

Hierarchical regularization for edge-preserving reconstruction of PET images

Johnathan M. Bardsley[†], Daniela Calvetti*, and Erkki Somersalo*

[†]Department of Mathematical Sciences, University of Montana, Missoula, MT 59812, USA.

* Department of Mathematics, Case Western Reserve University, Cleveland, OH 44106. USA.

E-mail: johnathan.bardsley@umontana.edu, daniela.calvetti@case.edu, erkki.somersalo@case.edu

Abstract. The data in PET emission and transmission tomography and in low dose X-ray tomography, consists of counts of photons originating from random events. The need to model the data as a Poisson process poses a challenge for traditional integral geometry-based reconstruction algorithms. Although qualitative a priori information of the target may be available, it may be difficult to encode it as a regularization functional in a minimization algorithm. This is the case, for example, when the target is known to consist of well defined structures, but how many, and their location, form and size are not specified. Following the Bayesian paradigm, we model the data and the target as random variables, and we account for the qualitative nature of the a priori information by introducing a hierarchical model in which the a priori variance is unknown and therefore part of the estimation problem. We present a numerically effective algorithm for estimating both the target and its prior variance. Computed examples with simulated and real data demonstrate that the algorithm gives good quality reconstructions for both emission and transmission PET problems in an efficient manner.

1. Introduction

Positron Emission Tomography (PET) is a functional imaging modality that has helped improve our understanding of the metabolism of various organs in a living body. In contrast to in vitro experiments with cell cultures, PET gives valuable information about the uptake of metabolites by cells in vivo and in situ, i.e., in their natural competitive environment. For example, PET imaging provided new, significant insight into the complex interplay between neurons and glial cells [22, 34] in the context of brain energetics. Moreover, it is used routinely to test if metastases have developed in cancer patients.

PET imaging is based on the detection of photon pairs emitted due to positron-electron annihilations resulting from the decay of the radioactive species used as a

marker. Since the PET signal is generated by spontaneous radioactive decay, the data is best modeled as a counting (Poisson) process [41]. Two observations are in order: the first is that the signal is, by its very nature, random; and the second is that for the inverse problem of estimating the emission density distribution complementary information about the source is needed in order to produce a reasonable and useful solution. The Bayesian statistical framework will allow us to conveniently address these two points ([11, 26, 29, 32]).

While the modeling of the observation noise as a random variable is generally accepted even in non-statistical frameworks, the peculiar feature of the Bayesian approach is the modeling of the primary unknown as a random variable. Ideally, the corresponding probability density, known as the *prior*, encodes the observer's prior knowledge about the unknown. In this context, randomness can be viewed – in the spirit of de Finetti [16] – as an expression of the observer's lack of certainty about the unknown. Since there is no unique way to interpret prior knowledge and assign prior probabilities, the question of whether a prior distribution is correct cannot be answered.

Since the classical works on Bayesian imaging ([6, 7, 24]), Bayesian methods have been employed extensively in imaging applications. In PET in particular, the Bayesian approach to image reconstruction has been extensively studied. The classical PET iterative method, known as *expectation-maximization* (EM) [31, 39], is an instance of the more general EM method of [17], which has a Bayesian flavor. Bayesian maximum a posteriori (MAP) estimation – in which a specific prior probability distribution is assumed on the unknown image – has received much attention in the PET literature, where it is also referred to as *penalized maximum likelihood* (PML). Work in this direction in the context of PET has focused primarily on modifications and improvements of EM with different penalty functionals (see the by-no-means exhaustive list [1, 20, 21, 25, 28, 26, 32, 35, 47] and the references therein).

Alternatively, an optimization method can be applied directly to the MAP problem. This is a challenging task due not only to the presence of a nonnegativity constraint on the image, but also because the negative-log Poisson likelihood is non-quadratic. The latter challenge is overcome in [19] by replacing the Poisson fit-to-data by a weighted least squares approximation, which makes the resulting approximate MAP problem quadratic. An efficient optimization algorithm for use on the penalized negative-log Poisson likelihood minimization problem implemented in the context of astronomical imaging in [5] has been shown to be effective for a number of different regularization functions [2, 3, 5]. We will make use of this algorithm here.

The question of the definition of the penalty (or regularization) function remains to be addressed. We want our regularization function to be edge-preserving. The standard approach for edge-preserving regularization based on total variation (TV) [2, 43] often leads to very challenging computational problems. A more computationally efficient method is obtained when a quadratic regularization functional is used within a hierarchical Bayesian framework, as proposed in [12]. The approach in [12] is similar to the one proposed in [14], but while in the latter an *improper prior* is used, in the former

the scaling of the prior is properly implemented. Here we extend this approach to the case in which the fit-to-data function is the negative-log Poisson likelihood.

Hierarchical approaches have been shown to be effective on a number of different applications, e.g., in recovering discontinuities in biological signals [15], particle size distributions in aerosol physics [42], temperature distributions in inverse heat conduction [45], blocky images in image deblurring [12, 14], missing information in image inpainting [13], as well as focal source identification in EEG/MEG [38, 8].

The paper is organized as follows. In the next section, we present the mathematical and statistical models for the PET problem. Our hierarchical regularization method is the topic of Section 3. The computational method used for the MAP estimation problem is outlined in Section 4. To conclude, computed results are presented in Section 5.

2. Model of PET signal

The PET modality is utilized to track the transport and uptake of certain metabolites which can be marked with radioactive isotopes in living organism. When the isotope decays, it emits a positron, which in turn annihilates with an electron, causing a pair of photons to propagate in opposite directions. If these two photons reach two different detectors within a sufficiently short time window they are counted as an event along the corresponding connecting line L , referred to as the Line Of Response (LOR).

2.1. Deterministic Model

In [31, 39], a mathematical/statistical model for this process is presented. It can be written in the discrete setting as a systems of M (# of LORs) linear equations with N (# of elements in the uniform $\sqrt{N} \times \sqrt{N}$ computational grid) unknowns:

$$I^{\text{emiss}} = \mathbf{A}^{\text{emiss}} \lambda + \beta, \quad (1)$$

where $I^{\text{emiss}} \in \mathbb{R}^M$ is the vector containing the expected number of events along each of the M LORs; $\mathbf{A}^{\text{emiss}}$ is the $M \times N$ forward model matrix; $\lambda \in \mathbb{R}^N$ is the discrete representation of the (unknown) photon emission density function; and β is the vector containing expected erroneous counts due to accidental coincidences and scattered events [37], which we assume known.

The coefficient matrix in (1) is of the form

$$\mathbf{A}^{\text{emiss}} = \mathbf{G} \mathbf{A}^{\text{Radon}}, \quad (2)$$

where $\mathbf{A}^{\text{Radon}}$ is the discrete $M \times N$ Radon transform matrix. Signal attenuation—due primarily to Compton scattering, but also to photoelectric absorption—is modeled by taking

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

with L_j is the j th LOR (see [37]). 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 ij th

element of A^{Radon} is the intersection length of the i th LOR with the j th computational grid element [29], and so is sparse.

Model (1) does not take into account detector efficiency or detector deadtime [37], however this will not effect our algorithm. For a complete model of PET data, see [37].

In order to use model (1), μ must be estimated. This motivates transmission PET, where the model is given, instead, by (see [31])

$$I^{\text{trans}} = I_0 \exp(-A^{\text{Radon}}\mu) + \beta. \quad (4)$$

Here $I^{\text{trans}} \in \mathbb{R}^M$ is the vector containing the measured intensities at each of the M sensors, $I_0 \in \mathbb{R}^M$ is the vector of source intensities (determined by a scan without the object and known as the *blank scan*); $\mu \in \mathbb{R}^N$ is the discrete representation of the attenuation function; and β is as in (1).

2.2. Statistical model

These deterministic models for the transmission and emission data are inadequate because the measured signal is noisy: the measurements are inaccurate, and the signal is itself the result of random scattering and emission of the photons. In the discussion to ensue, random variables are denoted by uppercase letters while their realizations are denoted by lowercase letters. For the sake of uniformity, we denote by $b \in \mathbb{R}^M$ the observed intensities I^{emiss} in the emission model and I^{trans} in the transmission model. Similarly, $x \in \mathbb{R}^N$ stands for the unknown density λ in the emission tomography problem and μ in the transmission problem. The matrices A^{emiss} and A^{trans} are denoted simply by A .

The data is typically modeled by a Poisson process in both the transmission and the emission PET problem. In the emission case [37], the standard statistical model corresponding to the discretized deterministic model (1) is given by

$$B_j \sim \text{Poisson}([Ax]_j + \beta_j), \quad j = 1, \dots, M, \quad (5)$$

where “ \sim ” means “is distributed as”, and $\text{Poisson}(\lambda)$ denotes a Poisson random vector with Poisson parameter vector λ .

In the remainder of the document, we will use π (with various subscripts) to denote probability density/mass functions.

Given (5), the probability mass function of the observable B conditioned on x is given by

$$\pi_{\text{emiss}}(b \mid x) = \prod_{j=1}^M \frac{([Ax]_j + \beta_j)^{b_j} e^{-([Ax]_j + \beta_j)}}{b_j!}. \quad (6)$$

We note that since Poisson random variables take on only discrete values, $\pi_{\text{emiss}}(b \mid x)$ should, in theory, be positive only for $b \in \mathbb{Z}_+^M$. However for the ease of both analysis and computation, we will treat it as a probability density defined on $\mathbb{R}_+^M \cup \{0\}$.

In the transmission PET case [46], the data-noise model has the form

$$B_j \sim \text{Poisson}(b_{0,j} e^{-[Ax]_j} + \beta_j), \quad j = 1, \dots, M, \quad (7)$$

where $b_{0,j}$ is the measured blank scan (I_0 in (4)). In this case, the probability mass function of B conditioned on x is of the form

$$\pi_{\text{trans}}(b | x) = \prod_{j=1}^m \frac{(b_{0,j} e^{-[Ax]_j} + \beta_j)^{b_j} e^{-(b_{0,j} e^{-[Ax]_j} + \beta_j)}}{b_j!}, \quad (8)$$

3. Hierarchical regularization

The Bayesian approach that we take in this paper is motivated from Bayes' Law, which provides an expression for the posterior probability of the random variable X conditioned on the observation b :

$$\pi(x | b) \propto \pi_{\text{prior}}(x) \pi_*(b | x), \quad (9)$$

where “ \propto ” means “is proportional up to an unimportant multiplicative factor”. Here π_* is given either by (6) or (8).

In our approach, which follows [12], we add another level, assuming that π_{prior} , and hence $\pi(x | b)$, depends upon a hyperparameter θ modeled as a random variable Θ with probability density $\pi_{\text{hyper}}(\theta)$. When this is the case, (9) becomes of the form

$$\pi(x, \theta | b) \propto \pi_{\text{hyper}}(\theta) \pi_{\text{prior}}(x | \theta) \pi_*(b | x). \quad (10)$$

Our goal is then to maximize the posterior density $\pi(x, \theta | b)$ with respect to x and θ subject to the constraint that $x \geq 0$. To do this, however, we must first define the probability densities π_{hyper} and π_{prior} .

3.1. The definition of the prior

It is realistic to assume that the emission activity vector and the attenuation vector (x in (5) and (7), respectively) will vary smoothly, except at edges between regions with different metabolic properties. Thus in regions of an image x not near an edge, we seek to penalize roughness, whereas at or near an edge such penalty is relaxed. Using the first derivative as a measure of roughness, we translate our assumptions mathematically as follows. Let X denote a multivariate random variable and assume that

$$\mathbf{L}_1 X, \mathbf{L}_2 X \sim \mathcal{N}(0, \mathbf{D}_\theta), \quad \mathbf{D}_\theta \stackrel{\text{def}}{=} \text{diag}(\theta_1, \dots, \theta_N), \quad (11)$$

where $\mathcal{N}(0, \mathbf{D}_\theta)$ denotes a Gaussian random vector with mean 0 and covariance \mathbf{D}_θ ; \mathbf{L}_1 and \mathbf{L}_2 are discretized horizontal and vertical first order partial derivatives with zero boundary conditions, respectively; and the θ_j 's are positive for $j = 1, \dots, N$. Thus, a decrease (increase) in the value of θ_j should correspond to an increase (decrease) in the roughness penalty.

In our definition of \mathbf{L}_1 and \mathbf{L}_2 , we intersperse first order forward and backward difference approximations of the x and y first partial derivatives, respectively, with zero boundary conditions. The backward difference approximation for \mathbf{L}_1 is defined by the 2D array operation $u(i, j) \leftarrow u(i, j) - u(i-1, j)$ with a zero boundary condition on the left, and the forward difference approximation is defined by $u(i, j) \leftarrow u(i+1, j) - u(i, j)$ with

a zero boundary condition on the right. These two approximations are interspersed in the iterative algorithm to ensue so that no boundary artifacts result in the reconstruction of x . The matrix \mathbf{L}_2 is defined analogously.

Assuming that $\mathbf{L}_1 X$ and $\mathbf{L}_2 X$ are independent random vectors we have that

$$\pi_{\text{prior}}(x | \theta) = \pi(\mathbf{L}_1 x, \mathbf{L}_2 x | \theta) \quad (12)$$

$$\begin{aligned} &= \pi(\mathbf{L}_1 x | \theta) \pi(\mathbf{L}_2 x | \theta) \\ &= \frac{|\mathbf{L}_1| |\mathbf{L}_2|}{(2\pi)^N |\mathbf{D}_\theta|} e^{-\frac{1}{2} x^\top (\mathbf{L}_1^\top \mathbf{D}_\theta^{-1} \mathbf{L}_1 + \mathbf{L}_2^\top \mathbf{D}_\theta^{-1} \mathbf{L}_2) x} \\ &\propto \left(\prod_{j=1}^N \theta_j^{-1} \right) e^{-\frac{1}{2} x^\top (\mathbf{L}_1^\top \mathbf{D}_\theta^{-1} \mathbf{L}_1 + \mathbf{L}_2^\top \mathbf{D}_\theta^{-1} \mathbf{L}_2) x}. \end{aligned} \quad (13)$$

Note that the invertibility of the matrices \mathbf{L}_1 and \mathbf{L}_2 , which greatly simplifies the design of the algorithm, is not strictly necessary, as was shown in [12].

If \mathbf{D}_θ is a constant multiple of the identity matrix, we obtain a rather standard regularization functional, which is very similar to that used in [19]. When \mathbf{D}_θ is different from the identity, a direct correspondence with the regularization functionals used in [3, 12, 14, 47] can be established in a straightforward manner. In the application to astronomical imaging in [3], where no hyper-model is used for the θ_j 's, the posterior distribution has the form of (9), rather than (10), as is also the case in [47]. The main difference between these two approaches concerns the interpretation of the θ_j 's. In [12, 14], where a hierarchical model is used and the θ_j 's are assumed to be random variables, it is necessary to use (10).

We finally remark that nothing in the methodology precludes the use of non multivariate Gaussian in (11) – for example, the more general Gibbs distribution [24, 28] – although the design of the corresponding algorithm may present new challenges.

3.2. Hyperprior definition

To complete the construction of the prior, we need to choose a hyperprior $\pi_{\text{hyper}}(\theta)$ conveying our belief about the θ_j 's, that is:

- The jumps in the true image should be sudden, hence the θ_j 's should be mutually independent.
- There is no obvious preference for the location of the jumps, therefore the θ_j 's should be identically distributed.
- Relatively few θ_j 's will be significantly large—indeed, most should be small—suggesting a hyperprior that allows rare outliers.

Under these conditions, a natural candidate for the hyperprior is the *gamma distribution*, defined as follows: if $\theta_j \sim \text{Gamma}(\alpha, \theta_0)$,

$$\pi_{\text{hyper}}(\theta) \propto \prod_{j=1}^N \theta_j^{\alpha-1} \exp\left(-\frac{\theta_j}{\theta_0}\right). \quad (14)$$

The mean and variance of the gamma distribution are $\alpha\theta_0$ and $\alpha\theta_0^2$, respectively. We note that this hyper-prior is used in both [12, 14]. In [12], the use of the inverse Gamma hyper-prior is also explored.

4. Iterative MAP computation

Taking all of the above into account, the MAP estimator is the minimizer, with respect to (x, θ) , of the negative-log of the posterior density (10) which, for the gamma hyper prior, is of the form

$$\begin{aligned}
 -\ln(\pi(x, \theta | b)) &\simeq -\ln(\pi_*(b | x)) + \frac{1}{2} \sum_{j=1}^N \frac{[\mathbf{L}_1 x]_j^2 + [\mathbf{L}_2 x]_j^2}{\theta_j} \\
 &\quad + \sum_{j=1}^N \frac{\theta_j}{\theta_0} - (\alpha - 2) \sum_{j=1}^N \ln \theta_j \\
 &= -\ln(\pi_*(b | x)) + \frac{1}{2} x^T \mathbf{C}_k x \\
 &\quad + \sum_{j=1}^N \frac{\theta_j}{\theta_0} - (\alpha - 2) \sum_{j=1}^N \ln \theta_j,
 \end{aligned} \tag{15}$$

where \simeq means “is equal up to an additive unimportant constant”, and $\mathbf{C}_k = \mathbf{L}_1^T \mathbf{D}_\theta^{-1} \mathbf{L}_1 + \mathbf{L}_2^T \mathbf{D}_\theta^{-1} \mathbf{L}_2$.

Although gradient-based algorithms could be used to minimize (15) directly, we propose a simple cyclic iterative algorithm that has been shown to be quite effective [12] for some classes of problems. The outline of the algorithm is as follows:

Iterative MAP Algorithm:

Step 0. Initialize $\theta = \theta^0$, $k = 1$.

Step 1. Update the estimate of the x vector by computing:

$$x^k = \arg \min_{x \geq 0} \left\{ T(x) \stackrel{\text{def}}{=} -\ln(\pi_*(b | x)) + \frac{1}{2} x^T \mathbf{C}_k x \right\}, \tag{16}$$

Step 2. Update the estimate of the θ vector by solving:

$$\theta^k = \arg \min_{\theta} \left\{ -\ln(\pi(x^k, \theta | b)) \right\}. \tag{17}$$

Step 3. Increase k by one and return to Step 1. Repeat until convergence.

In our implementation, we took θ^0 in Step 0 to be a constant vector with values θ_0 . The updated θ^k in Step 2 can be computed analytically. In particular, for the gamma hyper-prior,

$$\theta_j^k = \theta_0 \left(\frac{\alpha - 2}{2} + \sqrt{\frac{[\mathbf{L}_1 x^k]_j^2 + [\mathbf{L}_2 x^k]_j^2}{2\theta_0} + \frac{(\alpha - 2)^2}{4}} \right). \tag{18}$$

This equation provides a natural interpretation of the parameters θ_0 and α . If $\alpha = 2$, (18) becomes $\theta_j^k = \beta^{-1} \sqrt{[\mathbf{L}_1 x^k]_j^2 + [\mathbf{L}_2 x^k]_j^2}$, where $\beta = \sqrt{2/\theta_0}$. Substitution into (15)

then yields total variation regularized solution. For $0 < \alpha - 2 \ll 1$, the quadratic regularization function in (16) resembles total variation with regularization parameter approximately equal to $\beta = \sqrt{2/\theta_0}$. In our experiments, we chose $\alpha = 2.01$. The choice of the value of θ_0 is then analogous to choosing the value of the regularization parameter in standard inverse problems. This is an important problem because the performance of the algorithm is quite sensitive to the value of θ_0 . Methods for the selection of the regularization parameter for PML problems in the Poisson case has recently been explored in [4], however we save the extension of these methods to the current hierarchical setting for a later work due to the subtlety of the the issues involved.

If distributions other than multi-variate Gaussian are used in (11), a closed form expression for the θ update may not be available, in which case an iterative method would have to be used to obtain a numerical approximation of (17).

In general, no closed form expression for x^k in (16) exists, and even in the event that it did, the dimensions of the problems of interest require the use of iterative solution methods (see e.g., [19]). The solution of PML problems of the type (16) has received a lot of attention in the PET literature, with the majority of approaches seeking to extend the EM methodology (see e.g., [1, 21, 20, 25, 28, 35, 47]). Here we take a different approach and apply the nonnegatively constrained iterative method of [3, 5] to (16).

4.1. The iterative method for solving (16)

The algorithm that we will use was first proposed in [5] in the context of standard Tikhonov regularization for negative-log Poisson likelihood estimation. More recently, in [3], the method was shown to be convergent for problems of the form (16) with T strictly convex. In the appendix, we show that in both the emission and transmission PET problems the functional T is strictly convex.

For sake of completeness, we now provide a brief description of the PML iterative method of [3, 5], which we refer to as *gradient projection-reduced Newton* (GPRN). The reader interested in the details of this algorithm should see [3, 5].

GPRN Algorithm:

- Step 0.** Set $k = 0$ and choose initial guess x^0 such that $x_j^0 = 1$ for $j = 1, \dots, N$.
- Step 1.** Apply gradient projection iterations (see below for more detail) to (16) with initial guess x^k until certain stopping criteria are satisfied. Output the updated $x^k \geq 0$.
- Step 2.** Compute a quadratic Taylor series approximation q_k of T centered at x^k , restricted to components with indices r such that $x_r^k > 0$ (see below for more detail). Use the conjugate gradient (CG) iteration to minimize q_k until the stopping criteria are met. Use the most recent CG iterate as a search direction in a projected backtracking line search. Output x^{k+1} .
- Step 3.** If the outer iteration stopping criteria have been met, end the GPRN iterations. Otherwise, set $k = k + 1$ and return to Step 1.

Step 1: gradient projection iteration

The gradient projection iteration [30] used in Step 1 of GPRN is defined as follows: given $x^i \geq 0$ compute x^{i+1} via

$$\begin{aligned} p^i &= -\nabla T(x^i), \\ \lambda^i &= \arg \min_{\lambda > 0} T(\mathcal{P}(x^i + \lambda p^i)), \\ x^{i+1} &= \mathcal{P}(x^i + \lambda^i p^i). \end{aligned} \tag{19}$$

Here $\mathcal{P}(x) = \max\{x, 0\}$, where the maximum is computed component-wise. In practice, the subproblem (19) is solved inexactly using a projected backtracking line search algorithm; see [3] for details. The gradient projection iterations are well-defined and convergent for problems of the form (16) [30, Section 5.4]. The stopping rules for the gradient projection iterations in our computed examples within Step 1 of GPRN are as in [3, 5, 33]. In the implementation in this paper, at most 5 gradient projection iterations were allowed per GPRN iteration. The form of the gradient of T for both the emission and transmission cases can be found in the appendix.

Step 2: reduced conjugate gradient iterations

The quadratic Taylor series approximation of T used in Step 2 of iteration k of the GPRN algorithm is of the form

$$q_k(p) = T(x^k) + \langle \nabla_{\text{red}} T(x^k), p \rangle + \frac{1}{2} \langle \nabla_{\text{red}}^2 T(x^k) p, p \rangle, \tag{20}$$

where

$$[\nabla_{\text{red}} T(x)]_i = \begin{cases} [\nabla T(x)]_i, & x_i > 0 \\ 0, & x_i = 0, \end{cases}$$

and

$$[\nabla_{\text{red}}^2 T(x)]_{ij} = \begin{cases} [\nabla^2 T(x)]_{ij}, & \text{if } x_i > 0 \text{ and } x_j > 0 \\ \delta_{ij}, & \text{otherwise.} \end{cases}$$

After an approximate minimizer p^k of q_k has been computed by the CG iterative method equipped with the stopping rule of [3, 5, 33], a backtracking line search is performed to guarantee that $T(x^{k+1}) < T(x^k)$. In our present implementation, at most 30 CG iterations were allowed per outer GPRN iteration. Expressions for both gradient and Hessian of T for both the emission and transmission cases can be found in the Appendix.

Discussion

When applied to nonnegatively constrained quadratic minimization problems, the GPRN algorithm is very similar to the method proposed in [33]. Step 2 is needed because gradient projection – being the nonnegatively constrained analogue of the steepest descent method [30] – has a possibly excruciatingly slow convergence rate. On the other hand, unlike the nonnegative analogues of conjugate gradient (CG) proposed in [36, 40], the gradient projection method is extremely efficient at identifying the active

set at the solution, i.e. at finding indices i for which $x_i^* = 0$, where x^* is the nonnegative minimizer of T . The separation of the active set identification problem (Step 1) from the implementation of the CG method (Step 2) for nonnegatively constrained quadratic problems, yields an algorithm with active set identification properties at least as good as gradient projection and convergence properties similar to that of CG for unconstrained quadratic minimization. This algorithm seems to be equally efficient when applied to the non-quadratic problem (16). We conclude with the remark that while preconditioning of the CG iterations in Step 2 is possible [5], it is significantly more challenging than in the unconstrained case.

The main computational bottleneck of the GPRN algorithm (and, for that matter, the EM-based methods) is the number of matrix-vector products needed to reach convergence. In fact, while for the EM method, two matrix-vector products were needed per iteration, within the GPRN method, the gradient projection iterations in Step 1 require four matrix-vector products per iteration and an additional product for each line search step and each iteration of the CG method in Step 2 requires two matrix-vector products per iteration. In view of the how computationally expensive each GPRN iteration is, stopping rules play an important role. In our computed examples we follow [33]. It has been shown in [5] that in general the GPRN method requires fewer matrix vector products than the EM algorithm to converge. A comparison between GPRN and the current state of the art PET statistical imaging methods (see, e.g. [1]) is in order and will be the topic of future work.

It was shown in [3] that the GPRN algorithm is convergent provided T in (16) is strictly convex, which we prove to be the case for the applications of interest in the Appendix. As long as the convexity of T is respected, the GPRN algorithm is convergent for non-quadratic, convex regularization functions, such as obtained from the more general Gibbs distributions [24, 28] as well as when total-variation regularization is used with the negative-log Poisson likelihood [2].

5. Computed examples

In this section, we test our methodology on both emission and transmission PET examples. First we consider an emission PET example similar the canonical one of [39]. In particular, we suppose that the emission density λ is given by the object on the upper-left in Figure 1, while the attenuation vector μ is assumed to be zero. Data generated using (5) with $\beta_i = 1$ for all i and Poisson noise generated using MATLAB's `poissrnd` function is plotted on the upper-right in Figure 1. We have assumed a 128×128 computational grid and a sensor geometry defined by 128 angles and 128 sensor pairs per angle. Thus $M = N = 128^2$ in (2), (3).

To reconstruct the phantoms, we applied fifteen iterations of the iterative MAP algorithm with initial guesses $x_j^0 = 1$ and $\theta_j^0 = \theta_0$ for $j = 1, \dots, N$. The GPRN algorithm was stopped as soon as the norm of the projected gradient was reduced by six orders of magnitude – a standard measure of convergence for constrained optimization

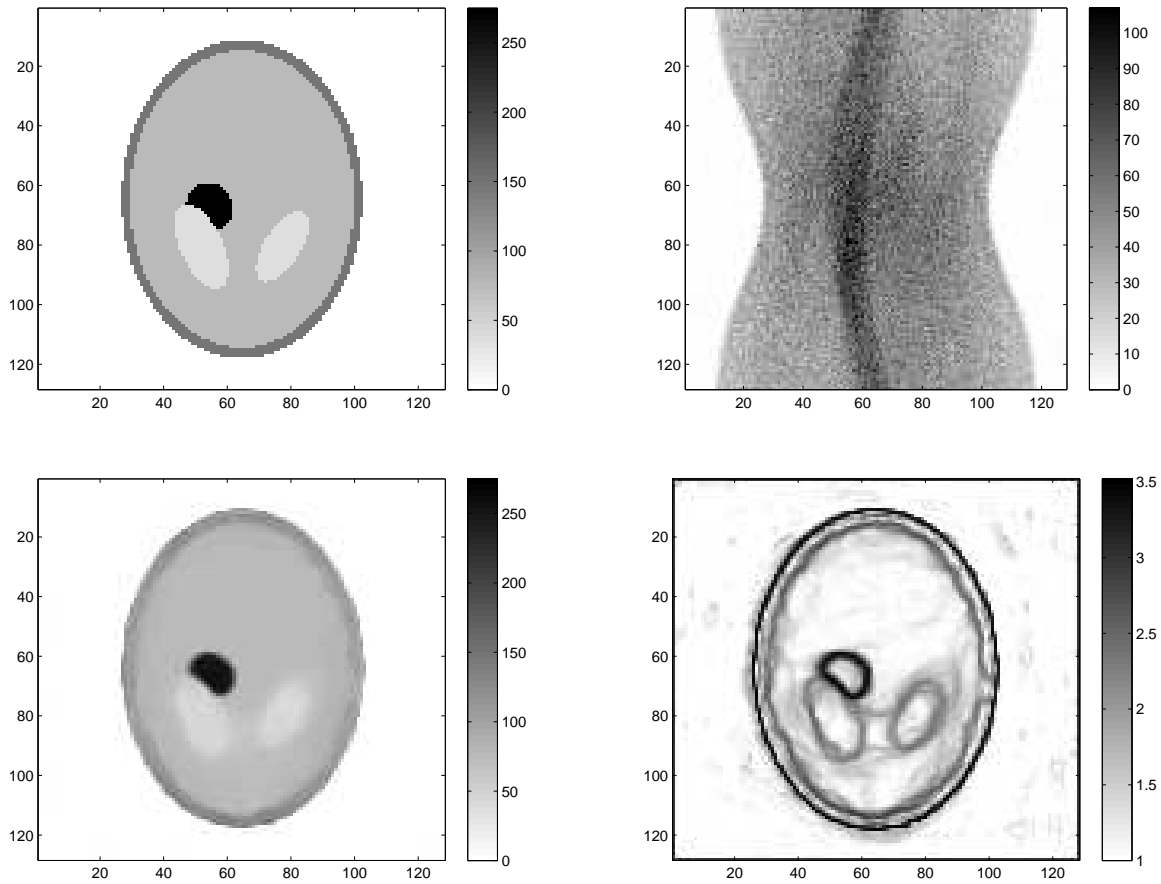


Figure 1. On the upper-left is the emission density phantom, and on the upper-right is the corresponding emission sinogram data. On the lower-left is the reconstructed emission density vector x , and on the lower-right is the reconstructed log-variance vector $\ln(\theta)$.

algorithms [30]. In the gamma hyperprior, we set $\alpha = 2.01$ and $\theta_0 = 1000$, which corresponds to a approximate regularization parameter value of $\sqrt{2/\theta_0} = 0.0447$. The selection of the values of the parameters of the hyperprior was rather heuristic, but since the algorithm is sensitive to the value of θ_0 , it is necessary to design solid selection criteria. The MAP estimation of the emission density vector x and of its variance vector θ are displayed in the bottom row of Figure 1.

In our second example, we generate synthetic transmission data using model (7) with $\beta_i = 1$ for all i and the phantom from our first example. The blankscan (b_0 in (7)) is defined $b_0(x, y) = 90(1 + x^2)$ on $-1 \leq x, y \leq 1$, which is the assumed domain of the computational grid. The phantom and transmission sinogram are displayed in the top row of Figure 2. To compute the MAP estimate of the attenuation vector x we apply the iterative MAP algorithm as in the previous example, except that we set $\theta_0 = 1$ and the GPRN algorithm was stopped as soon as a decrease of 5 orders of magnitude in the projected gradient was attained. The computed results are shown in the bottom row of

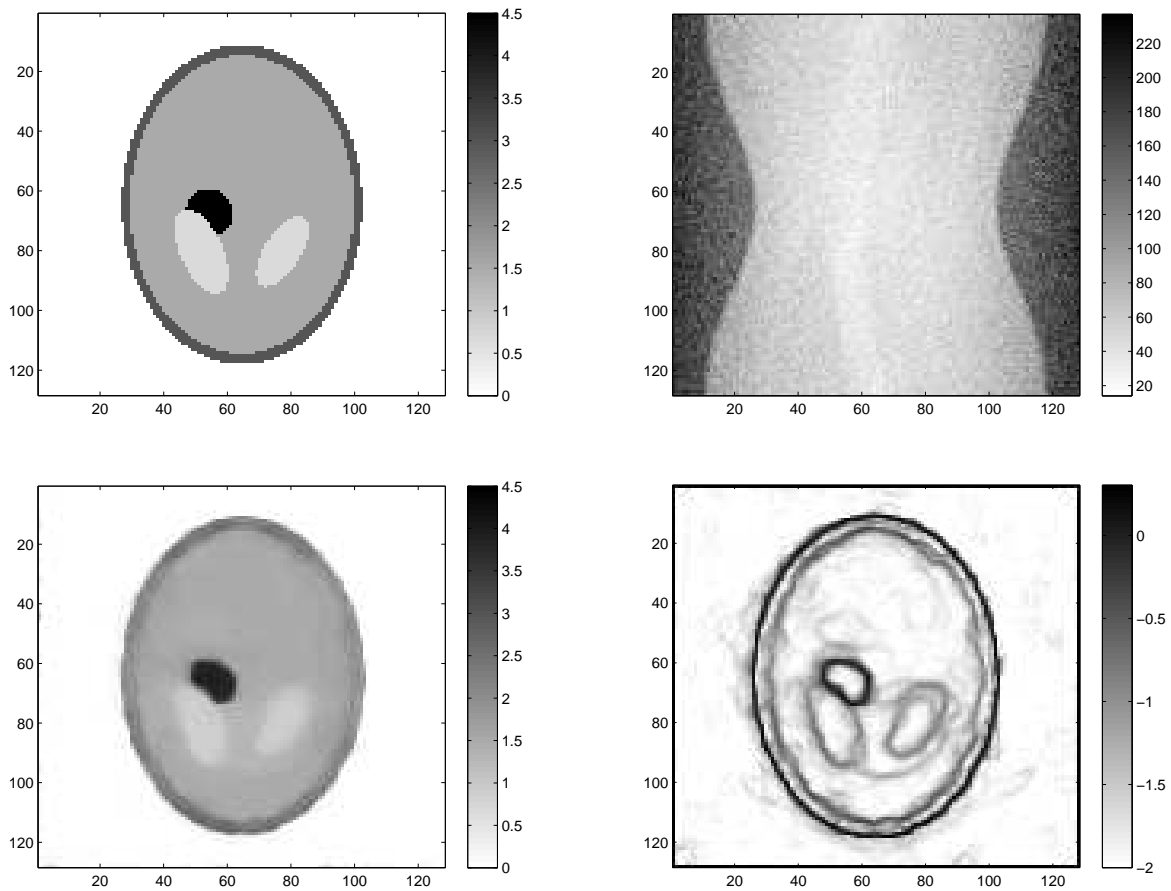


Figure 2. On the upper-left is the transmission density phantom, and on the upper-right is the corresponding transmission sinogram data. On the lower-left is the reconstructed emission density vector x , and on the lower-right is the reconstructed log-variance vector $\ln(\theta)$.

Figure 2.

To illustrate the performance and convergence rate of the algorithm, the plot in Figure 3 shows the relative error,

$$e_k = \|x_{\text{true}} - x_k\| / \|x_{\text{true}}\|,$$

at each iterative MAP outer iteration of the emission PET (left) and transmission PET (right).

In our last computed example we used real transmission PET data from [18], which we assume follows the statistical model (7) with $A = A^{\text{trans}}$. We performed our estimation on a 128×128 uniform computational grid, and are given the sensor geometry, which consisted of 160 sensors pairs per 192 angles. Since we had no information about the β_i 's, we assumed that they were zero. The blank scan b_0 downloaded from [18] is shown in Figure 4 along with the corresponding blurred, noisy sinogram. We reconstruct the image using the same implementation of the iterative MAP algorithm as in the previous example. The reconstructed image and corresponding variance array θ are plotted in

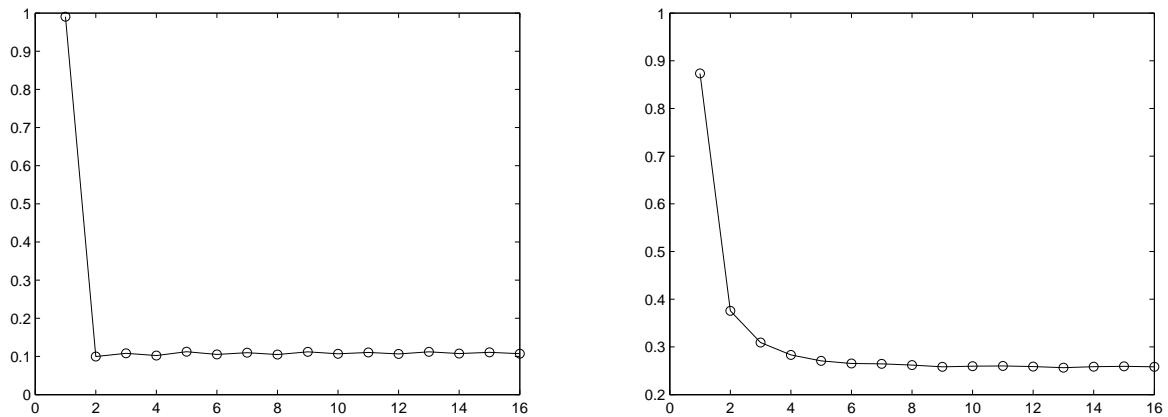


Figure 3. A plot of the relative error per iterative MAP algorithm iteration for the synthetic emission PET example (left) and the synthetic transmission PET example (right).

the bottom row of Figure 4.

The computing time required per outer iteration, in our examples and using our MATLAB implementation, varied between 9 and 30 seconds.

6. Discussion and conclusions

The article proposes a computational method for estimating absorption and emission densities of an unknown body from weak transmission signals consisting of random photon counts. The prior information that the object consists of blocky structures is implemented as a hierarchical conditionally Gaussian prior model. The conditional normality of the prior is an attractive way of including available information as it makes it possible to design computationally efficient algorithms.

In the present work, the Bayesian framework is used to design an iterative algorithm for computing an approximation of the maximum a posteriori estimate. As the solution of the inverse problem in the Bayesian framework is not a single estimate but in fact the posterior probability distribution, the present work does not exploit the full potential of the approach. Important additional information, such as credibility bounds for the object boundaries could also be addressed in this context, using Markov Chain Monte Carlo (MCMC) methods [12]. The design of an efficient variant of MCMC sampling methods for PET problems and of visually suggestive ways to display the results will be the topic of future research.

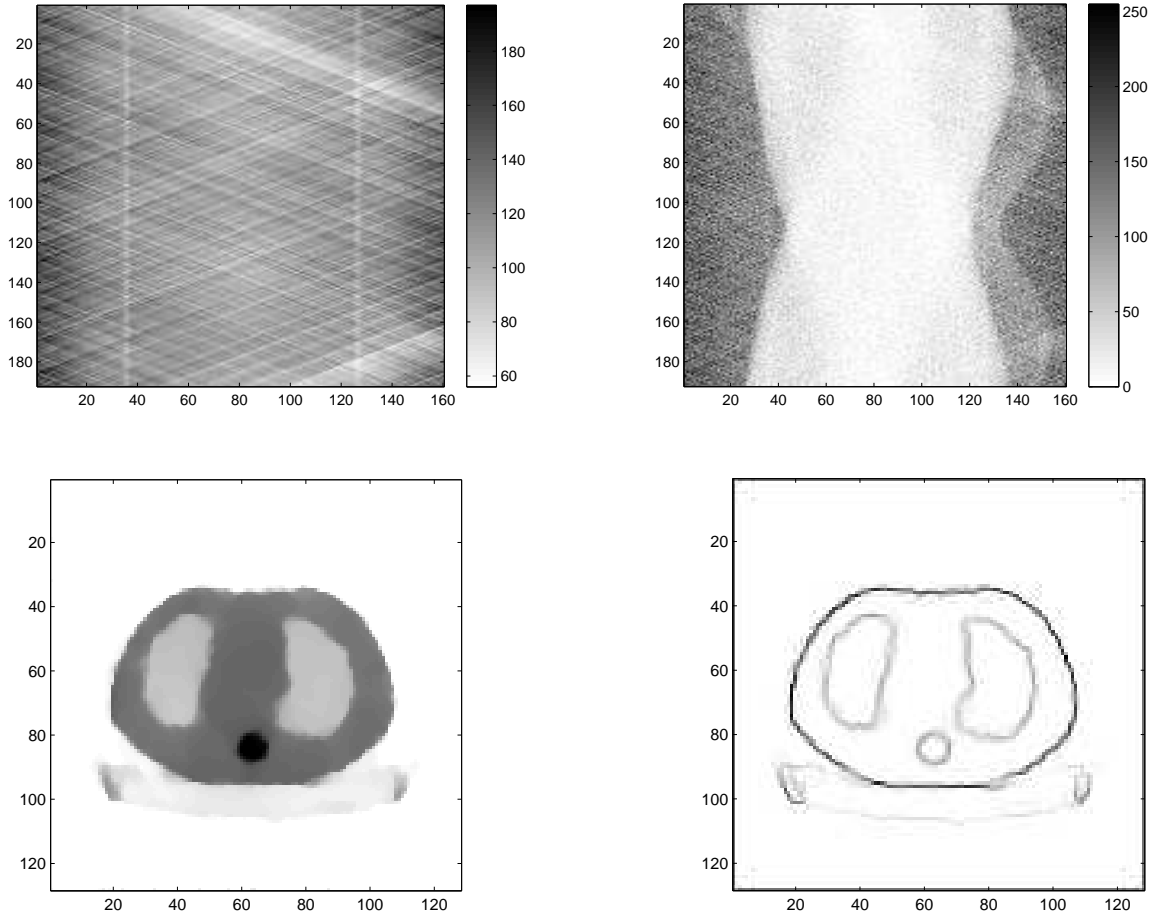


Figure 4. In the upper-left is the blank scan (b_0 in (7)). In the upper-right is the transmission sinogram. In the lower-left is the iterative MAP reconstruction of x . In the lower-right is the reconstructed value of θ .

7. Appendix

In the emission PET case ($\pi_* = \pi_{\text{emiss}}$), we have

$$-\ln \pi_{\text{emiss}}(b | x) \simeq \sum_{j=1}^M \{[Ax]_j + \beta_j - b_j \ln([Ax]_j + \beta_j)\} \quad (21)$$

$$\stackrel{\text{def}}{=} \ell_{\text{emiss}}(b | x); \quad (22)$$

whereas in the transmission PET case ($\pi_* = \pi_{\text{trans}}$), we have

$$-\ln \pi_{\text{trans}}(b | x) \simeq \sum_{j=1}^M \{b_{0,j} e^{-[Ax]_j} + \beta_j - b_j \ln(b_{0,j} e^{-[Ax]_j} + \beta_j)\} \quad (23)$$

$$\stackrel{\text{def}}{=} \ell_{\text{trans}}(b | x). \quad (24)$$

The gradient and Hessian of the regularization function defining T in (16) are given by $C_k x$ and C_k , respectively. The gradient and Hessian of the negative-log likelihood

functions are more involved. In particular, in the emission PET case we have

$$\nabla \ell_{\text{emiss}}(b | x) = \mathbf{A}^\top \left(\frac{\mathbf{A}x - (b - \beta)}{\mathbf{A}x + \beta} \right), \quad (25)$$

$$\nabla^2 \ell_{\text{emiss}}(b | x) = \mathbf{A}^\top \text{diag} \left(\frac{b}{(\mathbf{A}x + \beta)^2} \right) \mathbf{A}, \quad (26)$$

where all vector multiplication and division are intended component-wise, while in the transmission PET case,

$$\nabla \ell_{\text{trans}}(b | x) = \mathbf{A}^\top \left(b_0 e^{-\mathbf{A}x} - \frac{bb_0 e^{-\mathbf{A}x}}{b_0 e^{-\mathbf{A}x} + \beta} \right), \quad (27)$$

$$\nabla^2 \ell_{\text{trans}}(b | x) = \mathbf{A}^\top \text{diag} \left(b_0 e^{-\mathbf{A}x} - \frac{\beta bb_0 \odot e^{-\mathbf{A}x}}{(e^{-\mathbf{A}x} + \beta)^2} \right) \mathbf{A}, \quad (28)$$

where “ \odot ” denotes component-wise product of vectors.

Note that (26) is positive semi-definite for all $x \geq 0$. Hence $\ell_{\text{emiss}}(b | x)$ is convex on the set $x \geq 0$. Since $\ell_{\text{trans}}(b | x)$ is a convex transformation of $\ell_{\text{emiss}}(b | x)$, it is also convex on the set $x \geq 0$. Moreover, since \mathbf{L}_1 and \mathbf{L}_2 are invertible and \mathbf{D}_{θ^k} is positive definite for all k , the regularization matrix \mathbf{C}_k is positive definite for all k . This implies that $\nabla^2 T(x)$, where T is defined in (16), is positive definite for all $x \geq 0$, hence T is strictly convex. Moreover, it can be shown that $\ell_{\text{emis}}(b | x)$ and $\ell_{\text{trans}}(b | x)$ are bounded below. Therefore, since \mathbf{C}_{θ^k} is positive definite, T is a coercive function (c.f. [3]) and hence has a unique minimizer.

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, vol. 2, no. 2, 2008, pp. 167 - 185.
- [3] 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.
- [4] Johathan M. Bardsley and John Goldes, *Regularization Parameter Selection Methods for Ill-Posed Poisson Maximum Likelihood Estimation*, accepted in Inverse Problems.
- [5] J. M. Bardsley and C. R. Vogel, *A Nonnegatively Constrained Convex Programming Method for Image Reconstruction*, SIAM Journal on Scientific Computing, 25(4), 2004, pp. 1326-1343.
- [6] J. Besag, *On the statistical analysis of dirty pictures (with discussions)*, J. Royal Stat. Soc., Ser. B, 48, 1986, pp. 259-302.
- [7] J. Besag, *Towards Bayesian image analysis*, J. of Appl. Stat., 16, 1989, pp. 395-406.
- [8] D. Calvetti, H. Hakula, S. Pursiainen and E. Somersalo, *Conditionally Gaussian hypermodels for cerebral source localization*, arXiv:0811.3185 (November 2008).
- [9] D. Calvetti and E. Somersalo, *Local regularization and Bayesian hypermodels*, Advanced Signal Processing Algorithms, Architectures and Implementations XV, edited by Franklin T. Luk, Proc. of SPIE Vol. 5910 (SPIE, Bellingham WA, 2005) 59100W-1 – 59100W-9.
- [10] D. Calvetti and E. Somersalo, *Gaussian hypermodels and recovery of blocky objects*, Inverse Problems, 23, 2007, pp.733-754.
- [11] D. Calvetti and E. Somersalo, *Introduction to Bayesian Scientific Computing - Ten Lectures on Subjective Computing*. Springer, New York 2007.

- [12] D. Calvetti and E. Somersalo, *Hypermodels in the Bayesian Imaging Framework*, Inverse Problems, **24**, 2008, 034013 (20pp) doi: 10.1088/0266-5611/24/3/034013.
- [13] D. Calvetti, F. Sgallari, and E. Somersalo, *Image inpainting and bootstrap priors*, Image and Vision Computing, **24**, 2006, pp. 782–793.
- [14] G. K. Chanta, N. P. Galatsanos, and A. C. Likas, *Bayesian Restoration Using a New Nonstationary Edge-Preserving Image Prior*, IEEE Transactions on Image Processing, 15(10), pp. 2987-2997, October 2006.
- [15] R.K. Dash, E. Somersalo, M.E. Cabrera and D. Calvetti *An efficient deconvolution algorithm for estimating oxygen consumption during muscle activities*, Comput. Meth. Prog. Biomed. **85**, 2007, pp.247–256.
- [16] B. de Finetti, *Theory of Probability*, Vol. 1, John Wiley & Sons, New York 1974.
- [17] A. P. Dempster, N. M. Laird, and D. B. Rubin, *Maximum Likelihood From Incomplete Data via the EM algorithm*, Journal of the Royal Statistical Society, B, 39, pp. 1-38, 1977.
- [18] Real transmission PET data can be downloaded on Jeffrey Fessler’s web site at www.eecs.umich.edu/~fessler/result/tr/pet,trans,2d,sino/.
- [19] Jeffrey Fessler, *Penalized Weighted Least Squares Image Reconstruction for Positron Emission Tomography*, IEEE Transactions on Medical Imaging, 13(2), pp. 290-300, June 1994.
- [20] 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.
- [21] J. A. Fessler and A. O. Hero, *Space-alternating generalized EM algorithms*, IEEE Transactions on Signal Processing, 42(10), pp. 2664-2677, Oct. 1994.
- [22] P.T. Fox and M.E. Raichle *Focal physiological uncoupling of cerebral blood flow and oxidative metabolism during somatosensory stimulation of human subjects*, Proc. Natl. Acad Sci USA **83**, 1986, pp.1140–1144.
- [23] A. Gelman, *Bayesian Data analysis (2nd ed.)*, CRC Press, Boca Raton 2004.
- [24] S. Geman and D. Geman, *Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images*, IEEE Trans. Pattern Anal. Mach. Intell., **6**, 1984, pp. 721–741.
- [25] 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.
- [26] 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.
- [27] J. Heiskala, I. Nissilä, T. Neuvonen, S. Järvenpää and E. Somersalo, *Modeling anisotropic light propagation in a realistic model of the human head*, Appl. Optics **44**, 2005, pp. 2049–2057.
- [28] 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.
- [29] Jari Kaipio and Erkki Somersalo, *Statistical and Computational Inverse Problems*, Springer 2005.
- [30] C. T. Kelley, *Iterative Methods for Optimization*, SIAM, Philadelphia, 1999.
- [31] K. Lange and R. Carson, *EM reconstruction algorithms for emission and transmission tomography*, Journal of Computer Assisted Tomography, 8, pp. 306-316, 1984.
- [32] S.-J. Lee, A. Rangarajan, and G. Gindi, *Bayesian image reconstruction in SPECT using higher order mechanical models as priors*, IEEE Transactions on Medical Imaging, 14(4), pp. 669-680, 1995.
- [33] J. J. Moré and G. Toraldo, *On the Solution of Large Quadratic Programming Problems with Bound Constraints*, SIAM Journal on Optimization, **1** (1991), pp. 93–113.
- [34] M.A. Mintun MA, A.G. Vlassenko, M.M. Rundle and M.E. Raichle, *Increased lactate/pyruvate ratio augments blood flow in physiologically activated human brain*, Proc. Nat. Acad. Sci. **101**, 2004, pp.659–664.
- [35] 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.
- [36] D. P. O’Leary, *A generalized conjugate gradient algorithm for solving a class of quadratic programming problems*, Linear Algebra and Its Applications **34** (1980), pp. 371-399.
- [37] John M. Ollinger and Jeffrey A. Fessler, *Positron-Emission Tomography*, IEEE Signal Processing Magazine, January 1997.
- [38] M. Sato, T. Yoshika, S. Kajihara, K. Toyama, N. Goda, K. Doya, and N. Kawato, *Hierarchical Bayesian estimation for MEG inverse problem*, NeuroImage, **23**, 2004, pp. 806–826.
- [39] 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.
- [40] A.N. Tikhonov, A. Goncharsky, V.V. Stepanov, A.G. Yagola, *Numerical Methods for the Solution of Ill-Posed Problems*, Springer, 1995.
- [41] Y. Vardi, L. A. Shepp, and L. Kaufman, *A stastical model for positron emission tomography*, Journal of the American Statistical Association, 80, pp. 8-37, 1985.
- [42] A. Voutilainen, V. Kolehmainen, and J. Kaipio, *Statistical inversion of aerosol size measurement data*, Inv. Probl. Eng., **9**, 2001, pp. 67–94.
- [43] Curtis R. Vogel, *Computational Methods for Inverse Problems*, SIAM, Philidelphia, 2002.
- [44] L. Wang, S. L. Jacques and L. Zheng, *MCML – Monte Carlo modeling of light transport in multi-layered tissues*, Computer Meth. Biomed. **47**, 1995, pp. 131–146.
- [45] J. Wang and N. Zabaras, *Hierarchical Bayesian models for inverse problems in heat conduction*, Inverse Problems **21**, 2005, pp. 183–206.
- [46] M. Yavuz and J. A. Fessler, *New statistical models for randoms-precorrected PET scans*, Information Processing in Medical Im., J Duncan and G Gindi, editor. Springer-Verlag, Berlin, pp. 190-203, 1997.
- [47] Daniel Yu and Jeffrey Fessler, *Edge-Preserving Tomographic Reconstruction with Nonlocal Regularization*, IEEE Transactions on Medical Imaging, 21(2), pp. 159-173, 2002.