Numerical Algorithms manuscript No.
(will be inserted by the editor)

A Computational Framework for Total Variation-Regularized Positron Emission Tomography

Johnathan M. Bardsley · **John Goldes**

Received: date / Accepted: date

Abstract In positron emission tomography, image data corresponds to measurements of emitted photons from a radioactive tracer in the subject. Such count data is typically modeled using a Poisson random variable, leading to the use of the negative-log Poisson likelihood fit-to-data function. Regularization is needed, however, in order to guarantee reconstructions with minimal artifacts. Given that tracer densities are primarily smoothly varying, but also contain sharp jumps (or edges), total variation regularization is a natural choice. However, the resulting computational problem is quite challenging. In this paper, we present a efficient computational method for this problem. We also introduce three regularization parameter choice methods for use on total variation-regularized negative-log Poisson likelihood problems, which to our knowledge, have not been presented elsewhere. We test the computational and regularization parameter selection methods on synthetic data.

Keywords total variation · positron emission tomography · inverse problems · statistical imaging

Mathematics Subject Classification (2000) 65J22 · 65K10 · 65F22.

1 Introduction

In the medical imaging modality known as positron emission tomography (PET), a radioactive tracer element is injected into a subject. The tracer then exhibits radioactive decay, emitting positrons, which annihilate with electrons, producing a pair of annihilation photons moving in diametrically opposite directions. If both photons reach the detector ring surrounding the subject an event is registered along the line connecting the corresponding detector pair. This line is known as the line of response (LOR). The collection of all such

Supported by the NSF under grant DMS-0915107

Department of Mathematical Sciences
University of Montana
Phone: +1-406-243-5311
Fax: +1-406-243-2674
E-mail: johnathan.bardsley@umontana.edu
E-mail: john.goldes@umontana.edu

events from a particular experiment corresponds to a PET data set. From such data, the task is to reconstruct the tracer density distribution within the subject.

In [18,23], a mathematical model for PET image formation is presented. If there are M LORs and N elements in the uniform $\sqrt{N}\times\sqrt{N}$ computational grid, we can write the model in matrix-vector notation:

$$\mathbf{b} = \text{Poisson}(\mathbf{A}\mathbf{x}+\boldsymbol{\gamma}), \tag{1}$$

where $\mathbf{b}\in\mathbb{R}^M$ is the vector containing the recorded number of events along each of the M LORs; Poisson$(\boldsymbol{\lambda})$ denotes a Poisson random vector with mean vector $\boldsymbol{\lambda}\in\mathbb{R}^M$; $\mathbf{A}$ is the $M\times N$ forward model matrix; $\mathbf{x}\in\mathbb{R}^N$ is the discrete representation of the (unknown) photon emission density function; and $\boldsymbol{\gamma}$ is the vector containing expected erroneous counts due to accidental coincidences and scattered events [21], which we assume is known.

The attenuation matrix $\mathbf{A}$ in (1) is of the form

$$\mathbf{A} = \mathbf{G}\mathbf{A}^{\text{Radon}}, \tag{2}$$

where

$$\mathbf{G} = \text{diag}(g_1, g_2, \ldots, g_M), \quad g_j = \exp\left(-\int_{L_j}\mu(s)\,ds\right). \tag{3}$$

Here, L_j is the jth LOR; μ is the absorption density function of the subject; and $\mathbf{A}^{\text{Radon}}$ is the discrete $M\times N$ Radon transform matrix. Note that g_j can be viewed as the probability that an emission event anywhere along line L_j is recorded by the detector, and that the ijth element of $\mathbf{A}^{\text{Radon}}$ is the intersection length of the ith LOR with the jth computational grid element [16], and so is sparse.

The matrix $\mathbf{A}$ accounts for attenuation due both to Compton scattering and photo-electric absorption. Our mathematical/statistical model does not take into account detector efficiency or detector deadtime [21], however this has no bearing on our discussion. For a complete model of PET data, see [21].

Assuming (1), the probability mass function of the observable $\mathbf{b}$ conditioned on $\mathbf{x}$ is given by

$$p(\mathbf{b}\mid\mathbf{x}) = \prod_{j=1}^{M}\frac{([\mathbf{A}\mathbf{x}]_j+\gamma_j)^{b_j}\,e^{-([\mathbf{A}\mathbf{x}]_j+\gamma_j)}}{b_j!}. \tag{4}$$

Thus, given data $\mathbf{b}$ arising from model (1), the maximum likelihood estimate of the true image $\mathbf{x}_e$ is obtained by maximizing $p(\mathbf{b}\mid\mathbf{x})$ with respect to $\mathbf{x}$, subject to the constraint $\mathbf{x}\geq\mathbf{0}$. Equivalently, we can solve

$$\mathbf{x}_{\text{ML}} = \arg\min_{\mathbf{x}\geq\mathbf{0}} T_0(\mathbf{x};\mathbf{b}), \tag{5}$$

where

$$T_0(\mathbf{x};\mathbf{b}) = \sum_{i=1}^{n}\left\{([\mathbf{A}\mathbf{x}]_i+\gamma_i) - b_i\ln([\mathbf{A}\mathbf{x}]_i+\gamma_i)\right\}. \tag{6}$$

Note that $T_0(\mathbf{x};\mathbf{b})$ is equal, up to an additive constant, to $-\ln p(\mathbf{b}\mid\mathbf{x})$.

Solving (5) directly yields tracer densities with unrealistic artifacts. Because of this, a regularization term is often added (see, e.g., [1,8,9,10,11,13,12,14,15,19,20,22,26]), resulting in the following modification of (5): compute

$$\mathbf{x}_\alpha = \arg\min_{\mathbf{x}\geq\mathbf{0}}\left\{T_\alpha(\mathbf{x}) \stackrel{\text{def}}{=} T_0(\mathbf{x};\mathbf{b}) + \alpha J(\mathbf{x})\right\}, \tag{7}$$

where α is known as the regularization parameter and J the regularization function.

In the PET literature, (7) is known as a penalized maximum likelihood (PML) problem, and such problems have been studied extensively; see, e.g., [1,8,9,10,11,13,14,19,20,26]. Our focus in this paper is on the use of the total variation (TV) regularization function for J. In the context of PET, this problem has been studied less extensively; see e.g., [12,15,22]. For us the TV regularization function that we will use in our computations has the form

$$J(\mathbf{x}) = \sum_{i=1}^{N} \sqrt{[\mathbf{D}_1\mathbf{x}]_i^2 + [\mathbf{D}_2\mathbf{x}]_i^2 + \beta}, \tag{8}$$

where $\mathbf{D}_1$ and $\mathbf{D}_2$ are discretizations of the vertical and horizontal directional derivatives, respectively, with boundary conditions chosen to guarantee their invertibility. The positive parameter β is included in order to guarantee that J is differentiable.

For the application of image deconvolution (or deblurring) in the presence of Poisson noise, problem (6), (7) is often solved. If total variation regularization is used in this setting, the resulting computational and theoretical problem is challenging and has been studied by the authors [2,6]. In this paper, we apply the computational method of [2] to the PET image reconstruction problem. In [2,7] this method was shown to be more efficient than several other state-of-the-art minimization methods.

Additionally, we introduce a means of estimating the regularization parameter in (7). In [2], no such method was discussed, while in [5], only quadratic regularization functions were considered. Here, we extend the ideas of [5] to the total variation case.

The paper is organized as follows. First, we present our computational method in Section 2, then move on to regularization parameter choice methods in Section 3. Finally, in Section 4, we test the algorithms on some synthetically generated examples.

2 The Computational Method

In this section, we outline a computationally efficient method for solving

$$\min_{\mathbf{x}\geq\mathbf{0}} T_\alpha(\mathbf{x}), \tag{9}$$

where T_α is defined by (7), (8). Before presenting the method, we must set notation, perform preliminary calculations, and introduce some basic techniques.

2.1 Preliminaries

The projection of a vector $\mathbf{x} \in \mathbb{R}^N$ onto the feasible set $\Omega := \{\mathbf{x} \in \mathbb{R}^N \mid \mathbf{x} \geq \mathbf{0}\}$ can be conveniently expressed as

$$\mathscr{P}(\mathbf{x}) := \arg\min_{\mathbf{v}\in\Omega} ||\mathbf{v}-\mathbf{x}|| = \max\{\mathbf{x},\mathbf{0}\},$$

where $\max\{\mathbf{x},\mathbf{0}\}$ is the vector whose ith component is zero if $x_i < 0$ and is x_i otherwise. The active set for a vector $\mathbf{x} \geq \mathbf{0}$ is defined

$$\mathscr{A}(\mathbf{x}) = \{i \mid x_i = 0\}, \tag{10}$$

and the complementary set of indices, $\mathscr{I}(\mathbf{x})$, is known as the inactive set.

The reduced gradient of T_α at $\mathbf{x} \geq \mathbf{0}$ is given by

$$[\nabla_{\text{red}} T_\alpha(\mathbf{x})]_i = \begin{cases} [\nabla T_\alpha(\mathbf{x})]_i, & i \in \mathscr{I}(\mathbf{x}) \\ 0, & i \in \mathscr{A}(\mathbf{x}), \end{cases}$$

where ∇ denotes the gradient operator; the projected gradient of T_α is given by

$$[\nabla_{\text{proj}} T_\alpha(\mathbf{x})]_i = \begin{cases} [\nabla T_\alpha(\mathbf{x})]_i, & i \in \mathscr{I}(\mathbf{x}), \text{ or } i \in \mathscr{A}(\mathbf{x}) \text{ and } \frac{\partial T_\alpha(\mathbf{x})}{\partial x_i} < 0, \\ 0, & \text{otherwise;} \end{cases}$$

and the reduced Hessian is given by

$$[\nabla^2_{\text{red}} T_\alpha(\mathbf{x})]_{ij} = \begin{cases} [\nabla^2 T_\alpha(\mathbf{x})]_{i,j}, & \text{if } i \in \mathscr{I}(\mathbf{x}) \text{ and } j \in \mathscr{I}(\mathbf{x}) \\ \delta_{ij}, & \text{otherwise,} \end{cases} \tag{11}$$

where ∇^2 denotes the Hessian operator.

The gradient and Hessian of T_0 defined by (6) have the form

$$\nabla T_0(\mathbf{x}) = \mathbf{A}^T \left(\frac{\mathbf{A}\mathbf{x} - (\mathbf{b} - \gamma)}{\mathbf{A}\mathbf{x} + \gamma} \right), \tag{12}$$

$$\nabla^2 T_0(\mathbf{x}) = \mathbf{A}^T \text{diag} \left(\frac{\mathbf{b}}{(\mathbf{A}\mathbf{x} + \gamma)^2} \right) \mathbf{A}, \tag{13}$$

where all vector multiplication and division are intended component-wise

The gradient and Hessian of the regularization function J have the form

$$\nabla J(\mathbf{x}) = \mathbf{L}_1(\mathbf{x})\mathbf{x}, \tag{14}$$

$$\nabla^2 J(\mathbf{x}) = \mathbf{L}_1(\mathbf{x}) + 2\mathbf{L}_2(\mathbf{x}), \tag{15}$$

where, if $\psi(t) := \sqrt{t + \beta}$, $\mathbf{D}\mathbf{x}^2 := (\mathbf{D}_1\mathbf{x})^2 + (\mathbf{D}_2\mathbf{x})^2$, and $\mathbf{D}_{xy}\mathbf{x} := \mathbf{D}_1\mathbf{x} \odot \mathbf{D}_2\mathbf{x}$,

$$\mathbf{L}_1(\mathbf{x}) = \begin{bmatrix} \mathbf{D}_1 \\ \mathbf{D}_2 \end{bmatrix}^T \begin{bmatrix} \text{diag}(\psi'(\mathbf{D}\mathbf{x}^2)) & \mathbf{0} \\ \mathbf{0} & \text{diag}(\psi'(\mathbf{D}\mathbf{x}^2)) \end{bmatrix} \begin{bmatrix} \mathbf{D}_1 \\ \mathbf{D}_2 \end{bmatrix}, \tag{16}$$

$$\mathbf{L}_2(\mathbf{x}) = \begin{bmatrix} \mathbf{D}_1 \\ \mathbf{D}_2 \end{bmatrix}^T \begin{bmatrix} \text{diag}((\mathbf{D}_1\mathbf{x})^2 \odot \psi''(\mathbf{D}\mathbf{x}^2)) & \text{diag}(\mathbf{D}_{xy}\mathbf{x} \odot \psi''(\mathbf{D}\mathbf{x}^2)) \\ \text{diag}(\mathbf{D}_{xy}\mathbf{x} \odot \psi''(\mathbf{D}\mathbf{x}^2)) & \text{diag}((\mathbf{D}_2\mathbf{x})^2 \odot \psi''(\mathbf{D}\mathbf{x}^2)) \end{bmatrix} \begin{bmatrix} \mathbf{D}_1 \\ \mathbf{D}_2 \end{bmatrix}.$$

For a more detailed treatment of these computations see [24]. We note that $\nabla^2 J(\mathbf{x})$ is positive semi-definite for all $\mathbf{x}$, and hence, J is a convex function.

Next, we introduce the algorithm, which has two stages; the first involves gradient projection iterations.

2.2 Gradient Projection Iteration

A key component of the iterative method introduced in [7] is the gradient projection iteration [17], which we present now: given $\mathbf{x}_k \geq \mathbf{0}$, we compute $\mathbf{x}_{k+1}$ via

$$\mathbf{p}_k = -\nabla T_\alpha(\mathbf{x}_k), \tag{17}$$

$$\lambda_k = \arg\min_{\lambda>0} T_\alpha(\mathscr{P}(\mathbf{x}_k + \lambda \mathbf{p}_k)), \tag{18}$$

$$\mathbf{x}_{k+1} = \mathscr{P}(\mathbf{x}_k + \lambda_k \mathbf{p}_k). \tag{19}$$

In practice, subproblem (18) is solved inexactly using a projected backtracking line search. In the implementation used here, we take the initial step length parameter to be

$$\lambda_k^0 = \frac{||\mathbf{p}_k||^2}{\langle \nabla^2 T_\alpha(\mathbf{x}_k)\mathbf{p}_k, \mathbf{p}_k\rangle}. \tag{20}$$

The quadratic backtracking line search algorithm described in [4] is then used to create a sequence of line search parameters $\{\lambda_k^j\}_{j=0}^m$, where m is the smallest positive integer such that the sufficient decrease condition

$$T_\alpha(\mathbf{x}_k(\lambda_k^j)) \leq T_\alpha(\mathbf{x}_k) - \frac{\mu}{\lambda_k^j}||\mathbf{x}_k - \mathbf{x}_k(\lambda_k^j)||^2 \tag{21}$$

is satisfied, where $\mu \in (0,1)$ and

$$\mathbf{x}_k(\lambda) = \mathscr{P}_\Omega(\mathbf{x}_k + \lambda \mathbf{p}_k). \tag{22}$$

Then we take $\lambda_k := \lambda_k^m$ in (18).

Since we plan on only taking a few gradient projection iterations per outer iteration of our algorithm, we need an effective stopping criteria. Following [4], we use

$$T_\alpha(\mathbf{x}_{k-1}) - T_\alpha(\mathbf{x}_k) \leq \gamma_{GP} \max\{T_\alpha(\mathbf{x}_{i-1}) - T_\alpha(\mathbf{x}_i) \,|\, i = 1,\ldots,k-1\}, \tag{23}$$

where $0 < \gamma_{GP} < 1$. We have found the value $\gamma_{GP} = 0.1$ to be effective.

2.3 The Reduced Lagged-Diffusivity Step

In practice, like its unconstrained counterpart *steepest descent*, the gradient projection iteration is very slow to converge. A robust method with much better convergence properties results if gradient projection iterations are interspersed with steps computed from the reduced Newton system

$$\nabla^2_{\text{red}} T_\alpha(\mathbf{x}_k)\ \mathbf{p} = -\nabla_{\text{red}} T_\alpha(\mathbf{x}_k). \tag{24}$$

This is the approach taken in [7], however when total variation regularization is used, a more computationally efficient method results if (24) is replaced by the reduced quasi-Newton system

$$(\nabla^2_{\text{LD}})_{\text{red}} T_\alpha(\mathbf{x}_k)\ \mathbf{p} = -\nabla_{\text{red}} T_\alpha(\mathbf{x}_k), \tag{25}$$

where $(\nabla^2_{\text{LD}})_{\text{red}} T_\alpha(\mathbf{x})$ is defined as in (11) with

$$\nabla^2_{\text{LD}} T_\alpha(\mathbf{x}) = \nabla^2 T_0(\mathbf{x}) + \alpha \mathbf{L}_1(\mathbf{x}) \tag{26}$$

and $\mathbf{L}_1(\mathbf{x})$ defined in (16). We note that if in (26) T_0 is the regular least squares likelihood function $\|\mathbf{A}\mathbf{x}-\mathbf{b}\|^2$, the unreduced linear system

$$\nabla^2_{\mathrm{LD}} T_\alpha(\mathbf{x})\,\mathbf{p} = -\nabla T_\alpha(\mathbf{x}_k)$$

is that which is solved in each iteration of the lagged-diffusivity fixed point iteration of [25, 24]. Thus we call (25), (26) the reduced lagged-diffusivity system.

Approximate solutions of (25), (26) can be efficiently obtained using conjugate gradient iteration (CG) [17] applied to the problem of minimizing

$$q_k(\mathbf{p}) = T_\alpha(\mathbf{x}_k) + \langle \nabla_{\mathrm{red}} T_\alpha(\mathbf{x}_k), \mathbf{p}\rangle + \frac{1}{2}\langle (\nabla^2_{\mathrm{LD}})_{\mathrm{red}} T_\alpha(\mathbf{x}_k)\,\mathbf{p}, \mathbf{p}\rangle. \tag{27}$$

The result is a sequence $\{\mathbf{p}_k^j\}$ that converges to the minimizer of (27). Even with rapid CG convergence, for large-scale problems it is important to choose effective stopping criteria to reduce overall computational cost. We use

$$q_k(\mathbf{p}_k^{j-1}) - q_k(\mathbf{p}_k^j) \le \gamma_{CG} \max\{q_k(\mathbf{p}_k^{i-1}) - q_k(\mathbf{p}_k^i)\,|\, i = 1,\ldots,j-1\}, \tag{28}$$

where $0 < \gamma_{CG} < 1$. Then the approximate solution of (27) is taken to be the $\mathbf{p}_k^{m_{\mathrm{CG}}}$ where m_{CG} is the smallest integer such that (28) is satisfied.

With $\mathbf{p}_k := \mathbf{p}_k^{m_{\mathrm{CG}}}$, we again apply a projected backtracking line search, only this time we use the much less stringent acceptance criteria

$$T_\alpha(\mathbf{x}_k(\lambda_k^m)) < T_\alpha(\mathbf{x}_k). \tag{29}$$

2.4 The Numerical Algorithm and its Convergence

Finally, we can present the numerical method

Gradient Projection-Lagged-Diffusivity (GPLD) Iteration

Step 0: Select initial guess $\mathbf{x}_0$, and set $k = 0$.
Step 1: Given $\mathbf{x}_k$.
  (1) Take gradient projection steps until either (23) is satisfied or $\mathrm{GP}_{\max}$ iterations have been computed. Return updated $\mathbf{x}_k$.
Step 2: Given $\mathbf{x}_k$.
  (1) Do CG iterations to approximately minimize the quadratic (27) until either (28) is satisfied or $\mathrm{CG}_{\max}$ iterations have been computed. Return $\mathbf{p}_k = \mathbf{p}_k^{m_{\mathrm{CG}}}$.
  (2) Find λ_k^m that satisfies (29), and return $\mathbf{x}_{k+1} = \mathbf{x}_k(\lambda_k^m)$.
  (3) Update $k := k+1$ and return to Step 1.

This method was first applied to total variation regularization in [2]. For a general strictly convex and coercive function T with Lipschitz continuous gradient, the iterative quasi-Newton scheme above was shown to converge to the unique nonnegative minimizer of T in [4]. Thus, in order to prove convergence of GPLD applied to (7), we must show that T_α, with T_0 defined by (6), is strictly convex, and coercive, with Lipschitz continuous gradient. We do this now.

For strict convexity, note that for T_0 defined by (6), we have

$$\begin{aligned}\langle \nabla^2 T_\alpha(\mathbf{x})\mathbf{x},\mathbf{x}\rangle &= \left\langle \mathbf{A}^T \operatorname{diag}\left(\frac{\mathbf{b}}{(\mathbf{A}\mathbf{x}+\gamma)^2}\right)\mathbf{A}\mathbf{x},\mathbf{x}\right\rangle + \alpha\left\langle (\mathbf{L}_1(\mathbf{x})+2\mathbf{L}_2(\mathbf{x})))\mathbf{x},\mathbf{x}\right\rangle \\ &\geq \bar{\alpha}_1(\mathbf{x})\|\mathbf{A}\mathbf{x}\|^2 + \alpha(\|\Lambda(\mathbf{x})^{1/2}\mathbf{D}_1\mathbf{x}\|^2 + \|\Lambda(\mathbf{x})^{1/2}\mathbf{D}_2\mathbf{x}\|^2), \end{aligned} \tag{30}$$

where $\Lambda(\mathbf{x}) = \operatorname{diag}(\psi'(\mathbf{D}\mathbf{x}^2))$ (see (16)) and $\bar{\alpha}_1(\mathbf{x}) = \min_i \left\{ b_i/([\mathbf{A}\mathbf{x}]_i+\gamma)^2 \right\}$. Then if we assume that the intersection of the nullspaces of $\mathbf{A}$, $\mathbf{D}_1$, and $\mathbf{D}_2$ is trivial, and $b_i > 0$ for all i, since $\mathbf{A}\mathbf{x} \geq \mathbf{0}$ for all $\mathbf{x} \geq \mathbf{0}$, the inequality (30) is strict, implying that $\nabla^2 T_\alpha$ is positive definite for all $\mathbf{x} \geq \mathbf{0}$. Hence T_α is strictly convex on $\mathbf{x} \geq \mathbf{0}$.

The coercivity of T_α is proved using the following application of Jensen's inequality: for $\mathbf{x} \geq \mathbf{0}$

$$\begin{aligned} T_\alpha(\mathbf{x}) &\geq \|\mathbf{A}\mathbf{x}+\gamma\|_1 - \|\mathbf{b}\|_\infty \ln\|\mathbf{A}\mathbf{x}+\gamma\|_1 + \alpha(\|\Lambda(\mathbf{x})^{1/2}\mathbf{D}_1\mathbf{x}\|_2^2 + \|\Lambda(\mathbf{x})^{1/2}\mathbf{D}_2\mathbf{x}\|_2^2) \\ &= \|\mathbf{A}\mathbf{x}+\gamma\|_1 - \|\mathbf{b}\|_\infty \ln\|\mathbf{A}\mathbf{x}+\gamma\|_1 + \frac{\alpha}{2}\|\mathbf{D}\mathbf{x}^2/\sqrt{\mathbf{D}\mathbf{x}^2+\beta}\|_1, \end{aligned} \tag{31}$$

where recall, $\mathbf{D}\mathbf{x}^2 = \mathbf{D}_1\mathbf{x}^2 + \mathbf{D}_2\mathbf{x}^2$. Now, since $\mathbf{A}$, $\mathbf{D}_1$ and $\mathbf{D}_2$ have trivially intersecting null spaces, from (31) we see that $\|\mathbf{x}\|_1 \to +\infty$ implies $T_\alpha(\mathbf{x}) \to +\infty$, establishing coercivity.

Our last task is to prove Lipschitz continuity of ∇T_α. In [4], it is proven that for T_0 defined by (6), ∇T_0 is Lipschitz continuous. Additionally, we have

$$\begin{aligned} \|\mathbf{L}_1(\mathbf{x}) - \mathbf{L}_1(\mathbf{y})\|_2 &= \|(\mathbf{D}_1^T\Lambda(\mathbf{x})\mathbf{D}_1 + \mathbf{D}_2^T\Lambda(\mathbf{x})\mathbf{D}_2)\mathbf{x} - (\mathbf{D}_1^T\Lambda(\mathbf{y})\mathbf{D}_1 - \mathbf{D}_2^T\Lambda(\mathbf{y})\mathbf{D}_2)\mathbf{y}\|_2 \\ &\leq \frac{1}{2\sqrt{\beta}}\left(\|\mathbf{D}_1^T\mathbf{D}_1(\mathbf{x}-\mathbf{y})\|_2 + \|\mathbf{D}_2^T\mathbf{D}_2(\mathbf{x}-\mathbf{y})\|_2\right) \\ &\leq \frac{\rho_{\max}}{2\sqrt{\beta}}\|\mathbf{x}-\mathbf{y}\|_2, \end{aligned}$$

where $\rho_{\max} = \max\{\rho(\mathbf{D}_1^T\mathbf{D}_1^T), \rho(\mathbf{D}_2^T\mathbf{D}_2)\}$. Thus ∇J is Lipschtiz continuous, implying ∇T_α is Lipschitz continuous.

Thus we have proved the following theorem.

Theorem 1 *Assume $\mathbf{b} > \mathbf{0}$. If T_0 is defined by (6), and $\mathbf{A}$, $\mathbf{D}_1$, and $\mathbf{D}_2$ have trivially intersecting nullspaces, GPLD is guaranteed to converge to the unique minimizer of (7).*

3 Regularization Parameter Selection Methods

Methods for estimating the regularization parameter when the data noise follows a Poisson distribution have been presented in [5]. The implementation of those methods involved writing a quadratic Taylor series approximation of (6),

$$\begin{aligned} T_0(\mathbf{x};\mathbf{b}) =& T_0(\mathbf{x}_e;\mathbf{b}) + T_0^{\text{wls}}(\mathbf{x};\mathbf{b}) \\ &+ \mathscr{O}(\|\mathbf{h}\|_2^3, \|\mathbf{h}\|_2^2\|\mathbf{k}\|_2, \|\mathbf{h}\|_2\|\mathbf{k}\|_2^2, \|\mathbf{k}\|_2^2) \\ &+ \mathscr{O}(\|\mathbf{h}\|_2^2\|\mathbf{k}_\alpha\|_2, \|\mathbf{h}\|_2\|\mathbf{k}\|_2\|\mathbf{k}_\alpha\|_2, \|\mathbf{k}\|_2^2\|\mathbf{k}_\alpha\|_2), \end{aligned} \tag{32}$$

where $\mathbf{h} = \mathbf{x} - \mathbf{x}_e$, $\mathbf{k} = \mathbf{b} - \mathbf{b}_e$, $\mathbf{b}_e = \mathbf{A}\mathbf{x}_e + \gamma$, $\mathbf{k}_\alpha = \mathbf{b}_\alpha - \mathbf{b}_e$ with $\mathbf{b}_\alpha = \mathbf{A}\mathbf{x}_\alpha + \gamma$, and

$$T_0^{\text{wls}}(\mathbf{x};\mathbf{b}) = \frac{1}{2}\|\mathbf{B}_\alpha^{-1/2}(\mathbf{A}\mathbf{x} - (\mathbf{b}-\gamma))\|_2^2 \tag{33}$$

with $\mathbf{B}_\alpha = \text{diag}(\mathbf{b}_\alpha)$. This approximation enabled the extension of methods for least-squares estimation to Poisson likelihood estimation.

The techniques detailed in [5] were used to estimate α for problems in which the regularization term was quadratic. Here we apply those techniques to estimating α in the context of total variation regularization.

Discrepancy Principle. The idea behind the discrepancy principle, as presented in [5], is to choose α so that $\mathbf{x}_\alpha$ computed from (7) satisfies

$$2T_0^{\text{wls}}(\mathbf{x}_\alpha;\mathbf{b}) \approx E(2T_0^{\text{wls}}(\mathbf{x}_e;\mathbf{b})), \tag{34}$$

where $E(\cdot)$ denotes the expected value function. The approximation (32) can be used to argue that $2T_0^{\text{wls}}(\mathbf{x}_e;\mathbf{b})$ is approximately described by a $\chi^2(M)$ distribution, and so it follows that $E(2T_0^{\text{wls}}(\mathbf{x}_e;\mathbf{b})) = M$. The discrepancy principle (DP) recommendation is therefore defined as

$$\alpha_{\text{DP}} = \text{argmin}_{\alpha\geq 0}\left(T_0^{\text{wls}}(\mathbf{x}_\alpha;\mathbf{b}) - M/2\right)^2, \tag{35}$$

where $\mathbf{x}_\alpha$ is computed from (7). Note that the presence of the non-quadratic regularization term did not require a modification of the procedure given in [5]. It only affected the computation of $\mathbf{x}_\alpha$.

Generalized Cross Validation. [5] also details how the generalized cross validation method is adapted using approximation (32) for use in regularized Poisson-likelihood. The objective is to minimize the function

$$GCV(\alpha) = MT_0^{\text{wls}}(\mathbf{x}_\alpha;\mathbf{b}) \Big/ \text{trace}(\mathbf{I}_n - \mathbf{B}_\alpha^{-1/2}\mathbf{A}\mathbf{A}_\alpha)^2, \tag{36}$$

subject to the constraint $\alpha > 0$. $\mathbf{x}_\alpha$ is computed from (7) and $\mathbf{A}_\alpha$ is the matrix which satisfies $\mathbf{x}_\alpha \approx \mathbf{A}_\alpha\mathbf{B}_\alpha^{-1/2}(\mathbf{b}-\gamma)$. Here the form of $\mathbf{A}_\alpha$ is given by

$$\mathbf{A}_\alpha = (\mathbf{D}_\alpha(\mathbf{A}^T\mathbf{B}_\alpha^{-1}\mathbf{A} + \alpha\nabla^2 J(\mathbf{x}_\alpha))\mathbf{D}_\alpha)^\dagger \mathbf{D}_\alpha\mathbf{A}^T\mathbf{B}_\alpha^{-1/2}, \tag{37}$$

where † denotes "pseudo-inverse", and $\mathbf{D}_\alpha$ is a diagonal matrix with $[\mathbf{D}_\alpha]_{ii} = 1$ if $[\mathbf{x}_\alpha]_i > 0$ and 0 otherwise. This form differs from that given in [5] in that the regularization term in (37) is a function of $\mathbf{x}_\alpha$.

Because computing the trace can be impractical, randomized trace estimation is used to approximate the the expression in the denominator in (36) [24]. Letting $\mathbf{v}$ be a realization of a random vector $\mathbf{V}$ whose components are independent and take on values of 1 and -1 with equal probability the trace approximation is given by

$$\text{trace}(\mathbf{I}_n - \mathbf{B}_\alpha^{-1/2}\mathbf{A}\mathbf{A}_\alpha)^2 \approx (\mathbf{v}^T(\mathbf{I}_n - \mathbf{B}_\alpha^{-1/2}\mathbf{A}\mathbf{A}_\alpha)\mathbf{v})^2. \tag{38}$$

Note that in practice $\mathbf{A}_\alpha$ is not directly computed; $\mathbf{A}_\alpha\mathbf{v}$ is found by solving a linear system using truncated conjugate gradient iterations.

Unbiased Predictive Risk Estimator A third method for selecting a value for α is the unbiased predictive risk estimate (UPRE). This method aims to select the value of α which minimizes the predictive risk $T_0(\mathbf{x}_\alpha, \mathbf{b}_e)$ [24]. Since $\mathbf{b}_e$ is not known, UPRE instead chooses the value which minimizes an estimate of the expected value of the predictive risk. In [5] approximation (32) is again used to motivate the form of the UPRE function,

$$\mathrm{UPRE}(\alpha) = T_0^{\mathrm{wls}}(\mathbf{x}_\alpha; \mathbf{b}) + \mathrm{trace}(\mathbf{B}_\alpha^{-1/2}\mathbf{A}\mathbf{A}_\alpha) - \frac{M}{2}, \tag{39}$$

where $\mathbf{A}_\alpha$ is defined in (37). The trace is estimated using the same method that is employed in GCV. The UPRE value of α is taken to be that which minimizes (39).

4 Numerical Experiments

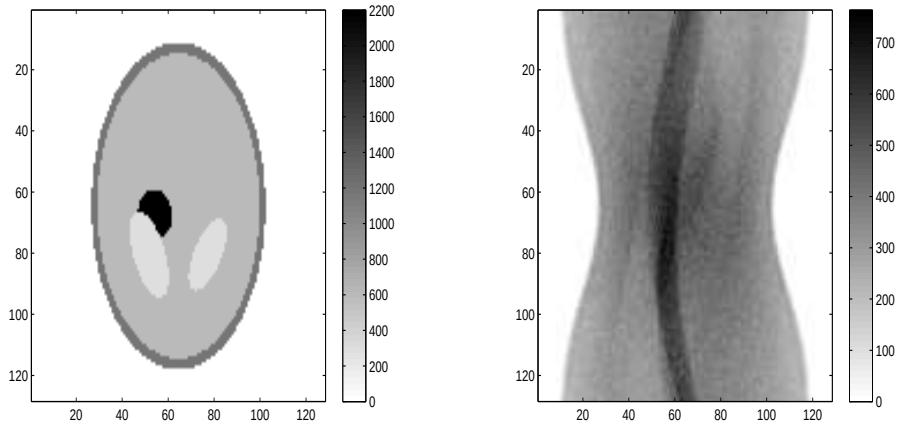

Fig. 1 *The true emission density* $\mathbf{x}_e$ *is plotted on the left and the data* $\mathbf{b}$ *is plotted on the right. The signal-to-noise ratio of* $\mathbf{b}$ *is 20.*

We test our approach on synthetically generated data. The true emission density $\mathbf{x}_e$ is shown in Figure 1. The noisy sinogram data $\mathbf{b}$, generated using statistical model (1) and MATLAB's `poissrnd` function, is shown on the right in Figure 1. We assumed that γ is a constant vector of 1s at all pixels, and that the density vector μ in (3) is zero. Our computational grid is defined by 128 detectors and angles, as well as a 128×128 uniform computational grid for the unknown emission density. Thus $M = N = 128^2$.

In order to test our approach on multiple data sets we vary the signal-to-noise ratio. The signal-to-noise ratio for data with statistical model (1) is defined as

$$SNR = \sqrt{\frac{\|\mathbf{A}\mathbf{x}_e + \gamma\|^2}{E(\|\mathbf{b} - (\mathbf{A}\mathbf{x}_e + \gamma)\|^2)}}, \tag{40}$$

with

$$E(\|\mathbf{b} - (\mathbf{A}\mathbf{x}_e + \gamma)\|^2) = \sum_{i=1}^{M} ([\mathbf{A}\mathbf{x}_e]_i + \gamma_i). \tag{41}$$

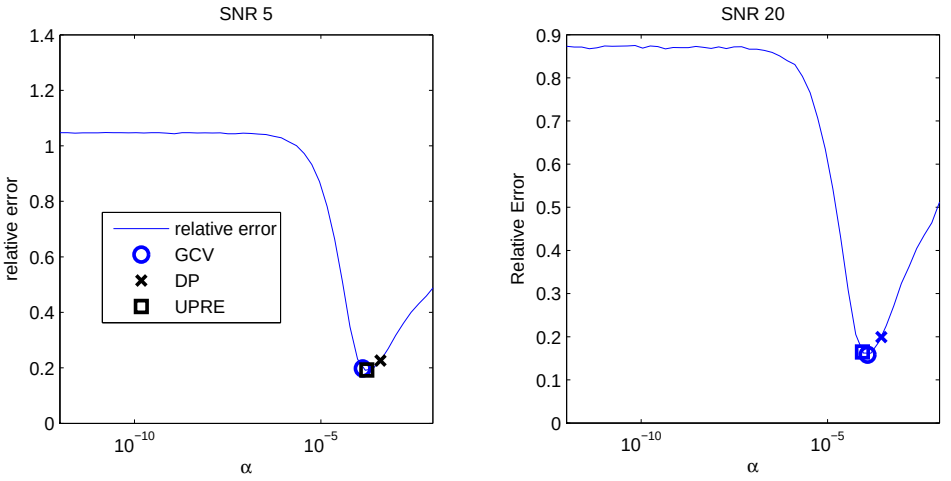

Fig. 2 *Plots of α versus relative error are shown. The plot on the left is from data with a SNR of 5 and the plot on the right is from data with a SNR of 20.*

To test the effectiveness of the regularization parameter selection methods, we plot the relative error

$$\frac{\|\mathbf{x}_\alpha - \mathbf{x}_e\|}{\|\mathbf{x}_e\|} \tag{42}$$

for a range of α values, together with the values of α chosen by the three methods. This can be seen in Figure 2. The regularized solution is found by solving (7). Note that in both cases of signal-to-noise ratio the three methods gave good recommendations for the value of α.

Figure 3 contains the reconstructions from the two data sets. Note the quality of the reconstructions is typical of total variation regularization.

In our implementation of GPLD, we used the parameters $\mathrm{GP}_{\max} = 5$, $\mathrm{CG}_{\max} = 30$, $\gamma_{\mathrm{GP}} = 0.1$, $\gamma_{\mathrm{CG}} = 0.1$, and we stopped iterations once

$$\|\nabla_{\mathrm{proj}} T_\alpha(\mathbf{x}_k)\|_2 / \|\nabla_{\mathrm{proj}} T_\alpha(\mathbf{x}_0)\|_2 < 10^{-5}. \tag{43}$$

In order to implement each of the three regularization parameter selection methods a minimization problem needs to be solved. DP requires the solution of (35), UPRE requires the minimizer of (39) to be computed, and GCV requires the minimizer of (36) to be computed. In all three cases the minimization is under the constraint $\alpha > 0$. To perform this we used Matlab's `fminbnd` function. `fminbnd` requires inputs for lower and upper bounds for the minimizer, which we set to 0 and 1.2 respectively. We set the `Tolx` parameter to 10^{-8}.

5 Conclusion

Given that tracer densities are typically smoothly varying, but contain edges, total variation (TV) regularization is a natural choice for positron emission tomography (PET) imaging problems. Nonetheless, due to the computationally challenging nature of the resulting problem, few researchers have studied this approach.

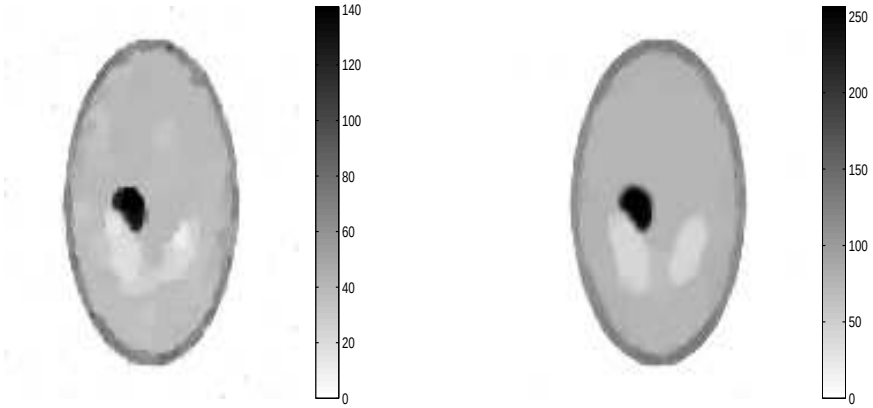

Fig. 3 *Plots of the reconstructions obtained from the two data sets with the UPRE recommendation for the regularization parameter with SNR 5 (on the left) and the DP recommendation for the regularization parameter with SNR 20 (on the right).*

In this paper, we have presented an efficient computational method for solving the TV-PET problem. Additionally, we have presented three methods for choosing the regularization parameter.

Our numerical experiments indicate that all of the methods work well.

References

1. Sangtae Ahn and Jeffrey Fessler, *Globally Convergent Image Reconstruction for Emission Tomography Using Relaxed Ordered Subsets Algorithms*, IEEE Transactions on Medical Imaging, 22(5), pp. 613-626, 2003.
2. Johnathan M. Bardsley, *An Efficient Computational Method for Total Variation-Penalized Poisson Likelihood Estimation*, Inverse Problems and Imaging, 2(2) (2008), pp. 167-185.
3. Johnathan M. Bardsley, Daniela Calvetti, and Erkki Somersalo, *Hierarchical regularization for edge-preserving reconstruction of PET images*, submitted, University of Montana, Department of Mathematical Sciences Tech. Report #1 2009.
4. Johathan M. Bardsley and John Goldes, *An Iterative Method for Edge-Preserving MAP Estimation when Data-Noise is Poisson*, accepted in the SIAM Journal on Scientific Computing.
5. Johnathan M. Bardsley and John Goldes, *Regularization parameter selection methods for ill-posed Poisson maximum likelihood estimation*, Inverse Problems, 25 (2009) 095005, doi: 10.1088/0266-5611/25/9/095005.
6. Johnathan M. Bardsley and Aaron Luttman, *Total Variation-Penalized Poisson Likelihood Estimation for Ill-Posed Problems*, Advances in Computational Mathematics, 31(1) (2009), pp. 35-59.
7. J. M. Bardsley and C. R. Vogel, *A Nonnnegatively Constrained Convex Programming Method for Image Reconstruction*, SIAM Journal on Scientific Computing, 25(4) 2004, pp. 1326-1343.
8. Jeffrey Fessler, *Penalized Weighted Least Squares Image Reconstruction for Positron Emission Tomography*, IEEE Transactions on Medical Imaging, 13(2), pp. 290-300, June 1994.
9. Jeffrey Fessler and Alfred O. Hero, *Penalized Maximum-Likelihood Image Reconstruction Using Space Alternation Generalized EM Algorithms*, IEEE Transactions on Image Processing, 4(10), pp. 1417-29, Oct. 1995.
10. J. A. Fessler and A. O. Hero, *Space-alternating generalized EM algorithms*, IEEE Transactions on Signal Processing, 42(10), pp. 2664-2677, Oct. 1994.
11. Peter Green, *Bayesian reconstructions from emission tomography data using a modified EM algorithm*, IEEE Transactions on Medical Imaging, 9(1), pp. 84-93, March 1990.

12. Hongbin Guoa, Rosemary A. Renauta, Kewei Chenb, and Eric Reimanb, *FDGPET parametric imaging by total variation minimization*, Computerized Medical Imaging and Graphics, 33 (2009), pp. 295303.
13. T. Herbert and R. Leahy, *A generalized EM algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs priors*, IEEE Transactions on Medical Imaging, 8(2), pp. 194-202, June 1989.
14. T. Hsiao, A. Rangarajan, and G. Gindi, *Bayesian image reconstruction for transmission tomography using deterministic annealing*, Journal of Electronic Imaging, Vol. 12, 7 (2003), doi:10.1117/1.1526103.
15. Elias Jonsson, Sung-cheng Huang, Tony Chan, *Total-Variation Regularization in Positron Emission Tomography*, Tech. Rep. 98-48, UCLA Group in Computational and Applied Mathematics, 1998.
16. Jari Kaipio and Erkki Somersalo, *Satistical and Computational Inverse Problems*, Springer 2005.
17. C. T. Kelley, *Iterative Methods for Optimization*, SIAM, Philadelphia, 1999.
18. K. Lange and R. Carson, *EM reconstruction algorithms for emission and transmission tomography*, Journal of Computer Assisted Tomography, 8, pp. 306-316, 1984.
19. S.-J. Lee, A. Rangarajan, and G. Gindi, *Bayesian image reconstruction in SPECT using higher order mechanical models as priors*, IEEE Tansactions on Medical Imaging, 14(4), pp. 669-680, 1995.
20. E. U. Mumcuoglu, R. Leahy, S. R. Cherry, Z. Zhou, *Fast Gradient-based methods for Bayesian reconstruction of transmission and Emission PET images*, IEEE Transactions on Medical Imaging, 13, pp. 687-701, 1994.
21. John M. Ollinger and Jeffrey A. Fessler, *Positron-Emission Tomography*, IEEE Signal Processing Magazine, January 1997.
22. Alex Sawatzky, Christoph Brune, Jahn Müller, and Martin Burger, *Total Variation Processing of Images with Poisson Statistics*, Computer Analysis of Images and Patterns, 5702 (2009), pp. 533-540.
23. L. A. Shepp and Y. Vardi, *Maximum likelihood reconstruction in positron emission tomography*, IEEE Transactions on Medical Imaging, vol. MI-1, pp. 113-122, 1982.
24. Curtis R. Vogel, *Computational Methods for Inverse Problems*, SIAM, Philidelphia, 2002.
25. C. R. Vogel and M. E. Oman, *A fast, robust algorithm for total variation based reconstruction of noisy, blurred images*, IEEE Transactions on Image Processing, **7** (1998), pp. 813-824.
26. D. F. Yu and J. A. Fessler, *Edge-Preserving Tomographic Reconstruction with Nonlocal Regularization*, IEEE Trans. on Medical Imaging, 21(2), 2002, pp. 159-173.