

# Hierarchical regularization for edge-preserving reconstruction of PET images

Johnathan M. Bardsley<sup>†</sup>, Daniela Calvetti<sup>\*</sup>, and Erkki Somersalo<sup>\*</sup>

**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 reconstruction for both emission and transmission PET problems at very low computational cost.

**Index Terms**—edge-preserving regularization, positron emission tomography, Bayesian statistical methods, nonnegatively constrained optimization

## I. INTRODUCTION

Positron Emission Tomography (PET) is a functional imaging modality that has increased significantly 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 [15], [21] in the context of brain energetics. Moreover, it is used routinely to the 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 process. Two observations are in order: the first is that the signal is, by its very nature, random; and the second is that the inverse problem of estimating the emission density distribution is so severely ill-posed that 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 ([9], [19]).

In the Bayesian framework, randomness is used to express the subject's degree of information, or the lack of it, about the quantities of interest, hence the notion of subjective probability. The randomness of a measured signal is often identified with observation noise, and it is encoded into the likelihood function. Since the likelihood function also contains information about how the measurement and the object of primary interest are related, it should also account for possible modeling mismatch.

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, even when the unknown is traditionally expected to be a well-defined physical quantity. In the Bayesian approach, as expressed by de Finetti [14], the randomness of the unknown is not a property of the object to be recovered, but is rather an expression of the subject's lack of information about it.

It is a particularly challenging task to express prior information that is qualitative in nature in terms of probability distributions. Moreover, there is no unique way to interpret the prior information and to assign the prior probabilities. Therefore, the question whether a prior distribution is correct cannot be answered, and in fact it is meaningless.

Since the classical works on Bayesian imaging ([4], [5], [17]), Bayesian methods have been employed extensively in imaging applications. In particular, hierarchical models have turned out to be useful tools for modeling qualitative prior information, and they have been successfully employed, e.g., in recovering discontinuities in biological signals [12], particle size distributions in aerosol physics [24], temperature distributions in inverse heat conduction [27], blocky images in image deblurring [10], missing information in image inpainting [11], as well as focal sources in EEG/MEG [23], [6].

In this work, we apply the Bayesian methodology to PET images with the prior information that the object is blocky. We consider both emission PET tomography, i.e., when the unknown is the distribution of the decaying marker, and transmission PET tomography, in which case the data consists of transmitted signals through an unknown absorbing density distribution. The latter problem is mathematically equivalent to the low dose X-ray tomography problem and is therefore of significant interest per se. The prior information in the former case corresponds to the assumption that the radioactive marker is absorbed by a target, for example a tumor with well defined boundaries, while in the latter case, we assume a priori a well

<sup>†</sup> Department of Mathematical Sciences, University of Montana. Email: bardsleyj@mso.umt.edu

<sup>\*</sup> Department of Mathematics, Case Western Reserve University. Email: daniela.calvetti@case.edu, erkki.somersalo@tkk.fi

structured object consisting, e.g., of nearly homogenous organs with well defined boundaries.

## II. MODEL OF PET SIGNAL

### A. Mathematical models for PET image formation

The transport and uptake of a substance such as glucose in an organ can be monitored by injecting the subject with a form of glucose containing a radioactive isotope. 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 are detected by two different detectors within a sufficiently short time window they are counted as an event along the connecting line, referred to as the Line Of Response (LOR). Assuming a sufficient number of events along a particular line, an approximation of the integral over the radioisotope distribution is obtained. Taking all such line integral approximations together yields a discrete approximation of the Radon transform of a cross section of the radioisotope distribution, which can be inverted to obtain an image (for more detail see [22]). Obtaining images in this manner is the basic idea behind positron emission tomography (PET).

PET imaging is routinely combined with anatomical imaging, which can be done with the PET detector itself—the so-called transmission PET problem—or by a computed tomography (CT) device integrated with the PET imaging system. Since the absorption and emission models are closely related, we integrate them here and consider both in the computational framework.

We denote by  $\Omega \subset \mathbb{R}^2$  the bounded domain containing the intersection of the object of interest. Let  $\mu : \Omega \rightarrow \mathbb{R}$  and  $\lambda : \Omega \rightarrow \mathbb{R}$  be two non-negative functions representing the *mass absorption* and *photon emission densities*, respectively. It is common to assume that the scattering of photons is negligible in comparison to the absorption of photons by the material within which the radioisotopes are contained. Let  $L$  denote a line passing through the target, parameterized by length  $s$ ; that is,  $L = \{x = x(s) \in \Omega, 0 \leq s \leq S\}$ . With a slight abuse of notation, we write  $\lambda(s) = \lambda(x(s))$  and  $\mu(s) = \mu(x(s))$ .

Consider first the transmission data: the attenuation of the intensity  $I = I(s)$  of the transmitted signal along the line segment  $ds$  of  $L$  is

$$dI = -I\mu ds, \quad (1)$$

and therefore, if  $I_0$  is the intensity of the source along line  $L$ , the received intensity  $I$  after the source photons have traversed the mass absorption distribution is obtained from the equation

$$-\ln\left(\frac{I}{I_0}\right) = -\int_{I_0}^I \frac{dI}{I} = \int_0^S \mu(s) ds. \quad (2)$$

Consider next the photon emission of the target. If the emission is proportional to the amount of the radioactive marker along the line segment  $ds$ , the contribution to the measured expected intensity must be  $\lambda(s)ds$ . On the other hand, the absorption of radiation due to the mass absorption density is given by (1), so the total change in the intensity is

$$dI = \lambda ds - \mu I ds. \quad (3)$$

We can solve this equation by the method of variation of constants. By writing

$$I(s) = J(s)\exp\left(-\int_0^s \mu(t)dt\right),$$

where  $J(s)$  is the net amount of radiation accumulated along the line segment  $[0, s]$ , and substituting this expression in (3), we find that  $J$  satisfies

$$dJ(s) = \lambda(s)\exp\left(\int_0^s \mu(t)dt\right) ds.$$

Therefore, by integrating from 0 to  $s$ , we find that

$$J(s) = \int_0^s \lambda(r)\exp\left(\int_0^r \mu(t)dt\right) dr,$$

and so the total radiation intensity  $I_+$  at  $s = S$  is

$$\begin{aligned} I_+ &= \left(\int_0^S \lambda(r)\exp\left(\int_0^r \mu(t)dt\right) dr\right) \exp\left(-\int_0^S \mu(t)dt\right) \\ &= \int_0^S \lambda(r)\exp\left(-\int_r^S \mu(t)dt\right) dr. \end{aligned} \quad (4)$$

This formula gives the intensity of the attenuated emission at the end of the line  $L$ .

There is an additional complication due to the identification of the LOR of an emission event, requiring that two photons are recorded almost simultaneously in opposite ends of the line. Using the notation above, the intensity along the line  $L$  at  $s = 0$  is

$$I_- = \int_0^S \lambda(r)\exp\left(-\int_0^r \mu(t)dt\right) dr. \quad (5)$$

Assume now that a photon pair is created at point  $x = x(s)$  of the line  $L$ , and one of the photons is detected, say, at  $s = S$  while the other one traveling in the opposite direction is absorbed. Since the event at  $s = S$  is missing a coincidence, the direction of arrival of this observed photon is unknown and the count needs to be discarded. Therefore, an appropriate model for the expectation of the photon count is, not  $I_+$  nor  $I_-$ , but instead

$$I = \int_0^S g(r)\lambda(r)dr, \quad (6)$$

where

$$g(r) = \min\left\{\exp\left(-\int_0^r \mu(t)dt\right), \exp\left(-\int_r^S \mu(t)dt\right)\right\}.$$

When comparing the last expression to (2), we notice that while the transmission problem leads to the classical Radon transform, the emission PET problem is a modified Radon transform problem where, instead of the line integrals of  $\lambda$ , we know the weighted line integrals (6).

### B. Discretization

In practical computations, we consider a two dimensional discrete model. The image area is modeled by the unit square,  $\Omega = [0, 1] \times [0, 1]$ , which we partition into an  $n \times n$  uniform grid. The mass absorption density  $\mu$  and the photon emission

density  $\lambda$  are approximated by piecewise constant functions, both vanishing outside the object of interest contained in  $\Omega$ . The rays passing through  $\Omega$  are marked with the same index  $j$ ,  $1 \leq j \leq M$ , and denoted by  $L_j$ . The integral of  $\mu$  along the line  $L_j$  is approximated by

$$\int_{L_j} \mu(s) ds \approx \sum_{k=1}^N a_{jk} \mu_k,$$

where  $N = n^2$ ,  $\mu_k$  is the value of  $\mu$  in the  $k$ th pixel and  $a_{jk}$  is the length of the intersection of the line  $L_j$  with the  $k$ th pixel, with the understanding that the length is zero if no intersection occurs. This approximation gives rise to a sparse tomography matrix  $A^{\text{trans}} = [a_{jk}] \in \mathbb{R}^{M \times N}$ .

In the transmission PET problem, the transmission intensity  $I_j^{\text{trans}}$  along the line  $L_j$  is measured directly, so that the model of interest is, from (2),

$$I_j^{\text{trans}} = I_{0,j} \exp \left( - \int_{L_j} \mu(s) ds \right), \quad (7)$$

which in the discrete setting becomes

$$I^{\text{trans}} = I_0 \exp(-A^{\text{trans}} \mu). \quad (8)$$

Here  $I^{\text{trans}} \in \mathbb{R}^M$  is the vector containing the measured intensities, while  $I_0 \in \mathbb{R}^M$  is the vector of source intensities. The source intensities are experimentally determined by a scan without the object, and is referred to in the literature as the *blank scan*.

To discretize the emission model, we use the same discretization for  $\Omega$ , denote by  $\lambda_k$  the value of the emission density  $\lambda$  in  $k$ th pixel, and approximate the integral (6) along the  $j$ th line as

$$I_j^{\text{emiss}} = \int_{L_j} g(r) \lambda(r) dr \approx \sum_{k=1}^N c_{jk} \lambda_k, \quad (9)$$

where  $c_{jk} = a_{jk} g_k$ , with  $g_k$  denoting the value of the function  $g$  in  $k$ th pixel. This leads to a discrete model

$$I^{\text{emiss}} = A^{\text{emiss}} \lambda, \quad (10)$$

where  $A^{\text{emiss}} = [c_{jk}] \in \mathbb{R}^{M \times N}$  is a sparse emission matrix with the same sparsity structure as the matrix  $A^{\text{trans}}$ .

### C. Data-noise model

The 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, we adopt the following notational conventions: random variables are denoted by uppercase letters while their realizations are denoted by lowercase letters. A normal distribution with mean  $x_0$  and covariance matrix  $\Gamma$  is denoted by  $\mathcal{N}(x_0, \Gamma)$ ; a Poisson random variable with Poisson parameter (mean and variance)  $\lambda$  is denoted by  $\text{Poisson}(\lambda)$ ; and “ $\sim$ ” expresses the distribution of a random variable, e.g.,  $X \sim \mathcal{N}(x_0, \Gamma)$ . The symbol “ $\propto$ ” is used to mean “proportional up to an unimportant multiplicative factor”, and  $\simeq$  for “equal up to an additive unimportant constant”.

For the sake of uniformity, we denote by  $y \in \mathbb{R}^M$  the observed intensities  $I^{\text{emiss}}$  in the emission model and  $I^{\text{trans}}$  in the transmission model. Similarly,  $x \in \mathbb{R}^N$  stands for the unknown density  $\lambda$  in the emission tomography problem and  $\mu$  in the transmission problem. The matrices  $A^{\text{emiss}}$  and  $A^{\text{trans}}$  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 [22], the standard statistical model corresponding to the discretized deterministic model (10) is given by

$$B_i \sim \text{Poisson}([Ax]_i + \beta_i), \quad i = 1, \dots, M, \quad (11)$$

where  $\beta_i$  corresponds to detected photons arising from physical processes other than those modeled by (9). Typically, these include random counts as well as counts due to scattering; recall that this was assumed not to occur in our derivation of (9).

Given (11), the probability density function of the observable  $B$  conditioned on  $x$  is given by

$$\pi_{\text{emiss}}(b | x) = \prod_{i=1}^M \frac{([Ax]_i + \beta_i)^{b_i} e^{-([Ax]_i + \beta_i)}}{b_i!}. \quad (12)$$

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

In the transmission PET case [28], the data-noise model has the form

$$B_i \sim \text{Poisson}(b_{0,i} e^{-[Ax]_i} + \beta_i), \quad i = 1, \dots, M, \quad (13)$$

where  $\beta_i$  is as in (11). In this case, the probability density function of  $B$  conditioned on  $x$  is of the form

$$\pi_{\text{trans}}(b | x) = \prod_{i=1}^m \frac{(b_{0,i} e^{-[Ax]_i} + \beta_i)^{b_i} e^{-(b_{0,i} e^{-[Ax]_i} + \beta_i)}}{b_i!}, \quad (14)$$

## III. 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  $b$ :

$$\pi(x | b) \propto \pi_{\text{prior}}(x) \pi_*(b | x). \quad (15)$$

Here  $\pi_*$  is given either by (12) or (14). In our approach, which follows [10], we add another level, assuming that  $\pi_{\text{prior}}$ , and hence  $\pi(x | b)$ , depends upon a hyperparameter  $\theta$  which is assumed to arise from a random variable  $\Theta$  with probability density  $\pi_{\text{hyper}}(\theta)$ . When this is the case (15) becomes of the form

$$\pi(x, \theta | b) \propto \pi_{\text{hyper}}(\theta) \pi_{\text{prior}}(x | \theta) \pi_*(b | x). \quad (16)$$

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

### A. The definition of the prior

To define  $\pi_{\text{prior}}(x | \theta)$  we make the assumption that

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

where the  $\theta_i$ 's are positive for  $i = 1, \dots, N$ ; and  $\mathbf{L}_1 X$  and  $\mathbf{L}_2 X$  are assumed to be independent random vectors, with  $\mathbf{L}_1$  and  $\mathbf{L}_2$  discretized first order partial derivatives with zero boundary condition along the vertical and horizontal axes, respectively, defined on the  $n \times n$  square computational grid. Our prior can then be written in the form

$$\begin{aligned} \pi_{\text{prior}}(x | \theta) &\propto \pi(\mathbf{L}_1 x, \mathbf{L}_2 x | \theta) & (17) \\ &= \pi(\mathbf{L}_1 x | \theta) \pi(\mathbf{L}_2 x | \theta) \\ &= \frac{\det(\mathbf{D}_\theta^{-1})}{(2\pi)^n} e^{-\frac{1}{2} x^T (\mathbf{L}_1^T \mathbf{D}_\theta^{-1} \mathbf{L}_1 + \mathbf{L}_2^T \mathbf{D}_\theta^{-1} \mathbf{L}_2) x} \\ &= \frac{1}{(2\pi)^N} \left( e^{-\frac{1}{2} x^T (\mathbf{L}_1^T \mathbf{D}_\theta^{-1} \mathbf{L}_1 + \mathbf{L}_2^T \mathbf{D}_\theta^{-1} \mathbf{L}_2) x} \right. \\ &\quad \left. \cdot e^{-\sum_{j=1}^N \log \theta_j} \right). & (18) \end{aligned}$$

It should be pointed out that the proportionality constant in (17) cannot be ignored because it depends upon the  $\theta_j$ 's; its computation in two-dimensions is a computationally challenging problem. This motivates the following change of variables: let

$$H = \mathbf{L}_1 X, \quad V = \mathbf{L}_2 X. \quad (19)$$

and replace (16) by

$$\pi(h, v, \theta | b) \propto \pi_{\text{hyper}}(\theta) \pi_{\text{prior}}(h, v | \theta) \pi_*(b | h, v). \quad (20)$$

Note that in this case, from (18),

$$\pi_{\text{prior}}(h, v | \theta) = \frac{1}{(2\pi)^N} e^{\sum_{j=1}^N \left\{ -\frac{v_j^2 + h_j^2}{2\theta_j} - \log \theta_j \right\}}. \quad (21)$$

To complete the construction of the prior, we need to choose a hyperprior  $\pi_{\text{hyper}}(\theta)$  conveying our belief about the  $\theta_j$ 's, which can be summarized as follows.

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

Given this, a natural candidate is the *gamma distribution*, defined as follows: if  $\theta_j \sim \text{Gamma}(\alpha, \theta_0)$ ,

$$\pi_{\text{hyper}}(\theta) \propto \prod_{j=1}^N \theta_j^{\alpha-1} \exp\left(-\frac{\theta_j}{\theta_0}\right). \quad (22)$$

Another candidate that will be discussed below is the *inverse gamma distribution*: if  $\theta_j \sim \text{invGamma}(\alpha, \theta_0)$ ,

$$\pi_{\text{hyper}}(\theta) \propto \prod_{j=1}^N \theta_j^{-\alpha-1} \exp\left(-\frac{\theta_0}{\theta_j}\right). \quad (23)$$

Both of these distributions are known to favor rare outliers,

which in the current model corresponds to favoring few but pronounced jumps in the reconstructed image.

### B. Definition of the likelihood

Next, we must rewrite the likelihood function in terms of the new variables. Note that if

$$\begin{bmatrix} H \\ V \end{bmatrix} = \begin{bmatrix} \mathbf{L}_1 \\ \mathbf{L}_2 \end{bmatrix} X \stackrel{\text{def}}{=} \mathbf{M} X,$$

and we compute the QR-decomposition of  $\mathbf{M}$ ,

$$\mathbf{M} = \mathbf{Q}\mathbf{R} = [\mathbf{Q}_1 \ \mathbf{Q}_2] \begin{bmatrix} \mathbf{R}_1 \\ 0 \end{bmatrix},$$

then

$$\mathbf{R}_1 X = \mathbf{Q}_1^T \begin{bmatrix} H \\ V \end{bmatrix}, \quad \mathbf{Q}_2^T \begin{bmatrix} H \\ V \end{bmatrix} = 0.$$

The equation on the left expresses the image  $X$  in terms of the increments, while that on the right is a compatibility condition (see [10]) stating that if the increment vectors  $V$  and  $H$  come from an image, the circulation around a vertex where four pixels meet has to vanish [8]. The likelihood can then be written

$$\pi(b | v, h) \propto \delta\left(\mathbf{Q}_2^T \begin{bmatrix} v \\ h \end{bmatrix}\right) \pi_*\left(b \mid \mathbf{R}_1^{-1} \mathbf{Q}_1^T \begin{bmatrix} h \\ v \end{bmatrix}\right),$$

where the Dirac delta  $\delta$  ensures that the support of the density is in the subspace perpendicular to the range of  $\mathbf{Q}_2$ ; and  $\pi_*$  is given either by (12) or (14). Note that, philosophically, the delta term constitutes a part of the prior because it is independent of the observation  $b$ .

## IV. ITERATIVE MAP COMPUTATION

The MAP estimator is the minimizer, with respect to  $(h, v, \theta)$ , of the negative-log of the posterior density

$$\begin{aligned} -\ln(\pi(h, v, \theta | b)) &\simeq -\ln(\pi_*(b | h, v)) + \frac{1}{2} \sum_{j=1}^N \frac{v_j^2 + h_j^2}{\theta_j} \\ &\quad + \sum_{j=1}^N \frac{\theta_j}{\theta_0} - (\alpha - 2) \sum_{j=1}^N \log \theta_j, \end{aligned}$$

$$\text{such that } \mathbf{L}_1^{-1} h, \mathbf{L}_2^{-1} v \geq 0 \text{ and } \mathbf{Q}_2^T \begin{bmatrix} h \\ v \end{bmatrix} = 0.$$

Although gradient-based algorithms could be used to solve this problem, we propose a simple cyclic iterative algorithm that has been found to be quite effective [10]. The outline of the algorithm is as follows:

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

**Step 1.** Update the estimate of the increments:

$$\begin{aligned} (h^k, v^k) &= \arg \min_{h, v} -\ln(\pi(h, v, \theta^{k-1} | b)), \\ \text{such that } &\mathbf{L}_1^{-1} h, \mathbf{L}_2^{-1} v \geq 0, \mathbf{Q}_2^T \begin{bmatrix} h \\ v \end{bmatrix} = 0. \end{aligned} \quad (24)$$

**Step 2.** Update the estimate of the  $\theta$  vector:

$$\theta^k = \arg \min_{\theta} -\ln(\pi(h^k, v^k, \theta | b)). \quad (25)$$

**Step 3.** Increase  $k$  by one and return to Step 2. Repeat until convergence.

A. *Step 1 of iterative MAP algorithm, increment  $((v, h))$  update*

It can be shown that solving problem (24) is equivalent to first solving

$$\arg \min_{x \geq 0} \left\{ T(x) \stackrel{\text{def}}{=} -\ln(\pi_*(b | x)) + \frac{1}{2} x^T C_k x \right\}, \quad (26)$$

where  $C_k = L_1^T D_{\theta_{k-1}} L_1 + L_2^T D_{\theta_{k-1}} L_2$ , and then setting  $v = L_1 x$ ,  $h = L_2 x$ .

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

$$\begin{aligned} -\ln \pi_{\text{emis}}(b | x) &\simeq \sum_{i=1}^M \{ [Ax]_i + \beta_i - b_i \ln([Ax]_i + \beta_i) \} \\ &\stackrel{\text{def}}{=} \ell_{\text{emis}}(b | x); \end{aligned} \quad (27)$$

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

$$\begin{aligned} -\ln \pi_{\text{trans}}(b | x) &\simeq \sum_{i=1}^M \left\{ b_{0,i} e^{-[Ax]_i} + \beta_i \right. \\ &\quad \left. - b_i \ln(b_{0,i} e^{-[Ax]_i} + \beta_i) \right\} \\ &\stackrel{\text{def}}{=} \ell_{\text{trans}}(b | x). \end{aligned} \quad (28)$$

1) *Existence and uniqueness of solutions for (26)*: The computational method that we will use for solving (26) requires the gradient and Hessian of the cost function  $T$  in (26). The gradient and Hessian of the regularization function 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, we have, in the emission PET case,

$$\nabla \ell_{\text{emis}}(b | x) = A^T \left( \frac{Ax - (b - \beta)}{Ax + \beta} \right), \quad (29)$$

$$\nabla^2 \ell_{\text{emis}}(b | x) = A^T \text{diag} \left( \frac{b}{(Ax + \beta)^2} \right) A, \quad (30)$$

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

$$\begin{aligned} \nabla \ell_{\text{trans}}(b | x) &= \\ &A^T \left( b_0 \odot e^{-Ax} - \frac{b \odot b_0 \odot e^{-Ax}}{b_0 \odot e^{-Ax} + \beta} \right), \end{aligned} \quad (31)$$

$$\begin{aligned} \nabla^2 \ell_{\text{trans}}(b | x) &= \\ &A^T \text{diag} \left( b_0 \odot e^{-Ax} - \frac{\beta \odot b \odot b_0 \odot e^{-Ax}}{(e^{-Ax} + \beta)^2} \right) A, \end{aligned} \quad (32)$$

where “ $\odot$ ” denotes Hadamard (component-wise) multiplication.

We note that both (30) and (32) are positive semi-definite for all  $x \geq 0$ . Moreover, since  $L_1$  and  $L_2$  are invertible and  $D_{\theta^k}$  is positive definite for all  $k$ , the regularization matrix  $C_k$  is positive definite for all  $k$ . This implies that  $\nabla^2 T(x)$ , where  $T$  is defined in (26), 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  $C_{\theta^k}$  is positive definite,  $T$  is a coercive function (c.f. [2]) and  $T$  has a unique minimizer over  $\mathcal{C}$ . In the next section, we outline a numerical method for computing this solution.

2) *A nonnegatively constrained, convex programming method*: We now present the computational method used for solving (26). It was originally introduced in [3] in the context of standard Tikhonov regularization for negative-log Poisson likelihood estimation. More recently, in [2], the method was shown to be convergent for a class of problems that includes—given our discussion in the previous section—problem (26).

Here we give a brief description of the method, and refer the interested reader to [2], [3] for more detail. An outline for the method, which we refer to as *gradient projection–reduced Newton*, follows:

#### GPRN Algorithm:

**Step 0.** Set  $k = 0$  and choose an initial guess  $x^0 \geq 0$ .

**Step 1.** Compute gradient projection iterations until certain stopping criteria are satisfied. Output the updated  $x^k \geq 0$ .

**Step 2.** Compute a quadratic Taylor series approximation of  $T$  based at  $x^k$  restricted to indices  $r$  such that  $x_r^k > 0$ . Use conjugate gradient iterations to compute an approximate minimizer of this function until certain 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.** Set  $k = k + 1$  and return to Step 1.

The gradient projection iteration [20] used in Step 1 of GPRN is computed as follows: given  $x^j \geq 0$ , we compute  $x^{j+1}$  via

$$p^j = -\nabla T(x^j), \quad (33)$$

$$\lambda^j = \arg \min_{\lambda > 0} T(\mathcal{P}(x^j + \lambda p^j)), \quad (34)$$

$$x^{j+1} = \mathcal{P}(x^j + \lambda^j p^j). \quad (35)$$

Here  $\mathcal{P}(x) = \max\{x, 0\}$  with the maximum computed component-wise. In practice, subproblem (34) is solved inexactly using a projected backtracking line search algorithm; see [2] for details. The gradient projection iterations are well-defined and convergent for problems of the form (26) [20, Section 5.4]. The stopping rules for the gradient projection iterations in our computed examples within Step 1 of GPRN are as in [2], [3].

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

$$q_k(p) = T(x^k) + \langle \nabla_{\text{red}} T(x^k), p \rangle + \frac{1}{2} \langle \nabla_{\text{red}}^2 T(x^k) p, p \rangle, \quad (36)$$

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 using the CG method, following [2], [3], a line search is performed in order to guarantee that  $J(x^{k+1}) < J(x^k)$ .

Before continuing, we note that GPRN has been shown to be very computationally efficient for regularized negative-log Poisson likelihood minimization problems. Extensive com-

parisons between it and other state of the art nonnegatively constrained minimization algorithms can be found in [3], where it is shown to be the most efficient of the algorithms considered.

### B. Step 2 of iterative MAP algorithm, variance ( $\theta$ ) update

The updated  $\theta_j$ , computed using the quadratic formula, is

$$\theta_j = \theta_0 \left( \eta + \sqrt{\frac{v_j^2 + h_j^2}{2\theta_0} + \eta^2} \right), \quad \eta = \frac{1}{2}(\alpha - 2). \quad (37)$$

## V. COMPUTED EXAMPLES

In this section, we apply our methodology to two PET examples. The first example is synthetic and will use the emission PET model (11) and negative-log likelihood function (27). In the second example, we use the transmission PET model (13), negative-log likelihood function (28), and real data made available by Jeffrey Fessler [13].

### A. Synthetically generated data: emission PET

We begin with the synthetically generated, emission PET problem. To avoid using the same forward model to generate the data that is used in the inverse solver, we produce the synthetic data by a Monte Carlo simulation. This simulation can also be used to validate the continuous model (6) and its discretized version.

We generate a test phantom of size  $128 \times 128$ . In Figure 1, the absorption density  $\mu$  and the emission density  $\lambda$  are plotted. To generate the photon counting data, we proceed as follows:

**Step 1.** Draw the emission point  $P \sim \text{Uniform}(D)$ , where  $D$  is the activity region in  $\Omega$ .

**Step 2.** Draw a random angle  $\Theta \sim \text{Uniform}([0, 2\pi])$  defining the direction  $\hat{e}$  of the photon propagation.

**Step 3.** Draw the photon free paths  $\ell_+$  and  $\ell_-$  in the directions  $\hat{e}$  and  $-\hat{e}$ , respectively.

**Step 4.** If the free paths are large enough for both photons to escape  $\Omega$ , register a count, otherwise no count is registered.

Above, only Step 3 defining the free paths follows the ideas in [26], [18]: If  $p$  is the realization of the random variable  $P$  of Step 1 above, we denote

$$m(s) = \mu(p + s\hat{e}), \quad s \geq 0.$$

In the continuous model representing the expectation of the photon count, the intensity of the radiation propagating from  $p$  to  $p + s\hat{e}$  is attenuated by the factor

$$a(s) = \exp\left(-\int_0^s m(s')ds'\right).$$

If  $N$  photons are emitted at  $p$  into the direction  $\hat{e}$ , the photon count at  $p + s^*\hat{e}$  should be the integer part of  $Na(s^*)$ , leading to the conclusion that the probability that a photon's free path is at least  $s^*$  must be  $a(s^*)$ ,

$$\text{P}\{s \geq s^*\} = a(s^*).$$

In particular, this implies that

$$\text{P}\{s \leq s^*\} = 1 - a(s^*).$$

Denoting by  $p(s)$  the probability density of the free path and by  $\Phi(s)$  the corresponding cumulative distribution, we find that

$$\Phi(s^*) = \int_0^{s^*} p(s)ds = \text{P}\{s \leq s^*\} = 1 - a(s^*).$$

Hence, the drawing from the density  $p$ , following the inverse cumulative distribution rule ([9]), can be done as follows:

$$\text{Draw } \xi \sim \text{Uniform}([0, 1]), \quad s = \Phi^{-1}(\xi).$$

To solve for the free path  $s$ , we need to consider the equation

$$\Phi(s) = 1 - a(s) = \xi \Rightarrow a(s) = 1 - \xi = t \sim \text{Uniform}([0, 1]),$$

or, equivalently,

$$\int_0^s m(s')ds' = \log(1/t),$$

which can be solved numerically for  $s$ . In the present application, however, we do not need to solve the free path of the photons; it is sufficient to check if the free path is large enough to reach the receiver. Denoting by  $S$  the maximum value of the line length parameter, the condition for the photon to arrive at the receiver is

$$\int_0^S m(s')ds' < \log(1/t),$$

otherwise the photon is absorbed. To guarantee the coincidence necessary for the LOR identification, the same check has then to be done with the photon traveling in the opposite direction, leading to rejection of both photons if any one of them fails to reach the receiver.

In practice, we draw the emission point by randomly picking a pixel from the activity area  $D$ , and the integrals are approximated in the same fashion as in the emission and transmission matrices. In Figure 1 we have plotted the emission sinogram obtained by simulating 100 000 positron emissions from the active region  $D$  indicated by dots in Figure 1 with the absorption density shown in the same figure. In our simulation, about 10% of emitted positrons do not result in a coincidence. To solve the inverse problem, we calculate the matrix  $A^{\text{emiss}}$  in the  $128 \times 128 = 16\,384$  grid, assuming that the absorption density  $\mu$  is known. The number of projection angles is 192 and the number of parallel rays per illumination direction is 160, leading to  $M = 30\,720$  data components. The size of the data is the same as the size of the transmission data made available in [13]. The hyperprior model for the variance  $\Theta$  in this example is the inverse gamma distribution. Figure 2 shows the MAP reconstruction obtained with the optimization algorithm with four outer iterations, i.e., alternating updates of the image and the variance vector. The computation time per outer iteration on a dual processor PC (2.5 GHz, 4GB) implemented with Matlab varies between 2 and 7 seconds.

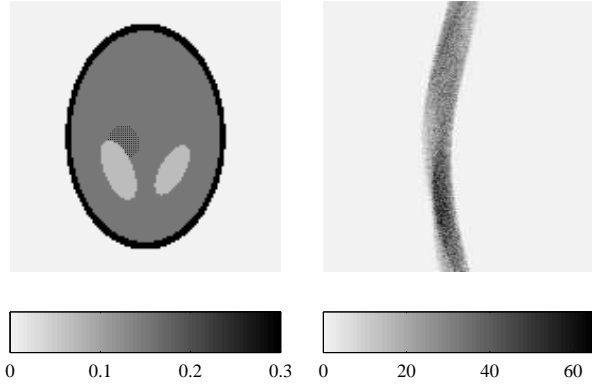


Fig. 1. The absorption phantom (left), with the emission area marked with dots. The Monte Carlo simulation of the emission data (right).

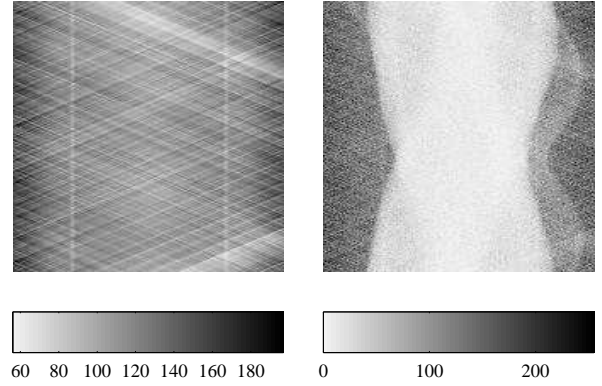


Fig. 3. The blank scan data (left) and the sinogram data of the phantom (right).

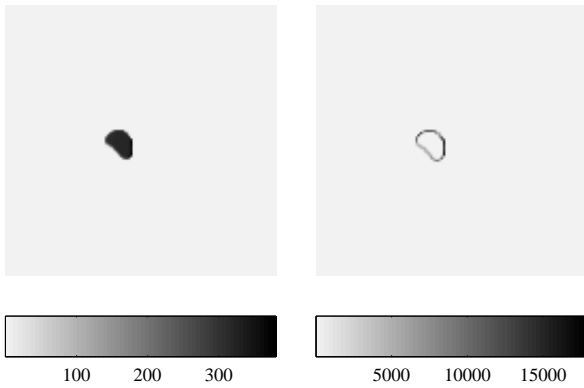


Fig. 2. The reconstructed emission intensity (left) and the variance parameter (right) after four alternating iterations.

### B. Real data: transmission PET

We end with an example computed using real transmission PET data [13]. Thus we assume that the data arises from the statistical model (13) with  $A = A^{\text{trans}}$ . Again, we assume a  $128 \times 128$  uniform computational grid, and 160 sensors and 192 angles defining the PET machine. Since we had no information about the  $\beta_i$ 's, we assumed that they were zero. Finally, the blank scan  $b_0$ , which is computed off line, is shown in Figure 3 with the corresponding blurred, noisy sinogram. We reconstruct the image using 7 iterations of the iterative MAP estimation algorithm outlined in Section IV. In this computed example we used the gamma hyperprior for the variance  $\Theta$ . The reconstructed image and corresponding variance array  $\theta$  are plotted in Figure 4. The computation time per outer iteration in this example varies between 9 and 30 seconds.

## VI. DISCUSSION AND CONCLUSIONS

The article proposes an efficient 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

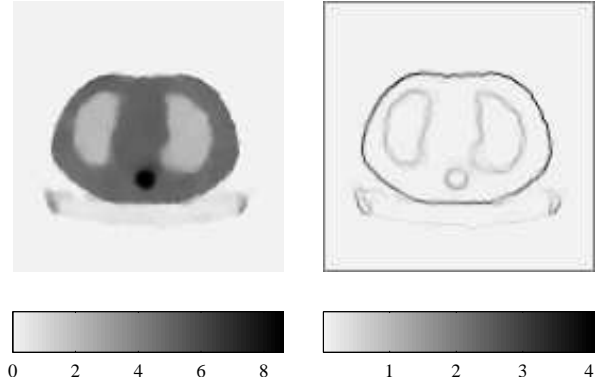


Fig. 4. Estimated absorption distribution corresponding to the data of Figure 3 (left) and the estimated prior variance (right).

the prior is an attractive way of including the prior information as it makes possible the design of an iterative algorithm with a relatively simple objective function. Unlike the popular total variation (TV) algorithms see, e.g., [1], [25], the objective function does not require additional regularization to compensate for the non-differentiability. Moreover, the TV-based methods are more computationally demanding. The connection between the TV method and the hyperprior approach has been discussed in [10].

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 only a single estimate but rather the whole 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 can also be addressed in this context, using Markov Chain Monte Carlo (MCMC) methods [10]. 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.

## VII. ACKNOWLEDGEMENTS

## REFERENCES

- [1] 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.
- [2] Johnathan M. Bardsley and John Goldes, *An Iterative Method for Edge-Preserving MAP Estimation when Data-Noise is Poisson*, submitted to SIAM Journal on Scientific Computing. University of Montana, Department of Mathematical Sciences Technical Report #26, 2008.
- [3] 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.
- [4] J. Besag, *On the statistical analysis of dirty pictures (with discussions)*, J. Royal Stat. Soc., Ser. B, **48**, 1986, pp. 259–302.
- [5] J. Besag, *Towards Bayesian image analysis*, J. of Appl. Stat., **16**, 1989, pp. 395–406.
- [6] D. Calvetti, H. Hakula, S. Pursiainen and E. Somersalo, *Conditionally Gaussian hypermodels for cerebral source localization*, arXiv:0811.3185 (November 2008).
- [7] 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.
- [8] D. Calvetti and E. Somersalo, *Gaussian hypermodels and recovery of blocky objects*, Inverse Problems, **23**, 2007, pp.733–754.
- [9] D. Calvetti and E. Somersalo, *Introduction to Bayesian Scientific Computing - Ten Lectures on Subjective Computing*. Springer, New York 2007.
- [10] 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.
- [11] D. Calvetti, F. Sgallari, and E. Somersalo, *Image inpainting and bootstrap priors*, Image and Vision Computing, **24**, 2006, pp. 782–793.
- [12] 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.
- [13] 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/](http://www.eecs.umich.edu/~fessler/result/tr/pet_trans_2d_sino/).
- [14] B. de Finetti, *Theory of Probability*, Vol. 1, John Wiley & Sons, New York 1974.
- [15] 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.
- [16] A. Gelman, *Bayesian Data analysis (2nd ed.)*, CRC Press, Boca Raton 2004.
- [17] 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.
- [18] 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.
- [19] Jari Kaipio and Erkki Somersalo, *Statistical and Computational Inverse Problems*, Springer 2005.
- [20] C. T. Kelley, *Iterative Methods for Optimization*, SIAM, Philadelphia, 1999.
- [21] 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.
- [22] John M. Ollinger and Jeffrey A. Fessler, *Positron-Emission Tomography*, IEEE Signal Processing Magazine, January 1997.
- [23] 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.
- [24] A. Voutilainen, V. Kolehmainen, and J. Kaipio, *Statistical inversion of aerosol size measurement data*, Inv. Probl. Eng., **9**, 2001, pp. 67–94.
- [25] Curtis R. Vogel, *Computational Methods for Inverse Problems*, SIAM, Philadelphia, 2002.
- [26] 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.
- [27] J. Wang and N. Zabaras, *Hierarchical Bayesian models for inverse problems in heat conduction*, Inverse Problems **21**, 2005, pp. 183–206.
- [28] 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.