

Augmented Lagrangian Methods for Penalized Likelihood Reconstruction in Emission Tomography

Daniel J. Lingenfelter and Jeffrey A. Fessler

Abstract—In emission tomography, the Poisson statistics of the observations make penalized-likelihood reconstruction with an ℓ_1 penalty more difficult than in the case where the observed data is Gaussian. Previously proposed methods for enforcing sparsity of the reconstructed image with respect to some transform use approximations of the ℓ_1 norm. These approximations facilitate the derivation of monotonic algorithms using optimization transfer methods. Recently, augmented Lagrangian methods have been applied to restoration of images corrupted by Poisson noise without requiring approximations to the ℓ_1 norm. This work extends previously derived augmented Lagrangian-based algorithms to penalized likelihood reconstruction for emission tomography with an exact ℓ_1 penalty. We compare the proposed algorithm to an incremental optimization transfer algorithm that performs penalized-likelihood reconstruction with a hyperbolic approximation to the ℓ_1 penalty. The results show that the proposed algorithm reduces the cost function nearly as quickly as the incremental optimization transfer algorithm. The potential advantage of the proposed method is that it solves the ℓ_1 penalized-likelihood problem exactly.

I. INTRODUCTION

Images are sometimes sparse in some basis or frame, which motivates the use of the ℓ_1 norm as a regularizer [1]. When the recorded data in an imaging problem is corrupted with Poisson noise, standard algorithms using an ℓ_2 data-fit term, such as those proposed in [1], do not account for the true statistics of the problem and can lead to suboptimal reconstructed images.

The ℓ_1 regularized (or penalized) likelihood image reconstruction problem for emission tomography with Poisson noise is of the form

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x} \succeq 0} \sum_{i=1}^{n_d} [A\mathbf{x}]_i + r_i - y_i \log([A\mathbf{x}]_i + r_i) + \beta \|C\mathbf{x}\|_1, \quad (1)$$

where n_d is the number of observations, n_p is the number of unknowns, $\hat{\mathbf{x}} \in \mathbb{R}^{n_p}$ is the estimated image, $\mathbf{y} \in \mathbb{R}^{n_d}$ is the observed data, $\mathbf{r} \in \mathbb{R}^{n_d}$ is the vector of known background intensities for each detector element, $C \in \mathbb{R}^{K \times n_p}$ is a linear transform under which the image is expected to be sparse, $A \in \mathbb{R}^{n_d \times n_p}$ is the system matrix, and $[\cdot]_i$ denotes the i th element of a vector.

Since (1) is difficult to minimize due to the unbounded curvature of the ℓ_1 norm, some previous work has focused on approximating the ℓ_1 norm by a smooth function. The regularizer in (1) is of the form

$$R(\mathbf{x}) = \sum_{k=1}^K \psi([C\mathbf{x}]_k), \quad (2)$$

where $\psi(\cdot) = |\cdot|$ for $k = 1, \dots, K$. Previous methods, e.g. [2] use smooth approximations to the absolute value function. These methods are advantageous because proper choice of the potential function leads to a strictly convex regularizer, which guarantees a unique solution. The disadvantage of incremental optimization transfer methods proposed in [2] is that they are not able to solve the exact ℓ_1 penalized-likelihood problem. An incremental optimization transfer algorithm was proposed in [2] to solve an approximation to (1) with smooth regularizers.

Incremental optimization transfer methods are also advantageous because the updates are guaranteed to preserve the sign of the estimate, so the algorithm, by nature, maintains a non-negative estimate of emission rate. Furthermore, incremental optimization transfer methods converge to a fixed point, unlike ordered subsets methods [3].

In [4], an algorithm, SPIRAL- ℓ_1 , is proposed that solves (1) for orthonormal C by forming a quadratic approximation of the log-likelihood at each iteration. The subproblem at each iteration can be efficiently solved in its dual form. It is shown in [4] that SPIRAL- ℓ_1 performs well compared to [2], but it requires that C be an orthonormal transform. This condition does not allow C to be a finite differencing matrix or redundant dictionary.

Recent advances in image denoising with Poisson noise have yielded algorithms with fast convergence, such as [5]. The algorithm in [5] is based on the alternating direction method of multipliers (ADMM) [6], which belongs to the family of augmented Lagrangian methods [7]. Unlike incremental optimization transfer methods, augmented Lagrangian-based methods do not always produce solution sequences that satisfy the nonnegativity constraint at every iteration. Thus, termination after a finite number of iterations can yield an image with negative intensities.

In image denoising, the system matrix A is typically an approximately circulant blur operator. The method of [5] exploits the assumption of an approximately circulant system matrix by proposing to exactly solve a linear system at each iteration using the fast Fourier transform (FFT). Since the system matrix A is not circulant in emission tomography problems, one must tolerate a higher computational cost per iteration than is required for image denoising. This will be discussed in more detail in the next section.

A competing method to that in [5] is given in [8]. The method of [8] requires fewer operations per iteration, but approximates the Poisson likelihood by the Gaussian likelihood using the Anscombe transform. In [5], it is shown that the

ADMM method leads to lower mean absolute error than the method in [8] in an image denoising problem. This method is compared to the ADMM method in [5].

In this work, we extend the ℓ_1 regularized shift-invariant restoration method in [5] to emission tomography reconstruction. This extension requires modifications to the PIDAL-FA algorithm proposed in [5] to compensate for ill-conditioned tomographic system matrices and to approximately solve the subproblems at each iteration because direct inversion is both ill-conditioned and computationally expensive.

We compare the proposed algorithm to an optimization transfer algorithm that uses the techniques of [9] to derive surrogate functions for the ℓ_1 penalty term. We also compare these algorithms to incremental optimization transfer algorithms, proposed in [2], using a hyperbolic approximation to the absolute value.

II. ALGORITHM

The proposed augmented Lagrangian-based minimization algorithm is based on the ADMM method [5] shown below.

```

1 Set  $k = 0$ , choose  $\mu > 0$ ,  $\mathbf{u}_0$  and  $\mathbf{d}_0$ 
2 repeat
3    $\mathbf{x}_{k+1} \in \arg \min_{\mathbf{x}} f_1(\mathbf{x}) + \frac{\mu}{2} \|\mathbf{G}\mathbf{x} - \mathbf{u}_k - \mathbf{d}_k\|_2^2$ 
4    $\mathbf{u}_{k+1} \in \arg \min_{\mathbf{u}} f_2(\mathbf{u}) + \frac{\mu}{2} \|\mathbf{G}\mathbf{x}_{k+1} - \mathbf{u} - \mathbf{d}_k\|_2^2$ 
5    $\mathbf{d}_{k+1} = \mathbf{d}_k - (\mathbf{G}\mathbf{x}_{k+1} - \mathbf{u}_{k+1})$ 
6    $k = k + 1$ 
7 until stopping criterion satisfied;

```

Algorithm 1: General ADMM algorithm

We extend the PIDAL-FA restoration algorithm proposed in [5] by incorporating three constants α_1 , α_2 , and $\alpha_3 > 0$ to adjust the spectrum of $G^T G$. Let

$$\mathbf{u} = [(\mathbf{u}^{(1)})^T, (\mathbf{u}^{(2)})^T, (\mathbf{u}^{(3)})^T]^T,$$

where $\mathbf{u}^{(1)} \in \mathbb{R}^{n_d}$, $\mathbf{u}^{(2)} \in \mathbb{R}^K$, and $\mathbf{u}^{(3)} \in \mathbb{R}^{n_p}$, and choose

$$G = \begin{bmatrix} \alpha_1 A \\ \alpha_2 C \\ \alpha_3 I_{n_p} \end{bmatrix}, \quad (3)$$

where I_{n_p} is the $n_p \times n_p$ identity matrix. We choose the functionals $f_1(\mathbf{x}) = \mathbf{0}$ and

$$f_2(\mathbf{u}) = -L\left(\frac{1}{\alpha_1} \mathbf{u}^{(1)}\right) + \beta \left\| \frac{1}{\alpha_2} \mathbf{u}^{(2)} \right\|_1 + \mathcal{I}_{\mathbb{R}^+}(\mathbf{u}^{(3)}),$$

where $L(\mathbf{x})$ denotes the Poisson log-likelihood and

$$\mathcal{I}_{\mathbb{R}^+}(\mathbf{x}) = \begin{cases} 0 & \mathbf{x} \in \mathbb{R}^+ \\ \infty & \text{else.} \end{cases}$$

Since $f_1(\mathbf{x}) = \mathbf{0}$, line 3 of Algorithm 1, becomes

$$\mathbf{x}_{k+1} = \arg \min_{\mathbf{x}} \|\mathbf{G}\mathbf{x} - \mathbf{u}_k - \mathbf{d}_k\|_2^2,$$

which is an ordinary least-squares problem with solution \mathbf{x}_{k+1} that satisfies linear system

$$G' G \mathbf{x}_{k+1} = G' \boldsymbol{\xi}_k, \quad (4)$$

where $\boldsymbol{\xi}_k \triangleq \mathbf{u}_k + \mathbf{d}_k$. Substituting (3) into (4), we obtain the system

$$\begin{aligned} & (\alpha_1^2 A' A + \alpha_2^2 C' C + \alpha_3 I_{n_p}) \mathbf{z}_{k+1} \\ & = \alpha_1 A' \boldsymbol{\xi}_k^{(1)} + \alpha_2 C' \boldsymbol{\xi}_k^{(2)} + \alpha_3 \boldsymbol{\xi}_k^{(3)}, \end{aligned} \quad (5)$$

which can be solved for \mathbf{z}_{k+1} by the conjugate gradient algorithm since $G' G$ is hermitian. Note that (5) can be solved without explicitly computing the matrices C and A provided that vector multiplication by A , A' , C and C' are defined.

The α parameters are critical to ensure convergence in a reasonable time because the speed of the conjugate gradient algorithm depends on the condition number of $G' G$. The constants α_1 , α_2 , and α_3 allow one to balance the influence of $A' A$, $C' C$, and I in (5). In emission tomography, $A' A$ is typically ill-conditioned, and the condition number of $C' C$ depends on the choice of C . One can gain efficiency in solving (5) by choosing C to be an orthonormal basis, such as wavelets, to make $C' C = I_{n_p}$, thus avoiding evaluation of a forward and backward transform in each conjugate gradient step.

The remaining steps of the modified ADMM algorithm are derived using the methods of [5], accounting for the conditioning constants. The modified ADMM algorithm for emission tomography is given in Algorithm 2, where $\text{soft}(x, \beta)$ is the soft thresholding function defined by

$$\text{soft}(x, \beta) = \begin{cases} x - \beta & x > \beta \\ x + \beta & x < -\beta \\ 0 & \text{else,} \end{cases}$$

for $\beta \geq 0$. Hereafter, we refer to Algorithm 2 as modified PIDAL-FA, using the name given in [5].

```

1 Choose  $\mathbf{u}_0^{(1)}$ ,  $\mathbf{u}_0^{(2)}$ ,  $\mathbf{u}_0^{(3)}$ ,  $\mathbf{d}_0^{(1)}$ ,  $\mathbf{d}_0^{(2)}$ ,  $\mathbf{d}_0^{(3)}$ ,  $\mu$ ,  $\beta$ ,  $\alpha_1$ ,  $\alpha_2$ ,  $\alpha_3$ . repeat
2    $\boldsymbol{\xi}_k = \mathbf{u}_k + \mathbf{d}_k$ 
3   Solve (5) for  $\mathbf{z}_{k+1}$  using conjugate gradient
4    $\boldsymbol{\nu}_k^{(1)} = \alpha_1 A \mathbf{z}_{k+1} - \mathbf{d}_k^{(1)}$ 
5    $\mathbf{b}_k = \boldsymbol{\nu}_k^{(1)} - \frac{\alpha_1}{\mu} - r_i \alpha_1$ 
6    $\mathbf{u}_k^{(1)} = \frac{1}{2} \left( \mathbf{b}_k + \sqrt{\mathbf{b}_k^2 - 4 \left( \frac{y_i - r_i}{\mu} + r_i \boldsymbol{\nu}_k^{(1)} \alpha_1 \right)} \right)$ 
7    $\boldsymbol{\nu}_k^{(2)} = \alpha_2 C \mathbf{z}_{k+1} - \mathbf{d}_k^{(2)}$ 
8    $\mathbf{u}_k^{(2)} = \text{soft} \left( \boldsymbol{\nu}_k^{(2)}, \frac{\beta}{\mu \alpha_2} \right)$ 
9    $\boldsymbol{\nu}_k^{(3)} = \alpha_3 \mathbf{z}_{k+1} - \mathbf{d}_k^{(3)}$ 
10   $\mathbf{u}_k^{(3)} = \max \left( \boldsymbol{\nu}_k^{(3)}, 0 \right)$ 
11 until stopping criterion satisfied;

```

Algorithm 2: Modified PIDAL-FA Algorithm

A. Convergence of ADMM Algorithm

The convergence of ADMM algorithms is proved in [6] and the result is also shown in [5]. The convergence theorem states that the ADMM algorithm converges to a fixed point

of the cost function provided that the solution is unique, and the sequence of errors in the solutions to lines 2 and 3 of Algorithm 1 are absolutely summable. This convergence result suggests solving line 3 of Algorithm 1 to increasing precision at each iteration to reduce computation in early iterations.

III. RESULTS

We compare the proposed algorithm to the incremental optimization transfer method proposed in [2] using the regularizer in (2) and the hyperbolic potential function, which approximates the absolute value function and is defined by

$$\psi(t) = \delta \left(\sqrt{1 + |t/\delta|^2} - 1 \right). \quad (6)$$

A. Simulation Setup

The software for the incremental optimization transfer algorithm [2] was obtained from the image reconstruction toolbox (IRT) at <http://www.eecs.umich.edu/~fessler/irt/irt>, and the software for the conjugate gradient algorithm is part of the *l1magic* toolbox available at <http://www.acm.caltech.edu/l1magic/>.

We tested the algorithm on a Zubal torso phantom [10] with size 128×128 pixels. The true phantom is shown in Figure 1.

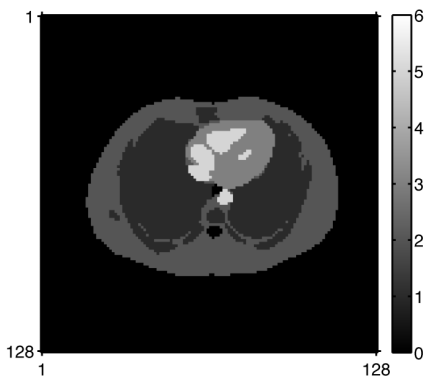


Fig. 1: True image

The simulated imaging system follows the strip-integral model and the system forward and backprojection operators were provided by the IRT. The mean number of total counts is 10^5 , the number of radial samples in the sinogram is 130, and the number of angular samples is 192. The mean background intensities r_i are 10% of the mean count rate for each corresponding detector bin. For all experiments, the matrix C is a first order finite-differences matrix that incorporates horizontal, vertical, and both diagonal differences. The regularization parameter $\beta = 0.1$ for all simulations because it was found empirically to produce a reasonable image when minimizing the true cost function in (1). All algorithms are initialized with an estimate of the image produced by filtered-back-projection.

The parameters used for the proposed algorithm are $\mu = \beta/\max(\mathbf{x}_{\text{true}})$, where \mathbf{x}_{true} is the true image, $\alpha_1 = 1/20$, $\alpha_2 = 2$, and $\alpha_3 = 10$. Conjugate gradient iterations are run until until the norm squared of the residual is $0.01 \cdot 2^{-i}$, in the i th iteration of the main algorithm. Varying these parameters within a factor of two did not affect convergence time significantly. The values of the α 's are related to the spectra of $A'A$, and $C'C$. Since A is ill-conditioned, the small value of α_1 reduces the largest eigenvalue of $G'G$. The large value of α_3 increases the smallest eigenvalue of $G'G$. These particular values were chosen by evaluating the spectral radius of $G'G$ with power iteration and inverse iteration. A reasonable choice of the α coefficients is critical for the conjugate gradient step of modified PIDAL-FA to converge to reasonable precision in a reasonable amount of time.

B. Simulation Results

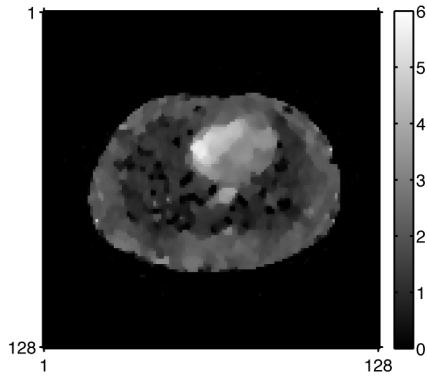
Figure 2 shows the reconstructed images using the incremental optimization transfer (EPL-INC) algorithm after 100 iterations, and Figure 3 shows the image reconstructed with the proposed algorithm. The image reconstructed with the proposed algorithm has sharp edges and large uniform regions, consistent with the finite-differencing penalty. The result of EPL-INC also has sharp edges, but there is less uniformity than the image reconstructed with the proposed algorithm.

The clinical utility of the image reconstructed with PIDAL-FA compared to those reconstructed with EPL-INC remains an open question. The images reconstructed with EPL-INC appear less uniform than the image reconstructed with PIDAL-FA, but it is more difficult to see small features in the image reconstructed with PIDAL-FA. We only tested PIDAL-FA with a finite-differencing penalty matrix C . A different choice for C that represents the image with greater sparsity could result in higher-quality reconstructed images.

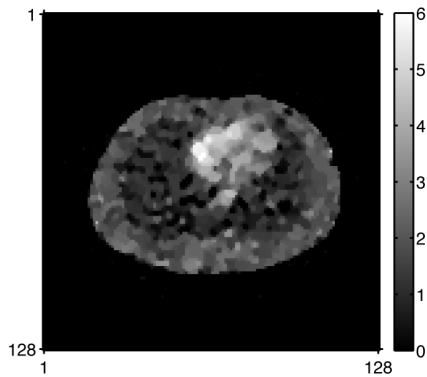
Figure 4 shows the cost function as a function of time for the two algorithms. The proposed algorithm doesn't maintain nonnegativity at each iterate, so we set all negative values to zero. The speed of these two algorithms appears to be similar for this problem, with the proposed algorithm being slightly slower initially. To achieve both speed and an exact solution to the ℓ_1 penalized likelihood problem, one could use EPL-INC for the initial iterations and then use PIDAL-FA to approach the exact solution.

IV. CONCLUSION

This work presents a modification of the PIDAL-FA algorithm proposed in [5], which is an augmented Lagrangian-based method for solving the ℓ_1 penalized likelihood problem with Poisson data. The proposed algorithm was shown to reduce the cost function at a similar speed as an incremental optimization transfer algorithm for solving an approximation to the ℓ_1 penalized likelihood problem. However, the incremental methods based on the approximate regularizer do not solve the true ℓ_1 penalized likelihood problem and their resulting images are not as sparse in the transform described by C , which may or may not be desirable in practice.



(a) $\delta = 0.1$



(b) $\delta = 0.001$

Fig. 2: Reconstructed images using EPL-INC with different values of the parameter δ

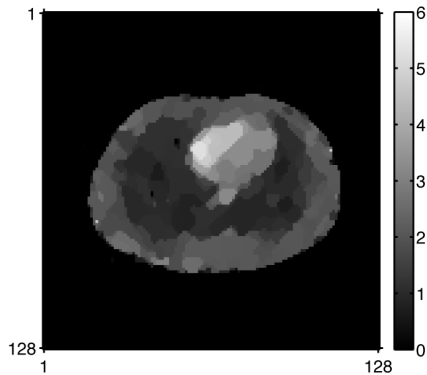


Fig. 3: Reconstructed image using PIDAL-FA

Future work will explore circulant approximations to G to accelerate convergence of the proposed algorithm and to apply the proposed algorithm to regularizers with other potential functions.

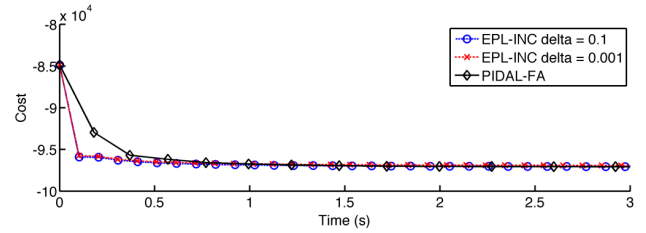


Fig. 4: Cost function. All negative values are set to zero prior to computing the cost.

ACKNOWLEDGEMENTS

The authors would like to acknowledge helpful discussions with Sathish Ramani about Augmented Lagrangian methods.

This work was supported by the U.S. Department of Homeland Security - DNDO under Grant Award Number 2008-DN-077-ARI007-04, NIH grant ROI-HL-098686, the MIT Lincoln Laboratory Fellowship, and the Innovative Signal Analysis Fellowship.

REFERENCES

- [1] E. Candes and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, February 2005.
- [2] S. Ahn, J. A. Fessler, D. Blatt, and A. O. Hero, "Convergent incremental optimization transfer algorithms: Application to tomography," *IEEE Trans. Med. Imag.*, vol. 25, no. 3, pp. 283–96, Mar. 2006.
- [3] H. M. Hudson and R. S. Larkin, "Accelerated image reconstruction using ordered subsets of projection data," *IEEE Trans. Med. Imag.*, vol. 13, no. 4, pp. 601–9, Dec. 1994.
- [4] Z. T. Harmany, R. F. Marcia, and R. M. Willett, "Spiral out of convexity: sparsity-regularized algorithms for photon-limited imaging," in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, ser. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, vol. 7533, February 2010.
- [5] M. A. T. Figueiredo and J. M. Bioucas-Dias, "Restoration of poissonian images using alternating direction optimization," *IEEE Transactions on Image Processing*, 2010.
- [6] J. Eckstein and D. P. Bertsekas, "On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Mathematical Programming*, vol. 55, no. 1, pp. 293–318, April 1992.
- [7] R. T. Rockafellar, "Augmented lagrangians and applications of the proximal point algorithm in convex programming," *Mathematics of Operations Research*, vol. 1, no. 2, pp. 97–116, 1976.
- [8] F. X. Dupe, J. M. Fadili, and J. L. Starck, "A proximal iteration for deconvolving poisson noisy images using sparse representations," *IEEE Transactions on Image Processing*, vol. 18, no. 2, pp. 310–321, February 2009.
- [9] A. R. De Pierro, "On the relation between the ISRA and the EM algorithm for positron emission tomography," *IEEE Trans. Med. Imag.*, vol. 12, no. 2, pp. 328–33, Jun. 1993.
- [10] I. G. Zubal, C. R. Harrell, E. O. Smith, Z. Rattner, G. Gindi, and P. B. Hoffer, "Computerized three-dimensional segmented human anatomy," *Medical Physics*, vol. 21, no. 2, pp. 299–302, 1994.