WAVELET REPRESENTATION AND TOTAL VARIATION REGULARIZATION IN EMISSION TOMOGRAPHY

Pavel Kisilev, Michael Zibulevsky, and Yehoshua Y. Zeevi

Dept. of Electrical Engineering, Technion, Israel Institute of Technology, Haifa 32000, Israel

ABSTRACT

A classical technique for reconstruction of Emission Tomography (ET) images from measured projections is based on the maximum likelihood (ML) estimation, achieved with the Expectation Maximization (EM) algorithm. We incorporate the wavelet transform (WT) and total variation (TV) based penalties into the ML framework, and compare performance of the EM algorithm and the recently proposed conjugate barrier (CB) algorithm. Using the WT- and TV-based penalties allows one to embed regularization procedures into the iterative process. In the case of the WT-based penalty, we impose a subset of wavelet coefficients with a desired resolution on the objective function. It appears that the CB algorithm outperforms substantially the EM algorithm in penalized reconstruction. Properties of the optimization algorithms along with WT- and TV-based regularization are demonstrated on image reconstructions of a synthetic brain phantom, and the quality of reconstruction is compared with standard methods.

1. INTRODUCTION

ET is a medical imaging technique that enables one to quantify a distribution of radioactivity within the body, and, as such, it is useful in detection and identification of pathological tissue. In this technique, radioactive tracers, injected into the body of a patient, emit photons, which are detected in distinct detector pairs, or bins. By counting the number of photons detected in the various bins, one measures the projection of the tracer distribution at different angles. A classical technique for the reconstruction of 2D and 3D ET images from measured projections is based on the maximum likelihood (ML) framework [1]. Utilizing particular properties of the Poisson process leads to the Expectation Maximization (EM) algorithm for ET reconstruction [2]. This algorithm provides reliable reconstruction results with high resolution. Alternatively, the ML reconstruction can be performed by the recently proposed conjugate barrier (CB) algorithm, which has several advantages over the EM.

Furthermore, in practice, it is desirable to carry out reconstruction on low statistics (i.e. noisy data). Under these circumstances, the maximum likelihood estimate at highest resolution contains high frequency noise even though the original image is known to be relatively smooth. Therefore, reconstruction of images from their projections requires some kind of regularization that usually represents a trade-off between accuracy and resolution.

Ideally, a lower resolution reconstruction should be applied to regions with no edges. On the other hand, keeping higher resolution components preserves local features in the reconstructed image. This provides desired regularization, so that a trade-off between increasing resolution and noise suppression is achieved. Such a strategy requires some prior knowledge of edges' location, which can be obtained from another imaging procedure (e.g. X-Ray tomography). In [5], the authors assumed that this prior information is available, and used this information to build a penalty template in the wavelet domain. In practice, such prior knowledge is rarely available. It was shown recently that the TV method appears to be one of the most successful regularization approaches to ill-posed problems (see for example [6] and [7]). The TV penalty represents kind of a weak prior about the object structure. In particular, it assumes that the underlying image contains edges, which is usually the case for medical images.

We utilize the wavelet transform (WT) and the Total Variation (TV) functional in our penalties, and show that they affect reconstruction in similar ways. The desired regularization is accomplished in a natural way; using either the WT- or the TV-based penalties allows one to embed regularisation procedures into the iterative process. This task is accomplished by either 1) penalizing for the lack of sparsity of the gradient of the reconstructed image, in the case of the TV-based penalty, or 2) imposing a new set of parameters on a subset of wavelet coefficients corresponding to desired resolutions, and penalizing for the lack of sparsity only this subset, in the case of the WT-based penalty.

It turns out that the penalized CB algorithm achieves the best trade-off between accuracy and resolution. In particular, it keeps improving the contrast while lowering the noise level with iterations. To our knowledge, such a result was not achieved with any of the existing ET reconstruction techniques. The WT-based penalty produces more natural reconstructed images at the earlier iterations, than the TV-based penalty, while the last provides the best trade-off between accuracy and resolution.

2. ML RECONSTRUCTION OF PET IMAGES

2.1. EM algorithm

L. Shepp and Y. Vardy [2] pioneered the ML image reconstruction in PET by application of the EM algorithm.

Let the total number of photons detected in each bin be $y(b)$, $b = 1,\ldots,B$. Let the body be divided into voxels (or pixels), and the number of photons generated independently within each voxel be $n(v)$, $v = 1,\ldots,V$. Generation of photons in each voxel is described by the Poisson process, characterized by the expected value of photons $\lambda(v)$.

Let $p(v,b)$ denote the probability of the event that a photon emitted from voxel v is detected in bin b, forming a matrix with VxB entries. The probability matrix values depend on various physical factors such as scanner geometry, detector efficiency, and the composition of the body being scanned. The issue of computing $p(v,b)$ will be discussed later.

The log-likelihood function for the measurements $y(b)$ is given by

$$L(\lambda) = \sum_{v} \lambda(v)p(v) - \sum_{b} y(b) \log \sum_{v} \lambda(v)p(v,b), \qquad (1)$$

where $\boldsymbol{\lambda} = \{\lambda(v),\ v = 1,\ldots,V\}$ is the set of unknown parameters, and $p(v) = \sum_{v} p(v,b)$ is the probability that an emission from v is detected.

To solve (1), the EM algorithm was applied to the PET reconstruction problem leading to the following formula

$$\lambda^{k+1}(v) = \frac{\lambda^{k}(v)}{p(v)} \sum_{b=1}^{B} \frac{y(b)p(v,b)}{\sum_{v'=1}^{V} \lambda^{k}(v')p(v',b)}, \quad \forall v \qquad (2)$$

for iteratively approximating a maximizer of $L(\boldsymbol{\lambda})$, where k denotes the k-th iteration.

2.2. Conjugate Barrier (CB) algorithm

The CB algorithm, recently proposed by Ben-Tal and Nemirovski [2], belongs to the general class of the Gradient Descent algorithms.

Let the function h be defined as:

$$h(\lambda) = \begin{cases} \|\lambda\|_p, & \lambda \in \Delta \\ +\infty, & \lambda \notin \Delta, \end{cases}$$

where $\|\cdot\|_p$ denotes the l_p norm, and Δ is the domain of valid values of λ. In our setting, wherein λ are the values of normalized image intensity

$$\Delta \equiv \{\lambda \in R^n \mid \lambda \geq 0,\ \Sigma_1^V \lambda(v) = \rho\},$$

where ρ is the total number of detector counts.

Let h^* be the so-called conjugate function of h:

$$h^*(\xi) = \sup_{\lambda \in \Delta} [\xi^T \lambda - h(\lambda)],$$

where ξ is called the conjugate image, so that $\xi, \lambda \in R^V$. The initial value ξ^0 can be initialized arbitrary (we use a matrix of ones).

The following two iterative steps summarize the CB algorithm:

$$\begin{aligned} &\text{step 1:} && \lambda^{k+1} = \nabla h^*(\xi^{k+1}) \\ &\text{step 2:} && \xi^{k+1} = \xi^k - \gamma^k \nabla L(\lambda^k), \\ &\text{where} && \gamma^k = \frac{\gamma^0}{\sqrt{k}} \left\|\nabla L(\lambda^k)\right\|_\infty \end{aligned} \qquad (3)$$

where $\nabla L(\lambda^k)$ is the gradient of the Log-likelihood function at the k-th iteration, γ^k is a positive step size at the k-th iteration, γ^0 is a small positive constant and values of λ^0 are initialized with the value $1/V$. (For more details see [2]).

The CB algorithm has several advantages over the EM: 1) its ordered sets version always converges 2) the rate of convergence is known and independent of the dimension of the problem, 3) the error bound is of order $O\left(\sqrt{\log V / K}\right)$, where K is the number of iterations. In addition, its computational cost is comparable to the EM.

3. TOTAL VARIATION AND WAVELET PENALTIES

As was mentioned before, natural PET data are usually very noisy due to a short acquisition time and various scatter effects. Exact minimization of the log-likelihood function of such noisy data leads to a very noisy reconstructed image. In such cases, various penalty functions reflecting smoothness of the noise-free image or other prior information are used in order to improve quality of reconstruction. In this case, penalized log-likelihood function takes the form

$$L_p(\lambda) = \sum_{v} \lambda(v)p(v) - \sum_{b} y(b) \log \sum_{v} \lambda(v)p(v,b) + \mu \mathrm{H}(\lambda), \qquad (4)$$

where $\mathrm{H}(\boldsymbol{\lambda})$ is a penalty function and μ is its weight parameter, which can be chosen based on the estimated signal to noise ratio (in this paper, we choose it experimentally).

The gradient of the penalized log-likelihood function is:

$$\nabla L_p(\lambda(v)) = p(v) - \sum_{b=1}^{B} \frac{y(b)p(v,b)}{\sum_{v'=1}^{V} \lambda(v')p(v',b)} + \mu \nabla \mathrm{H}(\lambda) \qquad (5)$$

In the case of the CB algorithm, the update of the conjugate image in the second step in (3) is performed according to the gradient of the penalized log-likelihood function (above).

The corresponding iterative formula for the penalized EM algorithm is

$$\lambda^{k+1}(v) = \frac{\lambda^{k}(v)}{p(v) + \mu \nabla \mathrm{H}(\lambda)} \sum_{b=1}^{B} \frac{y(b)p(v,b)}{\sum_{v'=1}^{V} \lambda^{k}(v')p(v',b)}, \quad \forall v \qquad (6)$$

Several penalty functions were proposed in the literature (see for example [3], [4]). These are usually applied in the original image domain. In this work we propose a new method that utilizes penalty function, defined in the wavelet domain. Our penalty function is intimately related to the Total Variation method.

3.1 Total variation penalty

In our study, we use the following formula for the TV penalty $\mathrm{T}(\lambda) = \Sigma_v \left|\nabla \lambda(v)\right|$. Since the expression for the TV penalty (above) is non-differentiable in locations where $\left|\nabla \lambda(v)\right| = 0$, we replace it with the approximation of the norm of gradient for a 2D image $\lambda(v)$:

$$\mathrm{H}_{TV}(\lambda(v)) = \sqrt{\lambda_x^2(v) + \lambda_y^2(v) + \eta},$$

where $\lambda_x(v)$ and $\lambda_y(v)$ are derivatives in directions x and y, wherein the index v in a discrete 2D case is a pair (i,j), i.e. pixel's coordinates, and η is the parameter which controls the smoothness of the penalty. The above approximation is crucial for smooth optimization methods. The CB algorithm is designed for non-smooth optimization, and it can work even for η=0.

It can be shown [11] that the gradient of H_{TV} is

$$\begin{aligned} \nabla_{\lambda(i,j)} \mathrm{H}_{TV} &= -\left(\frac{\partial \mathrm{H}_{TV}}{\partial \lambda_x(i,j)} - \frac{\partial \mathrm{H}_{TV}}{\partial \lambda_x(i-1,j)}\right) \\ &\quad -\left(\frac{\partial \mathrm{H}_{TV}}{\partial \lambda_y(i,j)} - \frac{\partial \mathrm{H}_{TV}}{\partial \lambda_y(i,j-1)}\right) \end{aligned}. \qquad (7)$$

3.2 Wavelet-based penalty

Let $\boldsymbol{\Phi}$ be a matrix constructed from orthonormal wavelet basis vectors, and let the discrete image $\boldsymbol{\lambda}$ be represented by this basis as follows:

$$\lambda = \Phi\alpha,$$

$$\alpha = \Phi^T\lambda.$$

(For detailed discussion of the wavelet transform, see [10]). Suppose that the wavelet coefficients $\boldsymbol{\alpha}$ of the image are known *apriori* to be sparse, i.e. only a small (unknown) part of them significantly differs from zero. Under this assumption, the sum of the absolute values of the coefficients (i.e. the l_1 norm) represents a natural penalty, which forces the coefficients of the reconstructed image to become sparse. Such a penalty is very popular in denoising and compression problems (see for example [8]). Suppose further that only a subset of the coefficients is known to be sparse. For example, for 2D images, the detail coefficients usually are sparse, while approximation ones are quite 'dense'. In this case, it is reasonable to penalize only detail coefficients, or even only a subset of the detail coefficients at high resolutions.

Since edges 'live' at high resolutions, in our experiments we penalize the detail coefficients of the WT at the highest resolutions by constructing the penalty:

$$\mathrm{H}_{WT} = \sum_{s\in S} |c_s|, \tag{8}$$

where S is the subset of the coefficients at the desired resolutions. Let $\boldsymbol{c} = \{c_s\}_{s\in S}$, and its norm is approximated, as before, by $|\boldsymbol{c}| = \sqrt{\boldsymbol{c}^2 + \zeta}$ with regularization parameter ζ. Taking into account that $\boldsymbol{c} = \Phi_S\lambda$, where matrix Φ_S is constructed from the wavelet basis vectors indexed on S, the gradient of the wavelet-based penalty can be written as

$$\nabla_\lambda \mathrm{H}_{WT} = \Phi_S^* \tilde{\boldsymbol{c}},$$

$$\tilde{\boldsymbol{c}} = \left[c_s / \sqrt{c_s^2 + \zeta}\right]_{s\in S},$$

where Φ_S^* is the adjoint of Φ_S (which coincides with Φ_S^T in the case of orthonormal wavelets).

In the case of the Haar wavelet basis, calculation of scalar products of the image with basis functions at the highest resolution is essentially equivalent to the calculation of gradients of the image. Therefore, the TV-based penalty can be considered as kind a WT penalty, in a particular case wherein the Haar coefficients at the highest resolution are penalized. Both, the sum of absolute values of corresponding subset of wavelet coefficients and the sum of absolute values of gradients, represent a measure of sparseness.

4. EXPERIMENTAL RESULTS

We carried out tests with the Shepp-Logan phantom; a model used in tomography for evaluating properties of reconstruction algorithms. The phantom was discretized into a 128x128 image. We slightly modified it by adding a hot spot (Figure 1), which we used for calculation of the contrast and noise suppression properties of reconstruction algorithms. Projection data were simulated as follows: we applied the radon transform to the phantom, using 60 angular and 185 radial samples of projections. These projection data were used as a mean rate of a Poisson process. Random samples of projection data were generated according to the above Poisson process, arriving at overall $1.2e^5$ detector counts.

We use the following parameters in order to determine which algorithm provides the best trade-off between contrast and noise control. The coefficient of variation (CV) is defined as the ratio of the standard deviation to the mean-value of the image over some region of interest (ROI). The contrast recovery (CR) for hot lesions in a cold background is defined as

$$CR_{hot} = \frac{H / C - 1}{H_{true} / C_{true} - 1},$$

where C and H are means taken over cold and hot ROI, respectively, and H_{true} / C_{true} is the real ratio of hot lesion to the background in the phantom.

To illustrate the contrast improvement and noise suppression properties of plain, TV- and WT-based penalized versions of EM and CB algorithms (we will refer to them as EMTV, EMWT, CBTV, CBWT, respectively), we compare their plots of CV versus CR with those characteristics of plain EM and CB algorithms (Figure 2). Each algorithm was iterated 50 times, with the penalty parameter $\mu = 0.005$. Generally speaking, when comparing two such curves, the lower curve achieves a better contrast-to-noise-intensity trade-off. It is clear that, when comparing plain versions of the algorithms, the CB has no advantages over the EM. Moreover, the EM algorithm achieves the same contrast as the CB at the earlier iteration. In contrast, when penalties are applied, the CB outperforms the EM by achieving a much better contrast-to-noise-intensity trade-off, although the CB versions achieve the same contrast as the corresponding EM versions slightly later. TV-based penalized versions of EM and CB algorithms outperform slightly the WT-based ones. The remarkable result is that, in the case of the CBTV algorithm, the curve is monotonically decreasing. This means that both, noise suppression and contrast, are improved with iterations.

In Figure 3, we show examples of images reconstructed by CB, CBWT and CBTV. The TV-based penalty provides a better contrast to noise trade-off than the WT-based one, but the reconstructed image looks more natural and artefacts-free when the WT-based penalty is applied. More experimental results can be found in [11].

5. CONLUSIONS

Numerical results and comparisons, concerning convergence properties and the quality of reconstruction of the proposed WT- and TV-based penalized algorithms versus plain algorithms indicate that using either WT or TV penalty significantly improves the contrast-to-noise trade-off. Penalties are in particular useful when they are applied to the CB algorithm, which outperforms the EM for all kinds of penalties and parameters. The combination of the CB algorithm with the TV penalty achieves the best contrast to noise trade-off, and, most importantly, the CBTV algorithm improves the contrast *and* suppresses noise at the same time, monotonically with increasing number of iterations. Our current research is concentrated on improving the performance of the WT-based algorithm, which, we feel, has more flexibility than the TV-based one.

6. REFERENCES

[1] L. Shepp, Y. Vardi, (1982). Maximum likelihood reconstruction for emission tomography. *IEEE Med. Imaging, 1, 113-122.*

[2] Ben-Tal, A. and Nemirovski, A. (1999). “The conjugate barrier method for non-smooth convex optimization”, TR #5/99, Oct. 1999, Minerva Optimization Center, Technion.

[3] J. A. Fessler, N. H. Clinthorne, W. L. Rogers. “Regularized emission image reconstruction using imperfect side information”. Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf., vol. 3, pp. 1991-1995, 1991.

[4] D. F. Yu, J. A. Fessler. “Three-dimensional non-local edge-preserving regularization for PET transmission reconstruction”. Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf., 2000. To appear. (oral)

[5] P. Kisilev, M. Jacobson, Y.Y. Zeevi. Utilizing Wavelet Transform for ML Reconstruction in Positron Emission Tomography. CCIT Report #522, Jan. 2001.

[6] T. F. Chan, C.K. Wong. Total Variation Blind Deconvolution, *IEEE Transactions on Image Processing, to appear.*

[7] E. Jonsson, S. Huang and T. Chan, Total Variation Regularization in Positron Emission Tomography, TR 9848, November 1998.

[8] S. Chen, D. Donoho, and M. Saunders. Atomic decomposition by basis pursuit. 1996. http:/www-stat.stanford.edu/~donoho/Reports/.

[9] P.Kisilev, M. Zibulevsky, and Y.Y. Zeevi, B. Pearlmutter. Multiresolution Framework for Blind Source Separation. CCIT Report #317, June 2000.

[10] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press,1998.

[11] P. Kisilev, M. Zibulevsky, Y.Y. Zeevi. Total Variation and Wavelet Regularization in Emission Tomography. CCIT Report #531, Jan. 2001.

Acknowledgement. Research supported in part by the Ollendorff Minerva Center, by the Fund for Promotion of Research at the Technion and by the Israeli Ministry of Science.

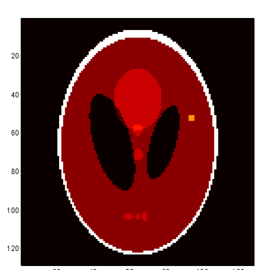

Figure 1. Modified Shepp-Logan phantom.

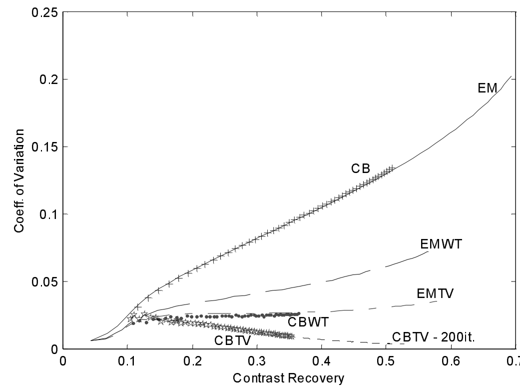

Figure 2. Coefficient of Variation versus Contrast Recovery; 50 iterations.

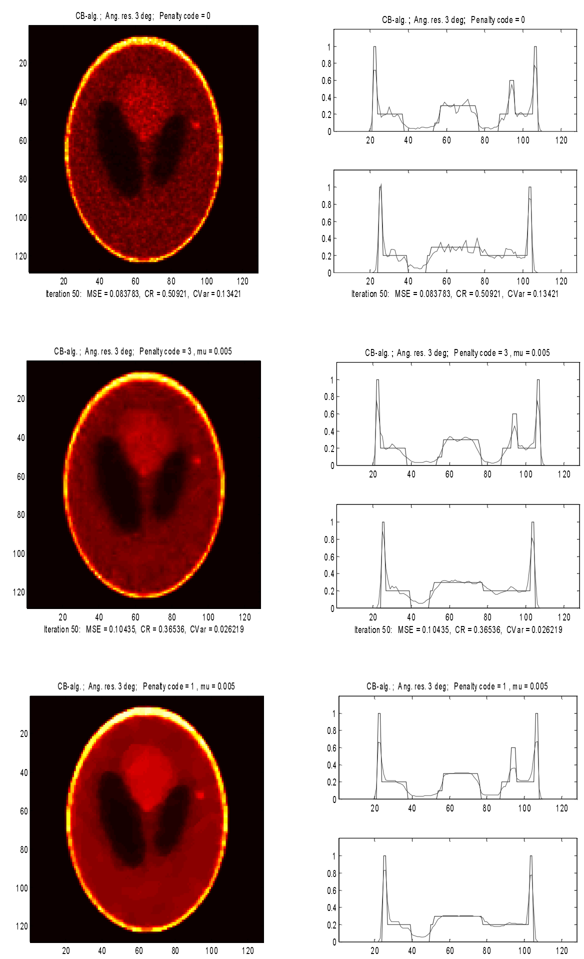

Figure 3. Image and its two slices reconstructed by the CB (upper raw), CBWT (middle raw), and CBTV (lower raw).