Techniques for regularization parameter and hyper-parameter selection in PET and SPECT Imaging

Johnathan M. Bardsley† and John Goldes†

(Received 00 Month 200x; in final form 00 Month 200x)

Penalized maximum likelihood methods are commonly used in positron emission tomography (PET) and single photon emission computed tomography (SPECT). Due to the fact that a Poisson data-noise model is typically assumed, standard regularization parameter choice methods, such as the discrepancy principle or generalized cross validation, can not be directly applied. In recent work of the authors, regularization parameter choice methods for penalized negative-log Poisson likelihood problems are introduced. In this paper, we apply these methods to the applications of PET and SPECT, introducing a modification that improves the performance of the methods. We then demonstrate how these techniques can be used to choose the hyper-parameters in a Bayesian hierarchical regularization approach.

Keywords: positron emission tomography, single photon emission computed tomography, regularization parameter selection, Bayesian statistical methods, Poisson noise

AMS Subject Classification: 65J22, 65K10, 65F22.

1. Introduction

Both positron emission tomography (PET) and single photon emission computed tomography (SPECT) scans provide information about the density of a metabolite (such as glucose) marked with a radioactive isotope in a living organism. In both cases, a radioactive tracer element is injected into the body. The tracer then exhibits radioactive decay, resulting in photon emission. In the case of PET, two photons are emitted in diametrically opposite directions, and if they are detected by the PET machine within a sufficiently short time window they are counted as an event along the line of response (LOR) connecting the two detectors. In the case of SPECT, single photons are emitted and for a detected photon, an event is recorded along a LOR determined using a technique known as columnation. The collection of all events along all LORs during an experiment constitutes a PET, or SPECT, data set. The image reconstruction problem, and the inverse problem of interest, is to reconstruct, from this data, the density of the radioactive metabolite within the organism being imaged.

In [15, 19, 27], mathematical/statistical models for PET and SPECT data formation are presented. We briefly describe these models here. If there are $M$ LORs and $N$ elements in the uniform $\sqrt{N} \times \sqrt{N}$ tracer density grid, we can write the mathematical/statistical model as follows:

$$b = \text{Poisson}(Ax + \gamma),$$

where $b \in \mathbb{R}^M$ is the vector containing the expected number of events along each
of the $M$ LORs; $\text{Poisson}(\lambda)$ denotes a Poisson random vector with mean $\lambda$. $A$ is the $M \times N$ forward model matrix; $x \in \mathbb{R}^N$ is the discrete representation of the (unknown) photon emission density function; and $\gamma$ is the vector containing expected erroneous counts due to accidental coincidences and scattered events [26], which we assume is known. We note that (1) is the dominant error model used in PET [26] and SPECT [15] due to the fact that photon counts tend to be low, making a Gaussian approximation of the statistical model significantly less accurate.

For both PET and SPECT, the forward model in (1) is defined by

$$[Ax]_i = \sum_{j=1}^{n^2} g_{ij} a_{ij}^{\text{Radon}} x_j,$$

where $a_{ij}^{\text{Radon}}$ is the $ij$th element of the Radon transform matrix (i.e., the intersection length of $L_i$ with pixel $j$), and $g_{ij}$ is the discrete impulse response function, which for PET is

$$g_{ij} = \exp \left( - \sum_{k=1}^{n^2} a_{ik}^{\text{Radon}} \mu_k \right),$$

where $\mu_k$ is the absorption density of the subject at pixel $k$, and for SPECT is

$$g_{ij} = \exp \left( - \sum_{k=j}^{n^2} a_{ik}^{\text{Radon}} \mu_k \right).$$

The absorption density $\mu_k$ in (2) and (3) must be estimated beforehand in order to solve the PET and SPECT inverse problems. Estimates of $\mu_k$ can be obtained using, for example, computed tomography or transmission tomography.

Thus the forward model matrix $A$ in (1) is defined by $[A]_{ij} = \exp \left( - \sum_{k=1}^{n^2} a_{ik}^{\text{Radon}} \mu_k \right) a_{ij}^{\text{Radon}}$ for PET, and $[A]_{ij} = \exp \left( - \sum_{k=j}^{n^2} a_{ik}^{\text{Radon}} \right) a_{ij}^{\text{Radon}}$ for SPECT.

Assuming statistical model (1), the probability mass function of the data $b$ conditioned on $x$ is given by

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

The maximum likelihood estimate of the true tracer density $x_*$ given $b$ is obtained by maximizing $p(b \mid x)$ with respect to $x$, subject to the constraint $x \geq 0$. Equivalently, we can solve

$$x_{ML} = \arg \min_{x \geq 0} T_0(x ; b),$$

where

$$T_0(x ; b) = \sum_{i=1}^{M} \left\{ ([Ax]_i + \gamma_i) - b_i \ln([Ax]_i + \gamma_i) \right\}.$$ 

Note that $T_0(x ; b)$ is equal, up to an additive constant, to $-\ln p(b \mid x)$. $x_{ML}$ is
known as the maximum likelihood estimate of $x_\ast$.

However, solving (5) directly yields tracer densities with unrealistic artifacts. Because of this, a regularization term is often added (see, e.g., [1, 12, 13, 15–17, 20, 23, 33]), which can incorporate prior knowledge about the true tracer density. This results in the following modification of (5): compute

$$\mathbf{x}_\alpha = \arg \min_{\mathbf{x} \geq 0} \left\{ T_\alpha (\mathbf{x}) \overset{\text{def}}{=} T_0 (\mathbf{x}; \mathbf{b}) + \frac{\alpha}{2} \mathbf{x}^T \mathbf{C} \mathbf{x} \right\}. \tag{7}$$

Here $\alpha$ is known as the regularization parameter, and $\mathbf{C}$ as the regularization matrix. To guarantee that $T_\alpha$ has a unique nonnegative minimizer, we make the assumption that the null-spaces of $\mathbf{A}$ and $\mathbf{C}$ intersect only trivially [5]. In the PET literature, (7) is known as a penalized maximum likelihood problem, and such problems have been studied extensively; see, e.g., [1, 12, 13, 15–17, 20, 23, 33].

We solve the nonlinear optimization problem (7) using the active set method introduced in [4]. This method intersperses gradient projection and conjugate gradient iterations, yielding a very efficient method for bound constrained, convex optimization problems. We note that provided $\mathbf{A}$ and $\mathbf{C}$ have nonintersecting null-spaces, (7) is strictly convex, and hence, has a unique nonnegative minimizer [6]. Moreover, in [6], it is shown that the method of [4] is guaranteed to converge to this unique minimizer.

One of our primary goals in this paper is to present some methods for choosing the parameter $\alpha$ in (7). But first, to put this work into context for the reader grounded in classical inverse problems, we briefly discuss the underlying continuous analogues of the above models and their classical theoretical analyses.

### 1.1. The connection to classical ill-posed problems

The expert in theoretical inverse problems may wonder how the above discussion relates to classical inverse problems. In that context, the underlying linear operator equation is

$$b = \mathbf{A} \mathbf{x}, \tag{8}$$

where $\mathbf{x}$ (the unknown) and $\mathbf{b}$ (the data) are functions contained in an appropriate function space, and $\mathbf{A}$ is a compact operator, which follows from the fact that the Radon transform is a compact operator [24, 25]. Thus (8) is a classical ill-posed problem.

The standard approach to solving an ill-posed problem is to solve a variational problem of the form

$$\min_{u \geq 0} \left\{ T_\alpha (\mathbf{A} \mathbf{x}; \mathbf{b}) \overset{\text{def}}{=} T_0 (\mathbf{A} \mathbf{x}; \mathbf{b}) + \alpha J (\mathbf{x}) \right\}, \tag{9}$$

where the fit-to-data function is given by $T_0 (\mathbf{A} \mathbf{x}; \mathbf{b}) = \| \mathbf{A} \mathbf{x} - \mathbf{b} \|_2^2$ and $\alpha$ and $J$ are the regularization parameter and functional, respectively. The theoretical analysis of such problems constitutes a significant literature in the field of inverse problems; see, e.g., [28, 31] and the references therein.

Classical analyses have been recently extended to the case of Poisson data [2, 7–9]. Here the functional analogue of the Poisson negative-log likelihood function (6)
is given by

\[ T_0(Ax; b) \overset{\text{def}}{=} \int_{\Omega} ((Ax + \gamma) - b \log(Ax + \gamma)) \, dy, \quad (10) \]

where \( y \) is the independent variable. Note that if \( A \) is invertible, the minimizer of \( T_0 \) is \( x = A^{-1} b \). The analysis found in [2], which takes into account both errors in the data \( b \) and in the operator \( A \), proves that (9), (10) defines a regularization scheme in the classical sense.

The focus of this paper is to present methods for choosing an appropriate value of \( \alpha \) in (7). In the next section, we present three methods, which are extensions of the well-known discrepancy principle, generalized cross validation, and unbiased predictive risk methods for regularization parameter selection in penalized least squares estimation [31]. We then test the methods on several examples in Section 4, and end with conclusions in Section 5.

2. Regularization Parameter Choice Methods

In this paper, we apply the regularization parameter selection methods of [5], which were implemented on applications from astronomical imaging, to the PET and SPECT cases. We note that other methods exist that could also be implemented on our problem; in particular, both the generalized discrepancy principle of [29] and the discrepancy principle of [10].

In [5], a quadratic approximation of \( T_0 \), as defined in (6), is computed and used to extend existing regularization parameter selection methods for least squares problems. The quadratic approximation is given by

\[ T_0(x; b) = T_0(x_e; b) + \frac{1}{2} (Ax - (b - \gamma))^T \text{diag}\left( \frac{1}{b_e} \right) (Ax - (b - \gamma)) + O(\|h\|_2^3; \|h\|_2^2; \|k\|_2^2; \|k\|_2^3), \quad (11) \]

where \( h = x - x_e \), \( k = b - b_e \), and \( b_e = Ax_e + \gamma \). Now we use an application of the mean value theorem to replace \( b_e \) in (11) with \( b_e = Ax_e + \gamma \), where \( x_e \) is computed from (7). Note that in [5] we proceeded similarly, but replaced \( b_e \) by \( b \) instead. We present this modification here because we have found it to be slightly more effective.

The mean value theorem is used by considering the \( i \)th component of \( \frac{1}{b_e - k_e} \) as a function of \( k_{\alpha_i} \), where \( k_{\alpha_i} = b_{\alpha_i} - b_e \). Letting \( r = Ax - (b - \gamma) \), the weighted sum of squares term in (11) can be written as

\[ \frac{1}{2} r^T \left[ r \odot \left( \frac{1}{b_{\alpha_i} - k_{\alpha_i}} \right) \right] = \frac{1}{2} r^T \left[ r \odot \left( \frac{1}{b_{\alpha_i} + \text{diag}\left( \frac{1}{(b_{\alpha_i} - k_{\alpha_i})^2} \right) k_{\alpha_i}} \right) \right] \]

\[ = \frac{1}{2} r^T \left( \frac{r}{b_{\alpha_i}} \right) + \frac{1}{2} r^T \left( \frac{r \odot k_{\alpha_i}}{(b_{\alpha_i} - k_{\alpha_i})^2} \right), \quad (12) \]

where \( \odot \) indicates component-wise multiplication, the square and quotient are taken component-wise, and \( 0 < |k_{\alpha_i}| < |k_{\alpha_i}| \). Noting that \( r = Ah - k \) and that \( b_{\alpha_i} \)
is bounded away from zero, we obtain

\[ r^T \left( \frac{r \odot k}{(b_\alpha - k_\alpha)^2} \right) = O(\|h\|^2\|k_\alpha\|, \|h\|\|k\|\|k_\alpha\|, \|k\|^2\|k_\alpha\|). \] (13)

Recalling (11), we have the approximation

\[ T_0(x; b) = T_0(x_e; b) + T^\text{wls}_0(x; b) \]
+ \( O(\|h\|^2\|k_\alpha\|, \|h\|\|k\|\|k_\alpha\|, \|k\|^2\|k_\alpha\|) \)
+ \( O(\|h\|^2\|k_\alpha\|, \|h\|\|k\|\|k_\alpha\|, \|k\|^2\|k_\alpha\|) \),

where

\[ T^\text{wls}_0(x; b) = \frac{1}{2} \|B_\alpha^{-1/2}(Ax - (b - \gamma))\|^2, \] (15)

with \( B_\alpha = \text{diag}(b_\alpha) \).

This approximation is modified from that of [5] in that the weighting term in (15) is now \( B_\alpha^{-1} \) instead of \( B^{-1} \).

In the three subsections below, we present three regularization parameter choice methods: the discrepancy principle, generalized cross validation (GCV), and unbiased predictive risk. It is likely that an L-curve method can be implemented for Poisson data, but we do not pursue that due to its non-convergence properties for least squares problems [31, 32].

The reader should also be aware that based on a continuous, deterministic analysis, i.e. assuming \( b = Ax + \eta \), where \( b, x \) and \( \eta \) are functions, it is shown in [21, 32] that GCV is non-convergent for ill-posed problems, due to the fact that it does not contain a dependence upon the noise level \( ||\eta|| \). On the other hand, the rigorous analysis of [30, 31], which is based on a partially-discrete model of the form \( b = A_n x + \eta \), where \( A_n \) maps functions into \( R^n \) and \( \eta \) is a white noise (stochastic) vector, proves that GCV is convergent. This partially discretized model can be motivated by the fact that there is no continuous white noise function in \( L^2 \). In any event, if the reader chooses to use GCV he/she should be aware of the assumptions required to guarantee convergence.

### 2.1. Discrepancy Principle

The idea behind the discrepancy principle, as presented in [5], is to choose \( \alpha \) so that \( x_\alpha \) computed from (7) satisfies

\[ T_0(x_\alpha; b) \approx E(T_0(x_e; b)), \] (16)

where \( E \) is the expected value function.

For the right-hand side of (16), from (14), we have

\[ E(T_0(x_e, b)) \approx T_0(x_e; b_e) + E \left( T^\text{wls}_0(x_e; b) \right). \] (17)

It can be argued (see [5]) that \( \|B_\alpha^{-1/2}(Ax_e - (b - \gamma))\|^2 \) is approximately \( \chi^2(M) \)
distributed, and hence that
\[ E\left(T_{wls}^0(x_e; b)\right) \approx M/2. \tag{18} \]

For the left-hand side of (16), from (14), we have
\[ T_0(x_0; b) \approx T_0(x_e; b) + T_{wls}^0(x_0; b). \tag{19} \]
Assuming \( T_0(x_e; b) \approx T_0(x_e; b_e) \), (17), (18), and (19) then imply that (16) will hold provided
\[ T_{wls}^0(x_0; b) \approx M/2. \tag{20} \]

Thus finally, based on (20), we can define the discrepancy principle choice of the regularization parameter:
\[ \alpha_{DP} = \arg \min_{\alpha \geq 0} \left( T_{wls}^0(x_0; b) - M/2 \right)^2, \tag{21} \]
where \( x_0 \) is computed from (7).

Recent work of Mead and Renaut [22] successfully implement a different \( \chi^2 \) criteria in the least squares case. This approach could very likely be extended to the Poisson setting, but we do not pursue that here.

### 2.2. Generalized Cross Validation

In [5], a quadratic approximation of \( T_0 \) is also used to extend the method of generalized cross validation (GCV) for regularization parameter choice in the least squares case [30, 31] to the case in which \( T_0 \) is the likelihood function. GCV can be viewed as an approximation of leave-one-out cross validation [30] for large-scale problems.

The method is as follows: choose \( \alpha \) to be the minimizer of
\[ \text{GCV}(\alpha) \overset{\text{def}}{=} \frac{MT_0^{wls}(x_0; b)}{\text{trace}(I_n - B^{-1/2}_\alpha A A_\alpha)^2} \tag{22} \]
subject to the constraint \( \alpha \geq 0 \). Here \( x_0 \) is computed from (7), and \( A_\alpha \) is a matrix satisfying \( x_0 = A_\alpha B^{-1/2}_\alpha (b - \gamma) \).

However, for (7), the data-to-regularized solution (or regularization) operator is nonlinear, and so \( A_\alpha \) must be a linear approximation satisfying \( x_0 \approx A_\alpha B^{-1/2}_\alpha (b - \gamma) \). In [5], the approximation
\[ A_\alpha = (D_\alpha(A^T B^{-1}_\alpha A + \alpha C)D_\alpha)^{1/2}D_\alpha A^T B^{-1/2}_\alpha \tag{23} \]
is used, where “\( \dagger \)” denotes psuedo-inverse, and \( D_\alpha \) is a diagonal matrix with diagonal entries \( [D_\alpha]_{ii} = 1 \) if \( [x_0]_i > 0 \) and \( [D_\alpha]_{ii} = 0 \) otherwise (see [5] for details).

Even with (23) in hand, however, the computation of the trace in (22) remains impractical, and so randomized trace estimation is used [14, 31]. Here the following fact is exploited: if \( v \) is a discrete white noise vector,
\[ \text{trace}(I_n - B^{-1/2}_\alpha A A_\alpha) \approx v^T v - v^T B^{-1/2}_\alpha A A_\alpha v. \tag{24} \]
In fact, equality holds in (24) if the right-hand side is replaced by its expected value. As is mentioned in [31], the optimal choice of \( v \) in (24) is to take \( v_i \) to be \(+1\) or \(-1\) with probability \(1/2\).

Finally, because the pseudoinverse in (23) is impractical to compute, \( A_\alpha v \) in (24) is approximated using a truncated conjugate gradient iteration applied to the linear system

\[
D_\alpha (A^T B_\alpha^{-1} A + \alpha C)D_\alpha x = D_\alpha A^T B_\alpha^{-1/2} v
\]

with a stopping rule based on the norm of the residual and a choice of maximum number of iterations.

Taking all of the above approximations of GCV(\( \alpha \)) into account, and calling the resulting approximate GCV function \( \tilde{GCV}(\alpha) \), we define the GCV choice of the regularization parameter by

\[
\alpha_{GCV} = \arg\min_{\alpha \geq 0} \tilde{GCV}(\alpha).
\]

2.3. Unbiased Predictive Risk Estimation

In [5], a quadratic approximation of \( T_0 \) is also used to extend the method of unbiased predictive risk estimation (UPRE) for regularization parameter choice in the least squares case [31] to the case in which \( T_0 \) is the likelihood function.

The motivation behind UPRE is as follows: we seek the value of \( \alpha \) that minimizes the predictive risk \( E(T_0(x_\alpha; b_\alpha)) \). However since \( b_\alpha \) is unknown, we minimize instead an unbiased estimator of the predictive risk. Following the arguments in [5], such an estimator can be well-approximated by choosing \( x_\alpha \) with \( \alpha \) approximately minimizing

\[
\text{UPRE}(\alpha) = T_0^{\text{wls}}(x_\alpha; b) + \text{trace}(B_\alpha^{-1/2} AA_\alpha) - M/2,
\]

where \( x_\alpha \) is computed from (7). The trace is estimated in the same fashion as for GCV, with the exception that in place of (24), we use

\[
\text{trace}(B_\alpha^{-1/2} AA_\alpha) \approx v^T B_\alpha^{-1/2} AA_\alpha v.
\]

This results in an approximate UPRE function \( \tilde{\text{UPRE}}(\alpha) \), and the UPRE choice of the regularization parameter is given by

\[
\alpha_{\text{UPRE}} = \arg\min_{\alpha \geq 0} \tilde{\text{UPRE}}(\alpha).
\]

Remark 1: It may, at first glance, seem that the regularization parameter choice methods (21), (26), and (29) do not depend upon the noise level. However, for Poisson data satisfying (1), the variance of \( b_\alpha \), and hence the noise-level at pixel \( i \), is \([Ax_\alpha + \gamma]_i\), which we approximate by \( b_\alpha = [Ax_\alpha + \gamma]_i \). This corresponds to the approximate noise model

\[
b = Ax_\epsilon + \eta, \quad \eta \sim N(0, \text{diag}(b_\alpha)),
\]

and the noise level corresponds to \( \|b_\alpha\| \). Hence, (21), (26), and (29) depend upon the noise level.
Note that in the case that $b_{\alpha} = \sigma^2 I$, the equation to be solved for the discrepancy principle (analogous to (20)), is the classical one

$$\|Ax_{\alpha} - b\|^2 = \sigma^2 M.$$

Finally, it remains to rigorously prove that the above methods are indeed convergent, i.e. that as the noise level goes to zero, the resulting $\alpha$ values obtained by our methods converge to zero in such a way that the regularization solution $x_{\alpha} \to x_e$. The Taylor series result (14) suggests that because the methods have been proven to be convergent for least squares problems, they are likely to also be convergent in the Poisson case. However, we do not pursue the convergence results here.

3. The Bayesian point-of-view and hierarchical regularization

In the Bayesian setting, $x_{\alpha}$ computed from (7) is also the maximum a posteriori (MAP) estimator:

$$x_{\alpha} = \arg\max_{x \geq 0} p(b \mid x)p_{\text{prior}}(x),$$

where $p(b \mid x)$ is defined in (4) and

$$p_{\text{prior}}(x) = \frac{\alpha |C|}{(2\pi)^n} \exp \left( -\frac{1}{2} x^T C x \right).$$

A very effective regularization function can be obtained, following [11], if we make use of a Bayesian hierarchical model for the prior. This amounts to replacing (30) by

$$\arg\max_{x \geq 0, \theta} p(b \mid x)p_{\text{prior}}(x \mid \theta)p_{\text{hyper}}(\theta),$$

with $p$ given by (4).

We obtain our model for $p_{\text{prior}}(x \mid \theta)$ by assuming that

$$D_1 x, D_2 x \sim \text{Normal}(0, D_\theta), \quad D_\theta = \text{diag}(\theta_1, \ldots, \theta_N),$$

where $D_1$ and $D_2$ are the vertical and horizontal discrete derivatives, respectively. Note that (33) can be motivated by the fact that PET/SPECT images are typically approximately piece-wise smooth (corresponding to small values of $\theta_i$), with sharp jumps in intensity (corresponding to larger values of $\theta_i$), at, e.g., tissue boundaries. Assuming that the two random vectors in (33) are independent, the prior then has the form

$$p_{\text{prior}}(x \mid \theta) = p(D_1 x \mid \theta)p(D_2 x \mid \theta)$$

$$\propto \det(D_\theta^{-1}) \exp \left( -\frac{1}{2} x^T (D_1^T D_\theta^{-1} D_1 + D_2^T D_\theta^{-1} D_2) x \right)$$

$$= \exp \left( -\frac{1}{2} x^T (D_1^T D_\theta^{-1} D_1 + D_2^T D_\theta^{-1} D_2) x - \sum_{i=1}^N \log \theta_i \right).$$

(34)
See [3] for details. Note that if \( \theta = \alpha^{-1} \mathbf{1} \) is deterministic, this is equivalent to using 
\[ C = D_1^T D_1 + D_2^T D_2 \] 
in (7), and we are back to classical regularization.

It remains only to determine the values of the \( \theta_i \)'s in (33) and (34), which requires the definition of \( p_{\text{hyper}}(\theta) \). As in [3], we assume

\[ \theta_i \sim \text{Gamma}(\alpha_0, \theta_0), \quad i = 1, \ldots, n, \]

where Gamma denotes the Gamma distribution, so that

\[ p_{\text{hyper}}(\theta) \propto \prod_{i=1}^{N} \theta_i^{\alpha_0-1} \exp \left( \frac{\theta_i}{\theta_0} \right). \tag{35} \]

Note that the Gamma distribution is motivated for use in this context in [3, 11] by the fact that variance values for the prior (i.e. the \( \theta_i \)'s) are reasonably assumed to be independent, identically distributed and small with a few large values, or outliers. The hyper-parameters \( \alpha_0 \) and \( \theta_0 \) can be chosen so that the mean is as small as one likes, and since the Gamma is a heavy-tailed distribution, outliers are allowed, making it an appropriate choice for the hyper-prior. We note, however, that various other distributions satisfy these properties.

The problem of choosing the hyper-parameters remains, and a method for automating the choice of \( \alpha_0 \) and \( \theta_0 \) will be presented below. We also note, in passing, that the Gamma distribution is conjugate to the Poisson distribution, which can be exploited, but we do not make use of this property here.

Once the hyper-parameters are defined, the MAP estimate is computed by minimizing 
\[ -\ln p(\mathbf{b} | \mathbf{x})p_{\text{prior}}(\mathbf{x} | \theta)p_{\text{hyper}}(\theta), \]
with respect to \( \mathbf{x} \geq 0 \) and \( \theta \). A simple cyclic iteration has been found to be effective for this problem [3]. The outline of the algorithm is as follows:

**Step 0:** Initialize \( \theta^0 = (\alpha_0 \theta_0) \cdot \mathbf{1}, \ k = 1. \)

**Step 1:** Update the estimate \( \mathbf{x} \):
\[ \mathbf{x}^k = \text{argmin}_{\mathbf{x} \geq 0} -\ln p(\mathbf{x}, \theta^{k-1} | \mathbf{b}). \]

**Step 2:** Update the estimate of \( \theta \):
\[ \theta^k = \text{argmin}_{\theta \geq 0} -\ln p(\mathbf{x}^k, \theta | \mathbf{b}). \]

**Step 3:** Increase \( k \) by 1 and return to **Step 1**. Repeat until convergence.

Note that optimization problem in **Step 1** has the form of (7), with

\[ C = D_1^T D_{\theta^{-1}} D_1 + D_2^T D_{\theta^{-1}} D_2, \]

and the update in **Step 2** has the analytic form

\[ \theta_j^k = \theta_0 \left( \frac{\alpha_0 - 2}{2} + \sqrt{\frac{|D_1 \mathbf{x}^k|_j^2 + |D_2 \mathbf{x}^k|_j^2}{2\theta_0} + \frac{(\alpha_0 - 2)^2}{4}} \right). \tag{36} \]

The simplicity of the update (36) is another reason to use the Gamma distribution.

Left unspecified is the choice of the parameters \( \alpha_0 \) and \( \theta_0 \) in the gamma hyper-prior. For \( 0 < \alpha_0 - 2 \ll 1 \), using (36) yields a regularization function closely related to total variation; note that when \( \alpha_0 = 2 \), total variation results. Since in our application we would like to encourage piece-wise smooth (total variation-like) reconstructions, in our experiments we chose \( \alpha_0 = 2.01. \)

For the choice of \( \theta_0 \), we note that the mean of Gamma(\( \alpha_0, \theta_0 \)) is \( \alpha_0 \theta_0 \), which should be approximately equal to \( \alpha^{-1} \), where \( \alpha \) is obtained by applying one of the
above regularization parameter selection methods to (7) with $C = D_1^T D_1 + D_2^T D_2$. Thus we advocate choosing $\theta_0 = 1/(\alpha_0)$.

4. Numerical Experiments

![Figure 1](image1.png)

Figure 1. $x_e$ is plotted on the left and $b$ is plotted on the right. The signal-to-noise ratio of $b$ is 20. Here the PET model is used.

We test our approach on synthetically generated data. The true emission density $x_e$ is shown in Figure 1. The noisy sinogram data $b$, generated using statistical model (1), MATLAB’s `poisrmd` function, and $A$ defined by the PET mathematical model, is shown on the right in Figure 1. We assumed that $\gamma$ is a constant vector of 1s at all pixels, and that the density vector $\mu$ was a vector of zeros in the PET case (typical for numerical experiments in the PET literature) and the indicator function on the support of $x_e$ in the SPECT case as is done in [15]. 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

$$SNR = \sqrt{\frac{\|Ax_e + \gamma\|^2}{E(||b - (Ax_e + \gamma)||^2)},}$$

with

$$E(||b - (Ax_e + \gamma)||^2) = \sum_{i=1}^{M}([Ax_e]_i + \gamma_i).$$

We test our methods on data with two different SNR values, 5 and 20.

MATLAB’s `fminbnd` function was used for computing approximate solutions to (21), (26), and (29). A more efficient method, exploiting the unique structure of our problem, is likely possible, however we do not pursue that here.

To test the effectiveness of the regularization parameter selection methods, we plot the relative error $\|x_{\alpha} - x_e\|/\|x_e\|$ for a range of $\alpha$ values, together with the values of $\alpha$ chosen by the three methods. This can be seen in Figure 2. The PET mathematical model is used, and the regularized solution $x_\alpha$ is calculated from (7).
with $C = L$, where $L$ is a discretization of the negative Laplacian (smoothing) operator. Note that when $\alpha$ is close to zero the relative error is large due to the presence of unrealistic artifacts, whereas when $\alpha$ is too large the relative error is large because the penalty term dominates the reconstruction.

In both cases, the GCV and UPRE methods yielded similar recommendations for $\alpha$, while the DP (discrepancy principle) yielded a recommendation that was slightly worse in terms of relative error. Figure 3 contains the reconstructions that were obtained from the two data sets using the DP and UPRE recommendations. The GCV recommendation yielded a reconstruction that was visibly very similar to that obtained from the UPRE recommendation.

If $\alpha$ is computed as above, i.e. using DP, GCV, or UPRE with $C = L$, an improved reconstruction can be obtained by taking $\alpha_0 = 2.01, \theta_0 = 1/(2.01\alpha)$ and iterating the hierarchical regularization algorithm of Section 3. Figure 4 shows the reconstructions $x^k$ and $\theta^k$ from this approach obtained after 6 iterations of the algorithm. Note the smoothness properties of the reconstructions. Also note that this completely automates the hierarchical method.

When the SPECT mathematical model is used, we obtain very similar results to those presented in Figures 2-4, and hence, do not show them here. This is likely due to the similarity in the models. However, we nonetheless present reconstructions obtained using the hierarchical approach, implemented precisely as in the previous paragraph for the PET case, in Figure 5.

5. Conclusions

We have presented regularization parameter choice methods for penalized negative-log Poisson likelihood problems arising in positron emission tomography (PET) and single photon emission computed tomography (SPECT). These methods are extensions of the discrepancy principle, generalized cross validation, and unbiased predictive risk estimation for regularization parameter choice in the least squares setting. The approach set forth here corresponds to a minor modification of that presented in [5].

The numerical results show that the methods yield good estimates of the regularization parameter and, moreover, that they can be used for choosing values of the hyper-parameters in the hierarchical regularization approach of [3]. The resulting
algorithms are effective for use on both PET and SPECT imaging examples.

References


Figure 4. On the left are the reconstructions of $x$ and $\theta$ obtained when the SNR is 5. On the left are the reconstructions of $x$ and $\theta$ obtained when the SNR is 20. In all cases, the PET model is used.

Figure 5. On the left is the reconstructed emission density $x$. On the right is the estimated value of $\theta$. Here the SNR is 10 and the SPECT model is used.
REFERENCES


