelity term $F(\mathbf{u}) = \sum (A\mathbf{u})_m - \mathbf{b}_m \cdot \log((A\mathbf{u})_m)$ to pose (1) as a Non-negative Quadratic Programming problem (NQP, see [17]) which can be solved with a similar approach as the IRN-NQP algorithm [14, 12]. The resulting algorithm has the following advantages:

- does not involve the solution of a linear system, but rather multiplicative updates only,
- has the ability to handle an arbitrary number of channels in (1), including the scalar (grayscale) and vector-valued (color) images ($C = \{r, g, b\}$ as special cases,
- most of its parameters are automatically adapted to the particular input dataset, and if needed, the norm of the regularization term (q in (1)) can be different than 1 ($0 < q \leq 2$).

2 Technicalities

We represent 2-dimensional images by 1-dimensional vectors: $\mathbf{u}_n$ ($n \in C$) is a 1-dimensional (column) or 1D vector that represents a 2D grayscale image obtained via any ordering (although the most reasonable choices are row-major or column-major) of the image pixel. For $C = \{r, g, b\}$ we have that $\mathbf{u} = [(\mathbf{u}_r)^T (\mathbf{u}_g)^T (\mathbf{u}_b)^T]^T$ is a 1D (column) vector that represents a 2D color image.

The gradient and Hessian of the data fidelity term in (1) $F(u) = \sum_k (A\mathbf{u})_k - \mathbf{b}_k \cdot \log((A\mathbf{u})_k)$ can be easily computed:

$$\nabla F(\mathbf{u}) = A^T \left(\mathbf{1} - \frac{\mathbf{b}}{A\mathbf{u}} \right) \tag{2}$$

$$\nabla^2 F(\mathbf{u}) = A^T \operatorname{diag}\left(\frac{\mathbf{b}}{(A\mathbf{u})^2} \right) A, \tag{3}$$

and its quadratic approximation, $Q_F^{(k)}(\mathbf{u})$, can be written as

$$Q_F^{(k)}(\mathbf{u}) \;=\; F(\mathbf{u}^{(k)}) + \mathbf{g}^{(k)^T}\mathbf{v}^{(k)} + \frac{1}{2}\mathbf{v}^{(k)^T} H^{(k)}\mathbf{v}^{(k)} \tag{4}$$

where $\mathbf{v}^{(k)} = \mathbf{u} - \mathbf{u}^{(k)}$, $\mathbf{g}^{(k)} = \nabla F(\mathbf{u}^{(k)})$ and $H^{(k)} = \nabla^2 F(\mathbf{u}^{(k)})$. Note that (4) remains the same for color images ($n \in C = \{r,g,b\}$) and scalar operations applied to a vector are considered to be applied element-wise, so that, for example, $\mathbf{u} = \mathbf{v}^2 \Rightarrow u[k] = (v[k])^2$ and $\mathbf{u} = \frac{1}{\mathbf{v}} \Rightarrow u[k] = \frac{1}{v[k]}$.

Also, for the case of color images, the linear operator A in (1) is assumed to be decoupled, i.e. A is a diagonal block matrix with elements A_n; if A is coupled (inter-channel blur) due to channel crosstalk, it is possible to reduced it to a diagonal block matrix via a similarity transformation [9].

In the present work $\frac{1}{q}\left\|\sqrt{\sum_{n\in C}(D_x\mathbf{u}_n)^2 + (D_y\mathbf{u}_n)^2}\right\|_q^q$ is the generalization of TV regularization to color images ($n \in C = \{r,g,b\}$) with coupled channels (see [3, Section 9], also used in [4] among others), where we note that $\sqrt{\sum_{n\in C}(D_x\mathbf{u}_n)^2 + (D_y\mathbf{u}_n)^2}$ is the discretization of $|\nabla\mathbf{u}|$ for coupled channels (see [4, eq. (3)]), and D_x and D_y represent the horizontal and vertical discrete derivative operators respectively.

3 The Poisson IRN-NQP Algorithm

In this section we succinctly describe of previous works (within the TV framework) that also use a second order Taylor approximation of the data fidelity term $F(\mathbf{u})$ to solve (1). We continue to summarized the derivation of the Poisson IRN-NQP (Iteratively Reweighted Norm, non-negative quadratic programming) algorithm, where we also provide a brief description of the NQP [17] problem to finally list the Poisson IRN-NQP algorithm.

3.1 Previous Related Work

Within the TV framework, there are several papers [1, 15, 18] that use a second order Taylor approximation of the data fidelity term $F(\mathbf{u})$ to make (1) a tractable problem, but they differ in the constrained optimization algorithm used to carry out the actual minimization and in whether the true Hessian of $F(\mathbf{u})$ or an approximation is used.

We must stress that all the above mentioned algorithms involve the solution of a linear system. Moreover, in [1] the minimization is carried out via a non-negatively constrained, projected quasi-Newton minimization algorithm; $\nabla^2 F(\mathbf{u})$ is used. In [18] a constrained TV algorithm, described in [2], plus the approximation $\nabla^2 F(\mathbf{u}) \approx \eta_k I, \eta_k > 0$ are used. In [15] a Expectation-Maximization TV approach is used; only the denoising problem is addressed.

3.2 Non-negative Quadratic Programming (NQP)

Recently [17] an interesting and quite simple algorithm has been proposed to solve the NQP problem:

$$\min_{\mathbf{v}} \; \frac{1}{2}\mathbf{v}^T\Phi\mathbf{v} + \mathbf{c}^T\mathbf{v} \;\; \text{s.t. } 0 \le \mathbf{v} \le \text{vmax}, \tag{5}$$

where the matrix Φ is assumed to be symmetric and positive defined, and vmax is some positive constant. The multiplicative updates for the NQP are summarized as follows (see [17] for details on derivation and convergence):

$$\Phi^+_{nl} = \begin{cases} \Phi_{nl} & \text{if } \Phi_{nl} > 0 \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad \Phi^-_{nl} = \begin{cases} |\Phi_{nl}| & \text{if } \Phi_{nl} < 0 \\ 0 & \text{otherwise,} \end{cases}$$

$$\mathbf{v}^{(k+1)} = \min\left\{\mathbf{v}^{(k)}\left[\frac{-\mathbf{c}+\sqrt{\mathbf{c}^2+\boldsymbol{\upsilon}^{(k)}\boldsymbol{\nu}^{(k)}}}{2\boldsymbol{\upsilon}^{(k)}}\right], \text{vmax}\right\} \tag{6}$$

where $\boldsymbol{\upsilon}^{(k)} = \Phi^{+}\mathbf{v}^{(k)}$, $\boldsymbol{\nu}^{(k)} = \Phi^{-}\mathbf{v}^{(k)}$ and all algebraic operations in (6) are to be carried out element wise. The NQP is quite efficient and has been used to solve interesting problems such as statistical learning [17] among others.

3.3 The Poisson IRN-NQP Algorithm

Our aim is to express (1) as a quadratic functional in order to solve a NQP problem. First we note that after algebraic manipulation, $Q_F^{(k)}(\mathbf{u})$ (see (4)) can be written as

$$Q_F^{(k)}(\mathbf{u}) \quad = \quad \frac{1}{2}\mathbf{u}^T H^{(k)}\mathbf{u} + \mathbf{c}^{(k)}\mathbf{u} + \zeta_F, \tag{7}$$

where ζ_F is a constant with respect to $\mathbf{u}$, and $\mathbf{c}^{(k)} = \left(g^{(k)} - H^{(k)^T}\mathbf{u}^{(k)}\right) = A^T\left(\mathbf{1} - 2\frac{\mathbf{b}}{A\mathbf{u}^{(k)}}\right)$.

While not derived so here, the ℓ^q norm of the regularization term in (1) can be represented by a equivalent weighted ℓ^2 norm (see [14, 12] for details):

$$R^{(k)}(\mathbf{u}) = \frac{\lambda}{2}\left\|W_R^{(k)^{1/2}}D\mathbf{u}\right\|_2^2 + \zeta_R = \frac{\lambda}{2}\mathbf{u}^T D^T W_R^{(k)} D\mathbf{u}, \tag{8}$$

where ζ_R is a constant with respect to $\mathbf{u}$, I_N is a $N\times N$ identity matrix, $\otimes$ is the Kronecker product, $C = \{gray\}$, $N = 1$ or $C = \{r,g,b\}, N = 3$ and

$$D = I_N \otimes [Dx^T Dy^T]^T \quad W_R^{(k)} = I_{2N}\otimes \Omega^{(k)}, \tag{9}$$

$$\Omega^{(k)} = \mathrm{diag}\left(\tau_{R,\varepsilon_R}\left(\sum_{n\in C}(D_x\mathbf{u}_n^{(k)})^2 + (D_y\mathbf{u}_n^{(k)})^2\right)\right). \tag{10}$$

Following a common strategy in IRLS type algorithms [16], the function

$$\tau_{R,\varepsilon_R}(x) = \begin{cases} |x|^{(q-2)/2} & \text{if } |x| > \varepsilon_R \\ \varepsilon_R^{(q-2)/2} & \text{if } |x| \le \varepsilon_R, \end{cases} \tag{11}$$

is defined to avoid numerical problems when $q<2$ and $\sum_{n\in C}(D_x\mathbf{u}_n^{(k)})^2 + (D_y\mathbf{u}_n^{(k)})^2$ has zero-valued components.

Combining (7) and (8) we can write the quadratic approximation of (1) as (the constant terms are left out)

$$T^{(k)}(\mathbf{u}) = \frac{1}{2}\mathbf{u}^T\left(H^{(k)} + \lambda D^T W_R^{(k)} D\right)\mathbf{u} + \mathbf{c}^{(k)^T}\mathbf{u}, \tag{12}$$

to finally note that by using $\Phi^{(k)} = H^{(k)} + \lambda D^T W_R^{(k)} D$ and $\mathbf{c}^{(k)} = A^T\left(\mathbf{1} - 2\frac{\mathbf{b}}{A\mathbf{u}^{(k)}}\right)$ we can iteratively solve (1) via (6).

It is easy to check that $\Phi^{(k)}$ is symmetric and positive define. Furthermore, we note that the proposed algorithm does not involve any matrix inversion: the fraction in the term $\mathbf{c}^{(k)} = A^T\left(\mathbf{1} - 2\frac{\mathbf{b}}{A\mathbf{u}^{(k)}}\right)$ is a point-wise division (a check for zero must be performed to avoid numerical problems) as well as in the term $H^{(k)} = A^T\,\mathrm{diag}\left(\frac{\mathbf{b}}{(A\mathbf{u})^2}\right)A$; also, there is no prohibitive restriction for the forward operator A (e.g. $A_{n,k}\ge 0$, see [8] and references therein). Two other key aspects of the Poisson IRN-NQP algorithm is that it can auto-adapt the threshold value ε_R and that it includes the *NQP tolerance* ($\varepsilon_{NQP}^{(k)}$), used to terminate the inner loop in Algorithm 1 (see [12] for details). Experimentally, $\alpha\in[1\,..\,0.5]$, $\gamma\in[$1e-3 .. 5e-1$]$, and $L=35$ give a good compromise between computational and reconstruction performance.

4 Experimental Results

For the case of grayscale images, we compared the Poisson IRN-NQP algorithm with two state of the art methods: PIDAL-FA [8] and DFS [7]. We use the mean absolute error (MAE = $\|\hat{\mathbf{u}} - \mathbf{u}\|_1/n$) as the reconstruction quality metric to match the results presented in [8, 7]. For vector-valued (color) images, to best of our knowledge, there is no other algorithm that explicitly includes the NHP model within the TV framework; due to space limitations we only present results based on the Poisson IRN-NQP algorithm for this case.

All simulations have been carried out using Matlab-only code on a 1.73GHz Intel core i7 CPU (L2: 6144K, RAM: 6G). Results corresponding to the Poisson IRN-NQP algorithm presented here may be reproduced using the the NUMIPAD (v. 0.30) distribution [13], an implementation of IRN and related algorithms.

The Cameraman and Peppers images (Fig. 4a and 4b respectively) were scaled to a maximum value $M \in \{5, 30, 100, 255\}$, then blurred by a 7×7 uniform filter and by a 7×7 out-of-focus kernel (2D pill-box filter) respectively and then Poisson noise was finally added; this matches experiment VI.C setup of [8] for the Cameraman case.

(a) (b)

Figure 1: Input test images: (a) Cameraman (grayscale 256×256), (b) Peppers (color 512×512).

Image		Cameraman			Peppers
		MAE			MAE
		Poisson-IRN-NQP	PIDAL-FA [8]$^{(*)}$	DFS [7]$^{(*)}$	Poisson-IRN-NQP
M	5	0.25	0.26	0.35	0.55
	30	1.21	1.22	1.47	2.56
	100	3.59	3.63	4.31	7.57
	255	8.34	8.45	10.26	17.53

Table I: Ten-trial average Cameraman's MAE for Poisson-IRN-NQP (8 outer loops) and the PIDAL-FA [8] and DFS [7] algorithms, and Peppers' MAE for Poisson-IRN-NQP. Cameraman and Peppers deconvolution elapsed 11.6s and 190s in average. $^{(*)}$ Results are taken from [8, Table II].

In Table I we show the ten-trial average Cameraman's MAE for the Poisson IRN-NQP, PIDAL-FA [8] and DFS [7] algorithms and the ten-trial average Peppers' MAE for the Poisson IRN-NQP. The quality reconstruction of the proposed algorithm outperforms DFS algorithm and has a slightly better performance than the PIDAL-FA algorithm. The Poisson IRN-NQP algorithm performed 8 outer loops (with $L = 35$) requiring 11.6s and 190s in average (for all cases), for the Cameraman and Peppers deconvolution respectively.

Initialize

$$\mathbf{u}^{(0)} = \mathbf{b}$$

for $k = 0, 1, ...$

$$W_F^{(k)} = \mathrm{diag}\left(\frac{\mathbf{b}}{(A\mathbf{u}^{(k)})^2}\right), \;\; H^{(k)} = A^T W_F^{(k)} A$$

$$\Omega_R^{(k)} = \mathrm{diag}\left(\tau_{R,\varepsilon_R}\left((D_x\mathbf{u}^{(k)})^2 + (D_y\mathbf{u}^{(k)})^2\right)\right)$$

$$W_R^{(k)} = \begin{pmatrix} \Omega_R^{(k)} & 0 \\ 0 & \Omega_R^{(k)} \end{pmatrix}$$

$$\Phi^{(k)} = H^{(k)} + \lambda D^T W_R^{(k)} D, \;\; \mathbf{c}^{(k)} = \text{-}A^T\left(\mathbf{1} - 2\tau_{F,\varepsilon_F}\left(\frac{\mathbf{b}}{A\mathbf{u}^{(k)}}\right)\right)$$

$$\mathbf{u}^{(k,0)} = \mathbf{u}^{(k)}$$

$$\varepsilon_{NQP}^{(k)} = \gamma \cdot \left(\frac{\|\Phi^{(k)}\mathbf{u}^{(k,0)} + \mathbf{c}^{(k)}\|_2}{\|\mathbf{c}^{(k)}\|_2}\right)^\alpha \qquad \text{(NQP tolerance)}$$

for $l = 0, 1, .., \mathrm{L}$

$$\boldsymbol{\upsilon}^{(k,l)} = \Phi^{+(k)}\mathbf{u}^{(k,l)}, \; \boldsymbol{\nu}^{(k,l)} = \Phi^{-(k)}\mathbf{u}^{(k,l)}$$

$$\mathbf{u}^{(k,l+1)} = \min\left\{\mathbf{u}^{(k,l)}\left[\frac{-\mathbf{c}^{(k)} + \sqrt{{\mathbf{c}^{(k)}}^2 + \boldsymbol{\upsilon}^{(k,l)}\boldsymbol{\nu}^{(k,l)}}}{2\boldsymbol{\upsilon}^{(k,l)}}\right], \mathrm{vmax}\right\}$$

if $\left(\frac{\|\Phi^{(k)}\mathbf{u}^{(k,l+1)} + \mathbf{c}^{(k)}\|_2}{\|\mathbf{c}^{(k)}\|_2} < \varepsilon_{NQP}^{(k)}\right)$ **break;**

end

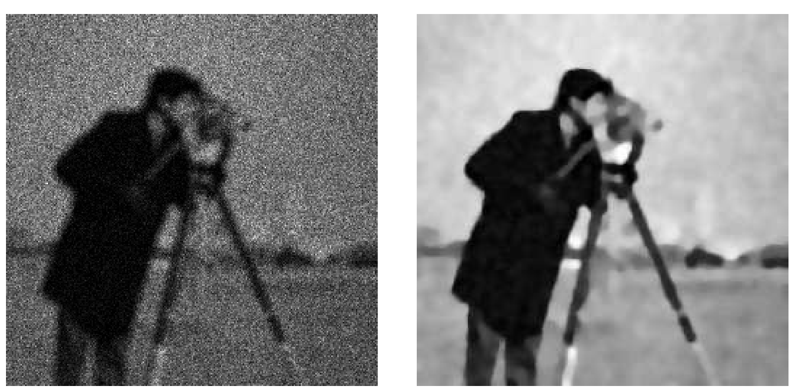
$$\mathbf{u}^{(k+1)} = \mathbf{u}^{(k,l+1)}$$

end

Algorithm 1: Poisson IRN-NQP algorithm.

5 Conclusions

The reconstruction quality of the proposed algorithm outperforms state of the art algorithms [8, 7] for grayscale image deconvolution corrupted with Poisson noise. One of the main features of this recursive algorithm is that it is based on multiplicative updates only and does not involves any matrix inversion, and to best of our knowledge, the proposed algorithm is the only one that explicitly includes the NHP model for color images within the TV framework.



(a) (b)

Figure 2: Blurred (7×7 uniform filter) Cameraman (a) with M=30 (SNR=4.6dB), and deconvolved version (b) via the Poisson-IRN-NQP (SNR=10.4dB, MAE=1.209).

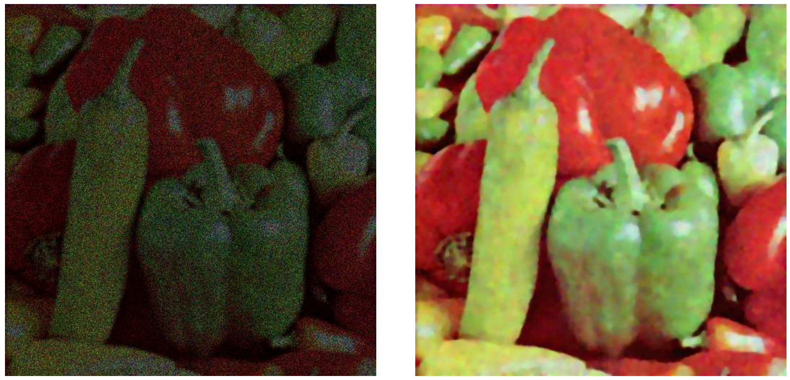

(a) (b)

Figure 3: Blurred (7×7 out-of-focus kernel) Peppers (a) with M=5 (SNR=-1.2dB), and deconvolved version (b) via the Poisson-IRN-NQP (SNR=13.3dB, MAE=0.558).

References

[1] J. Bardsley and A. Luttman. Total variation-penalized Poisson likelihood estimation for ill-posed problems. *Advances in Comp. Math.*, 31:35–59, 2009.

[2] A. Beck and M. Teboulle. Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *IEEE TIP*, 18:2419–2434, Nov. 2009.

[3] A. Bonnet. On the regularity of edges in image segmentation. *Annales de l'institut Henri Poincaré (C) Analyse non linéaire*, 13(4):485–528, 1996.

[4] X. Bresson and T. Chan. Fast dual minimization of the vectorial total variation norm and applications to color image processing. *J. of Inverse Problems and Imaging*, 2(4):455–484, 2008.

[5] R. Chan and K. Chen. Multilevel algorithm for a Poisson noise removal model with total-variation regularization. *Int. J. Comput. Math.*, 84:1183–1198, August 2007.

[6] N. Dey, L. Blanc-Féraud, C. Zimmer, Z. Kam, P. Roux, J.C. Olivo-Marin, and J. Zerubia. Richardson-Lucy algorithm with total variation regularization for 3d confocal microscope deconvolution. *Microscopy Research Technique*, 69:260–266, april 2006.

[7] F. Dupé, J. Fadili, and J. Starck. A proximal iteration for deconvolving Poisson images using sparse representations. *IEEE TIP*, 18(2):310–321, Feb. 2009.

[8] M. Figueiredo and J. Bioucas-Dias. Restoration of Poissonian images using alternating direction optimization. *IEEE TIP*, 19(12):3133 – 3145, 2010.

[9] N. Galatsanos, A. Katsaggelos, T. Chin, and A. Hillery. Least squares restoration of multichannel images. *IEEE TSP*, 39(10):2222–2236, 1991.

[10] E. Jonsson, S. Huang, and T. Chan. Total-variation regularization in positron emission tomography, UCLA CAM Report 98-48, 1998.

[11] I. Rodrigues and J. Sanches. Convex total variation denoising of Poisson fluorescence confocal images with anisotropic filtering. *IEEE TIP*, 20(1):146 – 160, 2011.

[12] P. Rodríguez. A non-negative quadratic programming approach to minimize the generalized vector-valued total variation functional. In *Proceedings of EUSIPCO'10*, Aalborg, Denmark, Aug. 2010.

[13] P. Rodríguez and B. Wohlberg. Numerical methods for inverse problems and adaptive decomposition (NUMIPAD). `http://numipad.sourceforge.net/`.

[14] P. Rodríguez and B. Wohlberg. Efficient minimization method for a generalized total variation functional. *IEEE TIP*, 18(2):322–332, February 2009.

[15] A. Sawatzky, C. Brune, J. Müller, and M. Burger. Total variation processing of images with Poisson statistics. In *CAIP'09*, volume 5702, pages 533–540. 2009.

[16] J. Scales and A. Gersztenkorn. Robust methods in inverse theory. *Inverse Problems*, 4(4):1071–1091, 1988.

[17] F. Sha, Y. Lin, L. Saul, and D. Lee. Multiplicative updates for nonnegative quadratic programming. *Neural Comput.*, 19(8):2004–2031, 2007.

[18] R. Willett, Z. Harmany, and R. Marcia. Poisson image reconstruction with total variation regularization. In *Proceedings of ICIP'10*, Hong Kong, Sep. 2010.