Interior-Point Methodology for Large-Scale Regularized Maximum Likelihood Reconstruction in Emission Tomography

Calvin A. Johnson
Division of Computer Research and Technology
National Institutes of Health
Bethesda, MD

Ariela Sofer
Department of Operations Research and Engineering
George Mason University
Fairfax, VA

DRAFT

June 9, 1997

Abstract

Interior-point methods possess strong theoretical properties and have been successfully applied to a wide variety of linear and nonlinear programming applications. This paper presents a class of algorithms, based on interior-point methodology, for performing regularized maximum likelihood reconstructions on 3-D emission tomography data. The algorithms solve a sequence of subproblems that converge to the regularized maximum likelihood solution from the interior of the feasible region (the non-negative orthant). We propose two methods, a primal method which updates only the primal image variables and a primal-dual method which simultaneously updates the primal variables and the Lagrange multipliers. Termination to a solution of desired accuracy is based on well-defined convergence measures. We demonstrate the rapid convergence of the interior-point methods using both data from a small animal scanner and Monte Carlo simulated data. We present a parallel implementation that permits the interior-point methods to scale to very large reconstruction problems.

1 Introduction

Positron emission tomography (PET) images can be reconstructed by an iterative process whereby the iterates maximize a Poisson likelihood objective function. Maximum likelihood (ML) reconstructions have been reported to improve both variance [71] and resolution [44] compared with filtered back projection reconstructions, in studies characterized by poor counting statistics. In transmission tomography for example, where the scan duration is often short and the count rate is therefore low, ML reconstructions based on a Poisson model of the observations have significantly improved the resulting attenuation maps [18, 52, 61]. However, since image reconstruction from projections is an ill-posed problem, ML estimates tend to overfit the data [1], resulting in increased noise and edge artifacts in later iterates of the expectation maximization (EM) algorithm [67].

One approach to treating the problem of noisy solutions is to terminate the EM algorithm early by using either a stopping heuristic (such as 50 iterations [22, 68, 72]) or an early termination criterion [50, 26, 70]. This approach requires either an arbitrarily defined stopping point (in the case of the heuristic) or a criterion which may or may not be well suited to the particular reconstruction problem. Another potential disadvantage of this approach is that early termination of the EM algorithm biases the solution towards the starting point, which is often a uniform field [1]. Moreover, Herman and Odhner have reported that later iterates of the EM algorithm may actually possess improved structural accuracy over early terminated solutions [30], and Miller and Wallis have reported improved contrast recovery in later iterates [51].

Alternatively, one can regularize the ML solution. This can be done either by post-filtering, by using the method of sieves [67], or by introducing a regularization term in the objective function. These methods reduce the "checkerboard" noise effect and edge artifact of the converged maximum likelihood solution at the expense of some resolution. The latter of the three approaches, commonly referred to as penalized maximum likelihood [16] or maximum *a posteriori* (MAP) reconstruction [49], has generated the most interest because, in addition to improving the noise properties of the converged solution, inclusion of a regularization term in the objective function also leads to faster convergence [47].

PET reconstructions of small animals are another class of problems characterized by poor counting statistics. Small animal studies require very high resolution at reasonable noise levels in order to distinguish small features in, say, a rat brain [65]. Small animal studies are potentially count-limited because higher tracer doses may cause biochemical saturation in systems of interest, e.g. receptor systems [45]. Other count-limiting factors include the need for reasonable scan durations and limitations on the sensitivity of the scanner [63, 64]. The low-count nature of small animal studies necessitates 3-D reconstructions to improve sensitivity [35]. The increased proportion of scattered events in 3-D, so prevalent in clinical 3-D PET reconstructions, is not significant in the small animal case because the small size of the object being imaged implies low

attenuation and hence a low proportion of scattered events. The low count rate also motivates the use of a Poisson model in conjunction with a maximum likelihood objective function for the reconstructions [35].

In recent years, much has been accomplished in improving the convergence of unregularized maximum likelihood reconstructions using 'block-iterative' methods. Hudson and Larkin's ordered subset EM (OSEM) [33] and Byrne's block-iterative algorithm [4] generate iterates that, at k iterations, very closely resemble the corresponding EM iterate at $k \cdot r$ iterations, where r is the number of subsets [36]. These algorithms do not converge to the maximum likelihood solution, however [3]. Another approach is based on row-action mathods [5]. Browne and DePierro [3] developed a row-action method , similar to the Algebraic Reconstruction Technique (ART) of Herman and Meyer [31], that converges to the maximum likelihood solution. This method, which Browne and DePierro dubbed the Row Action Maximum Likelihood Algorithm (RAMLA), can be implemented as a slight modification of OSEM and enjoys the apparent rapid convergence of OSEM. RAMLA does not, however, generalize to include a regularization term in the objective function.

There has also been significant recent activity in developing algorithms for improving the convergence of MAP and penalized maximum likelihood reconstructions. Mumcuoglu et al. introduced a quadratic penalty function to impose non-negativity (but did not follow a trajectory as we shall discuss) and maximized the MAP objective function via a nonlinear conjugate gradient with a line search [52]. Lalush and Tsui investigated a preconditioned nonlinear conjugate gradient method [42]. In a separate paper, Mumcuoglu investigated an active set method for MAP reconstruction [53]. These methods have resulted in only modest convergence acceleration.

Coordinate ascent and grouped coordinate ascent methods have recently been reported to significantly improve convergence over MAP-EM [29, 2, 17, 18, 62]. Using a Fourier domain analysis to study the convergence properties of coordinate ascent, Sauer and Bauman [62] have shown that coordinate ascent iterates converge faster in higher frequencies than MAP-EM. Fessler et al. have shown that a carefully selected grouped update strategy corresponds to choosing hidden data spaces with reduced Fisher information, thereby leading to asymptotically faster convergence[17, 18]. The grouped coordinate ascent methods can be parallelized, unlike the single coordinate ascent methods. We shall consider these methods again in Section 6.

In this paper we introduce a class of algorithms for performing large-scale (3-D) ML and MAP reconstructions. These algorithms depart entirely from the EM philosophy; rather, they are based on interior-point methodology from nonlinear optimization theory [60]. Particular attention is devoted in this paper to computational issues including per-iteration costs and parallel computer implementation issues. Computational tests are presented on actual data from a small-animal scanner and Monte Carlo simulated data. We approach the problem of reconstructing a PET image from a constrained optimization perspective and demonstrate that interior-point methods can significantly improve the convergence to the fully converged MAP reconstruction problem, compared

with MAP-EM, and can improve convergence to a less accurate solution as well. We demonstrate that the interior point methods are suitable for even the largest of 3-D problems.

The paper is organized as follows. In Section 2 we review the principles of ML reconstruction (including regularized ML), with an emphasis on understanding the objective function, and briefly summarize the EM algorithm. Section 3 presents a parallel implementation of the EM algorithm that can also be applied to the interior point methods. Two interior point methods for regularized ML reconstruction are proposed in Section 4. The first of these methods is a classical method that imposes the non-negativity constraints through a logarithmic barrier function. The *barrier* method is a primal method in that it only solves for the primal image variables. The second algorithm we propose is a *primal-dual* method that simultaneously updates both the primal variables and Lagrange multipliers (dual variables). Much of the material developed in Section 4.1 for the barrier method also applies to the primal-dual method in Section 4.2. A special implementation of the interior-point algorithms that exploits sparsity in the observation data is discussed in Section 4.3. Large-scale 3-D computational studies using the interior-point algorithms and a comparison with MAP-EM results are presented in Section 5. Some concluding remarks are made in Section 6.

2 Maximum Likelihood Reconstructions

The goal of ML estimation, as applied to emission tomography, is to find the estimates of the expected number of emission events that maximize the probability of the set of observations $y_j, \quad j = 1, \ldots, N$. As is customary, we form a finite parameter space by imposing a grid of boxes or *pixels* over the emitting object [28]. Given our set of observations or *projections* we seek to estimate $\theta_i = E\{\mathbf{x}_i\}, \ i = 1, \ldots n$, the expected number of counts emitted from pixel i. The number of radioactive events emitted from pixel i are assumed to be Poisson-distributed random variables with means θ_i. By a Bernoulli thinning process with probability $C_{i,j}$, the number of events emitted from pixel i and detected in projection line j, $\mathbf{X}_{ij} = \mathbf{x}_i C_{i,j}$ are also independent Poisson variables (the realizations X_{ij} are the complete but unobservable data set in EM [13]). The observations y_j are realizations of sums of independent Poisson variables $\mathbf{y}_j = \sum_i C_{i,j}\mathbf{X}_{i,j}$ that can also be assumed to be Poisson distributed. The *system matrix* $C \in \Re^{n \times N}$ of the probabilities $C_{i,j}$ can be computed based on geometric considerations and the spatially-variant resolution function of the scanner.

The projections y_j are thus assumed to be realizations of independent Poisson variables with means $m_j = \sum_i C_{i,j}\theta_i$. A more accurate model of the observations would be

$$m_j = \sum_i B_{i,j}\theta_i + r_j + s_j$$

where r_j and s_j are known values for random and scatter coincidences, and

$B_{i,j} = A_j C_{i,j}$ where A_j are known attenuation coefficients. In small-animal studies, the emitting object is small enough that attenuation, and therefore random and scattered coincidences, do not significantly corrupt the simple model of Poisson observations. For example, the reconstruction of a uniform cylinder in Figure 1(d) (in Section 5) demonstrates relatively uniform activity in the cylinder despite the absence of attenuation correction.

Given our simplified Poisson model, the likelihood may be written as

$$P\{y|\theta\} = \prod_j \frac{e^{-m_j} m_j^{y_j}}{y_j!} = \prod_j \frac{e^{-\sum_i C_{i,j}\theta_i} \left(\sum_i C_{i,j}\theta_i\right)^{y_j}}{y_j!}.$$

The ML objective function is formed by taking the log likelihood

$$f_{ML}(\theta) = \log P\{y|\theta\} = \sum_j \left(-\sum_i C_{i,j}\theta_i + y_j \log \sum_i C_{i,j}\theta_i - \log(y_j)! \right),$$

that is,

$$f_{ML}(\theta) = \sum_j \left(-\left(C^T\theta\right)_j + y_j \log\left(C^T\theta\right)_j - const \right). \tag{1}$$

Defining $q = Ce_N \in \Re^n$ to be the sum of the columns of C, where $e_N \in \Re^N$ is a vector of 1's, and defining

$$\hat{y} = C^T\theta$$

to be a *forward projection,* we can write the gradient and Hessian of the objective function, respectively, as

$$\nabla f_{ML}(\theta) = -q + C\,\hat{Y}^{-1} y, \tag{2}$$

and

$$\nabla^2 f_{ML}(\theta) = -C\,Y\hat{Y}^{-2}C^T. \tag{3}$$

where $Y = \mathrm{diag}(y_j,\ j = 1,\ldots,N)$ and $\hat{Y} = \mathrm{diag}(\hat{y}_j,\ j = 1,\ldots,N)$. The Hessian is known to be negative semidefinite for this problem (since $y_j/\hat{y}_j \geq 0 \quad \forall j$) and the objective function (1) is concave. Thus, any local maximum will also be a global maximum.

Equation (2) sheds some insight into the computational costs associated with maximizing the objective function. Given a current ($k-$th iteration) set of estimates θ^k, computing the gradient requires first computing a forward projection $\hat{y}^k = C^T\theta^k$ and then computing a *back projection* $C\left(\hat{Y}^k\right)^{-1} y$ from the forward projection. The costs of forward projecting and back projecting are similar and together dominate the computation associated with iterative reconstruction methods, especially in large scale. We shall revisit this computational structure, which is common to all iterative reconstruction methods.

Since the underlying activity distribution is non-negative, the ML reconstruction problem is a constrained optimization problem with lower bound constraints:

$$\begin{aligned} \max \quad & f_{ML}(\theta) \\ \text{s.t.} \quad & \theta \geq 0. \end{aligned} \tag{4}$$

2.1 The EM Algorithm

The expectation maximization (EM) method, first presented by Dempster, Laird, and Rubin [13] for ML estimation, is an iterative algorithm for computing ML estimates when the observations can be viewed as incomplete data. Shepp and Vardi [66] and Lange and Carson [46] applied the EM method to emission and transmission tomography problems, respectively. The EM algorithm has been proven to converge to the optimal solution of (4) [69].

The EM algorithm of Shepp and Vardi requires a positive initial solution $\theta^0 > 0$ be selected. The *E step* of the algorithm takes the expectation of the complete data log likelihood $Q\left(\theta, \theta^k\right) = E\left\{P\left[X|y, \theta^k\right]\right\}$. For unregularized ML, a closed-form solution exists to the *M step* (maximization of $Q\left(\theta, \theta^k\right)$) :

$$\theta_i^{k+1} = \frac{\theta_i^k}{q_i} \sum_{j=1}^{N} \frac{C_{i,j} y_j}{\hat{y}_j}. \tag{5}$$

The M step comprises a separable maximization of a strictly concave function $Q\left(\theta, \theta^k\right)$ that has a unique maximum θ^{k+1} and guarantees a monotonic increase in $f_{ML}(\theta)$. Equation (5) may be rewritten as

$$\theta_i^{k+1} = \theta_i^k + \frac{\theta_i^k}{q_i}\left(\sum_j C_{i,j} \frac{y_j}{\hat{y}_j^k} - q_i\right) = \theta_i^k + \frac{\theta_i^k}{q_i} \frac{\partial f_{ML}\left(\theta^k\right)}{\partial \theta_i}, \tag{6}$$

which may be expressed in vector form as

$$\theta^{k+1} = \theta^k + \operatorname{diag}\left(\frac{\theta_i^k}{q_i}\right) \nabla f_{ML}\left(\theta^k\right). \tag{7}$$

From (7) we see that EM is essentially a preconditioned steepest ascent method with a unit step length [38, 39]. The diagonal (EM) preconditioner guarantees that the solution at every iteration resides in the interior of the feasible region. As such, EM behaves like an interior-point method. It has been observed that EM converges slowly. The cost of an EM iteration is the cost of a gradient evaluation, one forward projection plus one back projection. The observed slow convergence and the high per-iteration cost of the EM algorithm have limited its usefulness in the clinic.

2.2 Maximum *a Posteriori* Reconstruction

Investigators have observed that the EM algorithm can produce noisy solutions at convergence. This observation has led numerous authors to devise methods to regularize the maximum likelihood objective function. As previously discussed, currently the most investigated approach is to introduce a Bayesian formulation [28, 47]. Given prior probabilities $P\{\theta\}$ and $P\{y\}$ for the image and observations, respectively, we define the posterior probability

$$P\{\theta|y\} = \frac{P\{y|\theta\}\,P\{\theta\}}{P\{y\}}.$$

The estimate of θ is then obtained by maximizing the posterior probability $P\{\theta|y\}$.

A common choice for the image prior is the Gibbs distribution $P\{\theta\} = e^{-\beta R(\theta)}$, although other priors (Gaussian, gamma) have been investigated [49, 47]. The popularity of Gibbs priors stems in part from their ability capture the local correlation property of images [27]. The energy function $R(\theta)$ is defined as a sum of neighborhood functions

$$R(\theta) = \frac{1}{2}\sum_i \sum_{l\in\mathcal{N}_i} \psi_{i,l}(\theta_i, \theta_l).$$

where $\mathcal{N}_i$ denotes the neighborhood of pixel i. In order to maintain concavity in the objective function, the potential function $\psi_{i,l}(\theta_i, \theta_l)$ should be convex with continuous first and second derivatives. The potential function is generally designed to discourage non-smoothness in a neighborhood. For the small animal studies, we have chosen the potential function $\psi_{i,l}(\theta_i, \theta_l) = \psi(\theta_i - \theta_l)$, where ψ is defined [48]

$$\psi(z) = |z| - \log(1 + |z|).$$

For maximum *a-posteriori* (MAP) reconstructions, the objective function (ignoring a constant term) is the log-posterior likelihood

$$f_{MAP}(\theta) = \log P\{\theta|y\} = f_{ML}(\theta) - \gamma R(\theta). \tag{8}$$

The MAP reconstruction problem may also be posed as a constrained optimization problem

$$\begin{array}{ll} \max & f_{MAP}(\theta) \\ \text{s.t.} & \theta \geq 0. \end{array} \tag{9}$$

We note for future reference the following:

$$\begin{aligned} \nabla f_{MAP}(\theta) &= \nabla f_{ML}(\theta) - \gamma \nabla R(\theta) = -q + C\,\hat{Y}^{-1} y - \gamma \nabla R(\theta) \\ \nabla^2 f_{MAP}(\theta) &= \nabla^2 f_{ML}(\theta) - \gamma \nabla^2 R(\theta) = -C\,Y\hat{Y}^{-2}C^T - \gamma \nabla^2 R(\theta). \end{aligned}$$

If the energy function is convex, $f_{MAP}(\theta)$ is concave, and $\nabla^2 f_{MAP}(\theta)$ is negative semidefinite.

In MAP-EM, the presence of the regularzing term in (8) precludes a closed form solution to the M-step. A number of authors have devised generalizations of EM for MAP. For comparison with our interior-point algorithms, we have implemented two such approaches: Green's "one step late" (OSL) algorithm [24, 25] and DePierro's generalized EM algorithm for penalized ML [14, 15]. The OSL algorithm approximates $R(\theta)$ with the constant $R\left(\theta^k\right)$, thereby producing a closed-form approximated update. OSL converges to the MAP solution for sufficiently small γ. DePierro's algorithm generalizes the EM approach to MAP by defining a strictly concave, separable, and twice continuously differentable "surrogate function" for the M-step (to include the regularization term). DePierro proved that successive maximizations of this surrogate function converge to the MAP solution.

Our experience has been that the iterates generated by OSL and the DePierro algorithm are very similar. The per-iteration costs of both algorithms are also similar. In both cases, a forward projection followed by a back projection must be performed at each iteration:

$$\nu = C\hat{Y}^{-1}y. \tag{10}$$

The vector $\nu \in \Re^n$ is then used in the simultaneous update (OSL or DePierro). The vast majority of the computational expense of a large-scale 3-D MAP-EM iteration is in performing the forward and back projections. The cost of performing DePierro's simultaneous update, even to high precision, is insignificant compared to the forward and back projection costs.

3 Parallel Computing

Large-scale 3-D maximum likelihood PET reconstructions are characterized by huge vector spaces and by sparsity in the system matrix [6]. The number of projection lines may be in the millions. The various parameters that affect the size of the reconstruction problem are listed in Table 1. In "thick-slice" reconstructions, list-mode data from the small animal scanner are binned into comparatively wide axial rings, whereas in the much larger "thin-slice" reconstructions, the list-mode data are binned into much narrower slices, resulting in nearly cubic voxels. Even in the thick-slice reconstructions, the full size of the system matrix is prohibitive. Storage of the system matrix thus requires the exploitation of its sparsity and symmetries. For every projection index j, only a narrow *chord* of pixels, representing just the region of support of projection line blurring, are nonzero. We typically define the boundary of the chords' region of support at 5% of maximum, resulting in cross sections of size 7×5 (thick-slice) and 7×7 (thin-slice), although we have also used more accurate (larger) cross sections. As indicated in the line marked 'nonzeros in C' in Table 1, incorporating sparsity in the storage of the system matrix, alone, is not enough to reduce the storage cost sufficiently; we must also take advantage of symmetries in the system matrix. These symmetries, which include rotation and reflection in the

	thick-slice	**thin-slice**
image size	$128\times128\times23$	$128\times128\times85$
axial detector rings	12	43
transverse angles	240	240
rays per angle	155	155
n	376,832	1.40×10^6
N	5.35×10^6	6.30×10^7
elements in C	2.02×10^{12}	8.82×10^{13}
nonzeros in C	1.87×10^{10}	6.14×10^{11}
storage required for C	361 MB	1.36 GB
cost of sparse gradient	3 minutes	10 minutes
cost of full gradient	7 minutes	4 hours

Table 1: Properties affecting computation, memory, and storage costs for two different-sized reconstruction problems. Gradient evaluation costs are based on a 2.5M count study on an IBM RS/6000 SP using 10 133-MHz processors.

transverse plane [38, 39], axial reflection [8], and parallel axial redundancies [34], account for a storage reduction factor of 104 for thick-slice reconstructions and 451 for thick-slice reconstructions.

Since in large-scale 3-D reconstructions, $N \gg n$, a reasonable data decomposition strategy would be to partition projection-space data across the processors. Indeed, as indicated in Table 1, the cost to memory of storing an 8-bit y vector and a single-precision $\hat{y}$ vector is 315 MB for a thin-slice reconstruction, Clearly then, it is necessary to partition in projection space and decompose the problem accordingly. In our implementation, the symmetries are indexed primarily in projection space. We can therefore decompose the problem by projection-space symmetry and assign each processor certain groups of symmetry-related projections. Specifically, we define $N = \sum_{l=1}^{m} b_l$, where m is the number of base-symmetry angles, $l = 1, \ldots, m$ is a particular symmetry index, and b_l is the number of symmetry-related projections in symmetry index l. We then define

$$\begin{aligned} y &= \left[\begin{array}{ccccc} y_1^T & y_2^T & : & : & y_m^T \end{array}\right]^T \\ \hat{y} &= \left[\begin{array}{ccccc} \hat{y}_1^T & \hat{y}_2^T & : & : & \hat{y}_m^T \end{array}\right]^T \\ C &= \left[\begin{array}{ccccc} C_1 & C_2 & : & : & C_m \end{array}\right] \\ \hat{Y}_l &= diag\left(\left(\hat{y}_l\right)_j\right) \end{aligned}$$

where $y_l, \hat{y}_l \in \Re^{b_l}$ and $C_l \in \Re^{n\times b_l}$. Image-space vectors such as θ and ν (the latter being defined by (10)) are replicated on all processors, but during a back projection operation, these vectors temporarily store information representing the back projection of particular processor's portion of projection space. Thus we introduce $\nu_{(d)} \in \Re^n$, $d = 1, \ldots, D$, where D is the number if processors, and

define a parallel version of the forward back projection operation:

$\nu_{(d)} = 0$
`for all` l `assigned to my processor` d
$\hat{y}_l^{(k)} = C_l^T \theta$
$\nu_{(d)} = \nu_{(d)} + C_l \hat{Y}_l^{-1} y$
`end for`
$\nu = \sum_{d=1}^{D} \nu_{(d)}$ `(global summation across processors)`

Using the update vector ν, we may then apply the appropriate image update for the EM algorithm being implemented (for example, in ML-EM we would apply (5).) The DePierro MAP-EM update may also be decomposed in parallel in image space (the n-dimensional vectors are "image-space" vectors while the N-dimensional vectors are "projection-space" vectors.) We have found that DePierro separable maximization is actually quite fast and that computation is dominated by the forward and back projections. Every EM iteration requires a full traversal through the columns of the C matrix (the "chords.") These chords are read from disk as needed. Storing only the base-symmetry chords saves some I/O costs but adds to the computational cost since the chord information must be shuffled about in accordance with the symmetry operators. We noted earlier that forward projection and back projection operations are common to all iterative methods, including the new interior-point algorithms presented in this paper. The parallelization scheme presented herein thus generalizes to our new methods. For simplicity, we shall henceforth drop the partitioning notation and only remark on the parallelization scheme in the context of computational issues of interest.

4 Interior-Point Methods

Many authors have recognized the need to improve the convergence of the EM algorithm. The ML-EM algorithm maintains feasibility at every iteration. In generalizing the EM method to MAP, maximization of separable "surrogate functions" permits easy enforcement of non-negativity constraints [14]. The non-negativity property is lost when one departs from an EM framework and attempts to maximize the objective function directly. Some work has been done in developing line searches while maintaining feasibility [38, 47]. Others have considered conjugate gradient techniques [43, 52]. Most of the methods reported to date have been explicitly unconstrained methods with modifications to maintain feasibility rather than methods intended for constrained optimization. A significant exception to this is Mumcuoglu, who has attempted to enforce non-negativity via a quadratic penalty term in the objective function [52] and has explored active-set methods [53].

In this paper, we shall introduce methods for solving the ML and MAP reconstruction problems by solving a sequence of subproblems that successively approach the constraint boundary from the interior of the feasible region and

are hence known as *interior-point methods*. We shall develop special algorithms to converge to the ML or MAP solution in an interior-point framework. These algorithms carefully exploit the computational features of the objective function and avoid excessive computation and storage. A widely-held belief is that Newton's method is too computationally prohibitive to be applied to large-scale optimization problems. Our algorithms can be viewed as alternatives based on Newton's method that do not require storage or factorization of the Hessian and do not require storage of anything larger than an image-space vector. The huge projection-space vectors are partitioned in parallel, permitting the interior-point methods to be successfully applied to even the largest of the reconstruction problems.

Before we develop the algorithms, let us define a general objective function that can be applied to either ML or MAP reconstructions:

$$f(\theta) = \begin{cases} -f_{ML}(\theta) & \text{for ML} \\ -f_{MAP}(\theta) & \text{for MAP.} \end{cases}$$

Thus for convenience of notation we repose the reconstruction problem as a constrained *minimization* problem:

$$\begin{aligned} \min \quad & f(\theta) \\ \text{s.t.} \quad & \theta \geq 0. \end{aligned} \tag{11}$$

Since the ML objective function is simply a specific case of MAP with $\gamma = 0$, and because in general we are more interested in the fully converged MAP solution than that of ML, we shall henceforth assume that $f(\theta) = -f_{MAP}(\theta)$. The necessary conditions for optimality of (11) are existence of Lagrange multipliers λ so that

$$\nabla \ell(\theta, \lambda) = \nabla f(\theta) - \lambda = 0 \tag{12}$$

$$\lambda_i \theta_i = 0, \quad i = 1, \ldots n \tag{13}$$

$$\lambda \geq 0 \tag{14}$$

where $\ell(\theta, \lambda) = f(\theta) - \lambda^T \theta$ is the Lagrangian function. The Lagrange multipliers are also termed the *dual variables* (the variables θ_i are the *primal variables*.) The vector θ is *primal feasible* if $\theta \geq 0$. The pair $\left[\theta^T, \lambda^T\right]^T$ is *dual feasible* if $\nabla f(\theta) - \lambda = 0$ and $\lambda \geq 0$. The condition $\lambda_i \theta_i = 0$ is known as *complementary slackness*. We see therefore that necessary conditions for optimality are primal feasibility, dual feasibility, and complementary slackness. We shall denote the optimal solution that satisfies (12-14) as (θ^*, λ^*).

4.1 The Logarithmic Barrier Method

Barrier methods, developed by Fiacco and McCormick [19], find the solution to a *constrained* optimization problem by solving a sequence of *unconstrained* optimization subproblems. The unconstrained subproblems involve an

auxiliary function that penalizes closeness to the constraint boundaries. Barrier methods are strictly feasible methods, i.e. the iterates lie in the interior of the feasible region ($\theta > 0$). They maintain feasibility by creating a barrier that keeps the iterates away from the boundary of the feasible region. Barrier methods were considered for many years to be ineffective because of their potential for structural ill-conditioning of the Hessian. Recent developments in linear programming [37, 41] and nonlinear programming [59], however, have revived interest in barrier methods.

The most widely used auxiliary function is the logarithmic barrier. Let us define the logarithmic barrier objective function for (11) as

$$F(\theta, \mu) = f(\theta) - \mu \sum_i \log \theta_i.$$

In so doing, we form a new objective function. We specify a strictly feasible starting point $\theta^0 > 0$ and solve a sequence of *unconstrained subproblems* based on the logarithmic barrier objective function

$$\min F(\theta, \mu_k) \tag{15}$$

for a decreasing sequence of *barrier parameters* $\mu_k > 0$. The solution to each subproblem resides in the interior of the feasible region, and for this reason the logarithmic barrier method is an interior-point method.

Consider the point $\theta(\mu)$ that is a minimizer of $F(\theta, \mu)$. Since the gradient of the barrier function at this point is zero,

$$\nabla F(\theta(\mu), \mu) = \nabla f(\theta(\mu)) - \mu \Theta^{-1} e_n = 0,$$

where $e_n \in \Re^n$ and Θ is a diagonal matrix with diagonal elements $\theta_i(\mu)$. Defining

$$\lambda_i(\mu) = \frac{\mu}{\theta_i(\mu)} \quad i = 1, \ldots, n, \tag{16}$$

we note that

$$\nabla F(\theta(\mu), \mu) = \nabla f(\theta(\mu)) - \lambda(\mu) = \nabla \ell(\theta(\mu), \lambda(\mu)) = 0.$$

Thus, the solution to each subproblem satisfies a perturbed version of the first-order necessary conditions for optimality:

$$\begin{aligned} \nabla f(\theta(\mu)) - \lambda(\mu) &= 0 && (17) \\ \lambda_i(\mu)\theta_i(\mu) &= \mu, \quad i = 1, \ldots, n && (18) \\ \lambda(\mu) &\geq 0. && (19) \end{aligned}$$

(In practice we do not solve the subproblems exactly but rather set a termination criterion based on the minimum multiplier estimate and $\|\nabla F(\theta, \mu)\|$.) Under mild conditions, it can be proven [19] that as $\mu \to 0$, $\theta(\mu) \to \theta^*$, and $\lambda(\mu) \to \lambda^*$, where (θ^*, λ^*) is the optimal solution to the constrained problem (11).

4.1.1 Solving the Unconstrained Subproblem

Most successful algorithms for unconstrained optimization are motivated by Newton's method. In the minimization of an unconstrained objective function $F(\theta)$, Newton's method iteratively updates the set of estimates via

$$\theta \leftarrow \theta + p,$$

where the *Newton direction* p is found by solving

$$\nabla^2 F(\theta) \cdot p = -\nabla F(\theta). \tag{20}$$

The most important advantage of Newton's method is that when started sufficiently close to the solution, the method has a quadratic rate of convergence. In its "pure form" however, Newton's method has some drawbacks. First, a step size must be found to guarantee descent. Next, safeguards must be incorporated in the event of non-convexity (although this drawback does not apply to the present problem). Finally, for many large problems, including ML PET reconstructions, forming or factorizing the Hessian is prohibitive due to the size of the problem and the amount of computation that would be required.

An important convergence theorem states that, in order for an unconstrained optimization method to achieve superlinear convergence, its search directions must approach the Newton direction in the limit as the solution is approached [60]. Thus we are motived to use a method that approximates the Newton direction without incurring the heavy computational costs and potential difficulties of the pure Newton method. A robust method that approximates the Newton direction, requires little storage, is computationally efficient (and thus is suitable for very large problems), and can safeguard against non-convexity is the *truncated-Newton* [12] method. It consists of an outer iteration and an inner iteration. The outer iteration computes a step size α and updates the solution estimate

$$\theta \leftarrow \theta + \alpha p.$$

The search direction, computed by an inner iteration, is an approximate or *truncated* solution to the Newton equations [56],[58]

$$\nabla^2 F(\theta)\, p \approx -\nabla F(\theta). \tag{21}$$

A linear conjugate gradient method [32] is commonly used to solve for (21); the method is truncated before the exact solution has been found, hence the name "truncated-Newton."

An equivalent statement of (21) is we seek to find the direction p that approximately minimizes the quadratic $Q(p) = \frac{1}{2} p^T \nabla^2 F(\theta)\, p + \nabla F(\theta)\, p$. A reasonable and effective truncation point for (21), based on the monotonicity of $Q(p)$, is proposed in [57]; the conjugate gradient is terminated at subiteration l if

$$\frac{Q\left(p^l\right) - Q\left(p^{l-1}\right)}{Q\left(p^l\right)} \leq \frac{1}{2l}. \tag{22}$$

The truncated-Newton method does not require storage of the Hessian, but rather only application of Hessian-vector product(s) in the conjugate gradient subiteration(s). From (3) we have that the Hessian-vector product has the form

$$\nabla^2 f_{ML}(\theta)\, v = -C\, Y \hat{Y}^{-2} C^T v \tag{23}$$

for arbitrary vector $v \in \Re^n$. Computationally, (23) consists of a forward projection ($C^T v$) followed by a diagonal scaling (it is reasonable to assume that $\hat{y}$ is already available from the computation of $\nabla f(\theta)$), followed by a back projection (premultiplication by C). The computation of $\nabla^2 F(\theta)\, v$ requires only minimal additional computation, as does the computation of the Hessian-vector product for MAP. To be explicit, since

$$\nabla^2 F(\theta) = \nabla^2 f(\theta) + \mu \Theta^{-2}, \tag{24}$$

then

$$\nabla^2 F(\theta)\, v = C\, Y \hat{Y}^{-2} C^T v + \gamma \nabla^2 R(\theta)\, v + \mu \Theta^{-2} v,$$

where $\nabla^2 R(\theta)\, v$ can be computed exactly without incurring significant computational expense. The forward and back projection operations in (23) dominate the computational cost of a conjugate gradient iteration. This operation is computationally similar to computing the gradient, or one EM iteration. The MAP terms in the objective function, gradient, and Hessian vector product are readily decomposed in image space and parallelized, further reducing the overall computation.

4.1.2 Preconditioning

When working in exact arithmetic, the number of iterations required by the conjugate gradient method to find an exact solution to $\nabla^2 F \cdot p = -\nabla F$ is equal to the number of distinct eigenvalues of $\nabla^2 F$. (Here and elsewhere in the paper, we use ∇F and $\nabla^2 F$ as a shorthand for $\nabla F(\theta)$ and $\nabla^2 F(\theta)$ respectively.) The conjugate gradient method converges linearly with rate constant $\left(\sqrt{\kappa - 1}\right) / \left(\sqrt{\kappa + 1}\right)$, where κ is the condition number of $\nabla^2 F$ [58]. To accelerate convergence, we would like to find a *preconditioner* M such that $M^{-1}\nabla^2 F$ has fewer distinct eigenvalues than $\nabla^2 F$ and a condition number closer to 1. We would then use the conjugate gradient method to solve a preconditioned system of equations

$$\left(M^{-1}\nabla^2 F\right) p \approx -M^{-1}\nabla F. \tag{25}$$

In practice, finding a good preconditioner is important to making the truncated-Newton method competitive [55]. We prefer to find a preconditioner such that the $\mathrm{cond}\left(M^{-1}\nabla^2 F\right) \approx 1$ and M^{-1} is easy to apply [23].

For ML and MAP it is possible to find the exact diagonal Hessian preconditioner (the diagonal matrix whose diagonal elements are exactly equal to the

diagonal elements of the Hessian) at a reasonable cost. For MAP we have for the diagonal terms of the Hessian (recalling again (3)):

$$\left(\nabla^2 F\right)_{i,i} = \sum_j \frac{C_{i,j}^2 y_j}{\hat{y}_j^2} + \gamma \frac{\partial^2 R\left(\theta\right)}{\partial \theta_i^2} + \frac{\mu}{\theta_i^2}. \tag{26}$$

Note that the first right-hand side term in (26) is similar in form to a back projection, although a bit more expensive due to the squaring operations.

We have found that in MAP reconstruction problems, solving (25) using the exact diagonal Hessian preconditioner to the tolerance defined in (22) usually requires fewer than 10 conjugate gradient iterations, regardless of the size of the problem! In many cases, only 3 or 4 conjugate gradient iterations are required. The directions produced by (25) with a diagonal Hessian preconditioner are well scaled (usually allowing step sizes of near 1) and lead to rapid descent. We have investigated a number of other preconditioners, including FFT-based preconditioners that model the approximately Toeplitz-block-Toeplitz nature of $\nabla^2 F$ with a circulant-block-circulant approximation [7, 9], high-pass filter approximations to the FFT-based preconditioner [10], the EM preconditioner $\mathrm{diag}(\theta_i/q_i)$ [43], and diagonal Hessian approximations [55]. These other preconditioners performed poorly in comparison to the exact diagonal Hessian preconditioner. In fact, the performance of these other preconditioners was only slightly better than that of the identity matrix (solving (21) without a preconditioner). It has been reported that block-circulant preconditioners are the most effective for solving conjugate gradient systems of equations in weighted least squares reconstructions [9]. We suspect that the observed poor performance of the block circulant preconditioners may be due to that diagonal barrier terms in $\nabla^2 F$ that further corrupt the approximately Toeplitz-block-Toeplitz nature of $\nabla^2 f_{ML}$. That is, after adding the regularizing and barrier terms to $\nabla^2 f_{ML}$, the block circulant approximation may no longer be very accurate.

4.1.3 Line Search

A potentially expensive component of the unconstrained minimization problem is the line search for the determination of the step size α. For ML and MAP reconstructions, knowledge of the structure of the objective function can lead to a substantial reduction in the cost of implementing a line search over a more naive approach. Specifically, after the search direction p has been found, and once a forward projection $\hat{w} = C^T p \in \Re^N$ has been computed, it is possible to compute the objective function and directional derivative values at the trial points $\left(\theta^k + \alpha p\right)$ at little additional cost. To see this, note that

$$\hat{y}^{k+1} = C^T \theta^{k+1} = C^T \left(\theta^k + \alpha p\right) = \hat{y}^k + \alpha \hat{w}.$$

Since (1) can be rewritten as

$$f_{ML}\left(\theta\right) = -\sum_j \left(\hat{y}_j^k - y_j \log \hat{y}_j^k\right),$$

we have [39] that

$$f\left(\theta^k + \alpha p\right) = \sum_j \left(\left(\hat{y}_j^k + \alpha\hat{w}_j\right) - y_j \log\left(\hat{y}_j^k + \alpha\hat{w}_j\right)\right) + \frac{1}{2}\gamma R\left(\theta^k + \alpha p\right). \quad (27)$$

Similar expressions exists for the directional derivatives [52]:

$$\begin{aligned} \frac{d}{d\alpha} f\left(\theta^k + \alpha p\right) &= \sum_j \left(w_j - \frac{y_j \hat{w}_j}{\hat{y}_j^k + \alpha\hat{w}_j}\right) + \gamma p^T \nabla_\theta R\left(\theta + \alpha p\right) \quad (28) \\ \frac{d^2}{d\alpha^2} f\left(\theta^k + \alpha p\right) &= \sum_j \left(\frac{y_j \hat{w}_j^2}{\left(\hat{y}_j^k + \alpha\hat{w}_j\right)^2}\right) + \gamma p^T \nabla_\theta^2 R\left(\theta + \alpha p\right) p \end{aligned}$$

Using equations (27) and (28), the function and directional derivative along p for various trial points of α can be evaluated at very little computational cost. After the initial forward projection to compute $\hat{w}$, no further projection operations are required during the line search. The forward projection $\hat{w}$ can be re-used, so that only one back projection is subsequently required to update the gradient.

The above observations and the well behaved convex nature of the objective function have permitted us to implement a highly accurate but low-cost Newton line search. We initialize the search with

$$\alpha_0 = \min\left(1, 0.9995\alpha_{\max}\right)$$

where $\alpha_{\max}$ is the maximum step length that does not violate non-negativity. We then take Newton steps on α :

$$\alpha \leftarrow \alpha - \frac{\frac{d}{d\alpha} f\left(\theta^k + \alpha p\right)}{\frac{d^2}{d\alpha^2} f\left(\theta^k + \alpha p\right)}$$

until the Wolfe condition

$$\left|p^T \nabla F\left(\theta^k + \alpha p\right)\right| \leq \eta \left|p^T \nabla F\left(\theta^k\right)\right|$$

is met. Due to the low cost of each step we have chosen a relatively strict Wolfe tolerance of $\eta = 0.05$. Since the vectors y, $\hat{y}$, and $\hat{w}$ are distributed across the processors, the summations in (27) and (28) are readily parallelized, further reducing computational expense. The MAP components parallelize easily through a decomposition in image space. We find this line search technique to be highly effective and, in no small part, responsible for the positive results we report. The Wolfe criterion to $\eta = 0.05$ is usually satisfied in only a few steps. Being that the directions are well scaled, the trial point $\alpha = 1$ satisfies the Wolfe criterion in many cases, precluding the need for a line search at all when that occurs.

4.1.4 Stabilization

A well known property of the logarithmic barrier method is that as $\mu \to 0$, the Hessian of the barrier objective function becomes increasingly ill conditioned [54]. This manifests itself as follows. Consider the index set $\mathcal{I}$ of binding constraints (the indices i for which $\theta_i = 0$) at the optimal solution. Define N to be the Jacobian matrix of the binding constraints (for non-negativity constraints, N is simply formed from columns of the identity matrix), $N^{\#}$ to be a left pseudo-inverse of N, and Z to be a basis matrix for the null space of N^T. Recalling (24) and denoting $\theta = \theta(\mu)$, we define $\Theta_{\mathcal{I}} = \text{diag}(\theta_i(\mu),\ i \in \mathcal{I})$ and $\Theta_{\mathcal{J}} = \text{diag}(\theta_i(\mu),\ i \notin \mathcal{I})$. For lower-bound constraints, we may rewrite (24) as

$$\nabla^2 F(\theta) = H + \mu N \Theta_{\mathcal{I}}^{-2} N^{\#} \tag{29}$$

where

$$H = \nabla^2 f(\theta) + \mu Z \Theta_{\mathcal{J}}^{-2} Z^T. \tag{30}$$

As $\mu \to 0$ and the optimal solution is approached, $\theta_i \to 0,\ i \in \mathcal{I}$, but $\theta_i,\ i \notin \mathcal{I}$ tend away from zero, and thus H remains relatively well conditioned. Thus we make the observation that as $\mu \to 0$, H remains well conditioned while $\nabla^2 F(\theta)$ becomes increasingly ill conditioned.

Nash and Sofer have developed an approximation to the Newton direction that avoids this structural ill conditioning [59]. For small μ we can approximate the Newton direction by $p \approx p_1 + \mu p_2$, where

$$\begin{aligned} p_1 &= -Z\left(Z^T H Z\right)^{-1} Z^T \nabla F & (31)\\ \zeta &= N^{\#}(H p_1 + \nabla F) & (32)\\ p_2 &= -\left(N^{\#}\right)^T \Lambda_{\mathcal{I}}^{-2} \zeta. \end{aligned}$$

where $\Lambda_{\mathcal{I}} = diag(\lambda_i(\mu),\ i \in \mathcal{I})$and $\lambda_i(\mu)$ are defined by (16). Equation (31) can be solved using the conjugate gradient method on the well-conditioned matrix $Z^T H Z$ and equation (32) requires one additional Hessian-vector product of the type previously described. The index set $\mathcal{I}$ is based on a prediction of the set of active constraints at the solution, $i \in \mathcal{I}$ if $\theta_i \leq 100\mu$. Nash and Sofer have proved that, for sufficiently small μ, the vector p computed using the above approximation yields a descent direction with respect to the logarithmic barrier objective function [59]. We have found through empirical evidence that for MAP reconstructions, a reasonable rule for incorporating stabilization is $\mu \leq 2 \cdot 10^{-4}$. This generally corresponds to the last barrier subproblem of 3-D reconstructions.

4.1.5 Initializing the Barrier Parameter

The choice of the initial barrier parameter may have a substantial effect on the algorithm. If the parameter is too small, the first subproblem may have extreme difficulty due to ill conditioning; if the parameter is too large, then many (unnecessary) subproblems will be required to solve the problem. Proper

initialization of the barrier parameter μ involves finding the most suitable point on the barrier trajectory based on the initial solution θ^o and the observation data y. Recalling the perturbed optimality conditions in (17), we write

$$\nabla F(\theta, \mu) = q - C\hat{Y}^{-1}y + \gamma \nabla R(\theta) - \mu \Theta^{-1} e_n = 0.$$

Post-multiplying by θ we arrive at

$$q^T\theta - y^T e_N + \gamma \theta^T \nabla R(\theta) = n\mu,$$

or

$$e_N^T \left(C^T\theta - y\right) + \gamma \theta^T \nabla R(\theta) = n\mu.$$

This suggests the following rule for initialization, which we find quite effective:

$$\mu_0 = \frac{e_N^T \left(C^T\theta^0 - y\right) + \gamma \theta^{0T} \nabla R\left(\theta^0\right)}{n}. \tag{33}$$

The initial solution we used most frequently was a uniform field, although other choices may be preferable. An alternative scheme for initializing μ is described in Section 4.3.

4.1.6 Extrapolation

The set of solutions $\theta(\mu)$ defines a unique differentiable trajectory. Thus the solution to the previous subproblems may assist in predicting a solution to the next subproblem. This prediction can serve as a starting point for the next subproblem. We have implemented linear, quadratic, and cubic extrapolation to predict the solutions to subproblems 3, 4, and 5+, respectively. If the extrapolated solution is infeasible, we move along the direction of extrapolation to the extrapolation $\theta \leftarrow \theta + \bar{\alpha}\Delta\theta$, where $\bar{\alpha} = 0.98\alpha_{\max}$ and $\alpha_{\max}$ is the maximum feasible step size to the constraint boundary. In MAP reconstructions, extrapolation generally eliminates one to three truncated Newton steps per subproblem.

4.1.7 Termination

Termination of the logarithmic barrier method should be based on closeness to the solution defined by the Karush-Kuhn-Tucker (KKT) first-order necessary conditions in (12-14). (The second-order necessary condition is satisfied automatically by the known convexity of $F(\theta)$.) We therefore need to define tolerances on the optimality criteria and estimates of the Lagrange multipliers for the purpose of defining a reasonable termination point. At the optimal solution, the Lagrange multipliers may be written

$$\lambda_i^* = \begin{cases} \frac{\partial f(\theta^*)}{\partial \theta_i} & \text{for } \theta_i^* = 0 \\ 0 & \text{for } \theta_i^* > 0 \end{cases}. \tag{34}$$

In our implementation, we define four termination rules, two for subproblem termination and two for final termination. A commonly used rule for subproblem termination is

$$\left\|\nabla F\left(\theta,\mu_k\right)\right\|_\infty \leq \varepsilon. \tag{35}$$

We have been able to achieve more consistent results by using

$$\frac{\left|F\left(\theta\left(\mu\right)^k\right) - F\left(\theta\left(\mu\right)^{k-1}\right)\right|}{\left|F\left(\theta\left(\mu\right)^k\right)\right|} \leq \varepsilon_1 \tag{36}$$

where $\varepsilon_1 = 10^{-6}$ for the thick-slice reconstructions. Specifying the accuracy of a subproblem solution is a tradeoff between time spent in the current subproblem and (potentially) time spent in the next. For MAP reconstruction, we have discovered that the barrier method works best with a relatively accurate subproblem solution. As such, our implementation of the barrier method is a "long-step" algorithm. The other requirement for subproblem termination is

$$\min\left(\lambda_i\left(\mu\right)\right) \geq \varepsilon_2. \tag{37}$$

Final termination is based on norms of the gradient of the Lagrangian and complementarity vector. The criteria for final termination are

$$\left\|\nabla_\theta \ell\left(\theta,\lambda\right)\right\|_\infty \leq \varepsilon_3 \tag{38}$$

$$\max\left|\theta_i \lambda_i\right| \leq \varepsilon_4. \tag{39}$$

For the logarithmic barrier method, the criteria in (37), (38), and (39) refer to the estimates of the dual variables at the solution as defined by (34). In practice, we find it necessary to fix $\lambda_i = 0$ if $\theta_i \geq \delta$, since some of the nonbinding θ_i can be quite large and the $\partial f/\partial\theta_i$ never reach exactly zero in finite precision. The parameter δ should be chosen to reflect the threshold below which variables are to be considered binding. We have found that for thick-slice 3-D reconstructions, reasonably accurate solutions are ensured if we choose $\varepsilon_2 = -10^{-5}$, $\varepsilon_3 = .02$ and $\varepsilon_4 = .002$.

4.2 Primal-Dual Methods

Primal-dual methods are another class of interior point methods. Unlike the logarithmic barrier method, which estimates only the primal variables, primal-dual methods simultaneously estimate both the primal variables θ and dual variables λ. Like the logarithmic barrier method, the primal-dual methods generate estimates of the variables that lie on a differentiable trajectory, and the points on this trajectory are characterized by a barrier parameter μ. Solutions at a particular point μ along the barrier trajectory satisfy the perturbed optimality conditions (17-19). Primal-dual methods have enjoyed great success in linear programming problems [41] and have very recently been proposed for nonlinear optimization [11, 20]. These methods maintain primal feasibility and

dual feasibility to within a specified tolerance while reducing the *duality gap* in the complementary slackness as μ is reduced and the optimal primal-dual solution is approached. The primal-dual algorithm for MAP reconstruction that we present here closely follows the method recently proposed by Conn, Gould, and Toint [11].

From the perturbed optimality conditions (17) and defining $v = \left[\begin{array}{cc} \theta^T & \lambda^T \end{array} \right]^T$, we have

$$g(v) = \left[\begin{array}{c} \nabla f(\theta) - \lambda \\ \Lambda\Theta e_n - \mu e_n \end{array} \right] = 0. \tag{40}$$

Taking a linear approximation to (40) and defining $p = \left[\begin{array}{cc} p_\theta^T & p_\lambda^T \end{array} \right]^T$,

$$g(v+p) \approx g(v) + \nabla g(v)\, p = 0,$$

we can simultaneously solve the Newton equation for the primal and dual directions:

$$\left[\begin{array}{ll} \nabla^2 f(\theta), & -I \\ \Lambda, & \Theta \end{array} \right] \left[\begin{array}{c} p_\theta \\ p_\lambda \end{array} \right] = - \left[\begin{array}{c} \nabla f(\theta) - \lambda \\ \Lambda\Theta e - \mu e \end{array} \right].$$

This yields the system

$$(\nabla^2 f(\theta) + \Theta^{-1}\Lambda) p_\theta = -\nabla f(\theta) + \mu\Theta^{-1} e_n \tag{41}$$

$$p_\lambda = -\lambda - \Theta^{-1}\Lambda p_\theta + \mu\Theta^{-1} e_n. \tag{42}$$

Observe that (41) is very similar in form to (20) and can be approximately solved using the conjugate gradient method in the same manner as (21). The formulae presented earlier for computing the Hessian-vector product and exact diagonal Hessian preconditioner also apply, with very minor modifications, to the conjugate gradient computation of p_θ. From (42), computing the dual direction vector involves only basic operations on vectors. We then move the primal and dual solution vectors according to

$$\left[\begin{array}{c} \theta^{k+1} \\ \lambda^{k+1} \end{array} \right] = \left[\begin{array}{c} \theta^k + \alpha_\theta p_\theta \\ \lambda^k + \alpha_\lambda p_\lambda \end{array} \right].$$

In this algorithm the primal and dual step sizes may be chosen independently, provided that certain conditions beyond non-negativity are satisfied for the dual step.

In primal-dual methods, a merit function that measures the progress of the algorithm needs to be defined. The primal step size is chosen in such a way as to maintain primal feasibility and to ensure sufficient decrease in the merit function. If the merit function is chosen properly, sufficient decrease in the merit function should permit a reduction in the tolerance of the dual feasibility condition $\|\nabla \ell(\theta, \mu)\| = 0$. There is a close connection between the primal-dual method and the logarithmic barrier method in that both methods have the same perturbed optimality conditions (17) at subproblem solution. This close connection to the logarithmic barrier method motivates the merit function

$$F(\theta,\mu) = f(\theta) - \mu \sum_i \log \theta_i.$$

Note that $F(\theta,\mu)$ is simply the logarithmic barrier function and that

$$\nabla_\theta F(\theta,\mu) = \nabla f(\theta) - \mu \Theta^{-1} e_n$$

is identical to the right-hand side of (41). Conn et al. have proved that $p_\theta^T \nabla_\theta F(\theta,\mu) \leq 0$ [11], so p_θ is a descent direction for the merit function. Thus the line search technique developed in Section 4.1.3, which maintains primal feasibility, applies equally well to finding α_θ for the primal-dual method.

The formula for the dual step size α_λ is as follows. If $\lambda^{(k)} + p_\lambda$ lies componentwise in the interval

$$\lambda^k + p_\lambda \in \tag{43}$$
$$\left[.01 \min\left(e_n, \lambda^k, \mu_k \Theta_{k+1}^{-1} e_n\right), \max\left(100 e_n, \lambda^k, 100 \mu_k^{-1} e_n, 100 \mu_k \Theta_{k+1}^{-1} e_n\right)\right]$$

then $\lambda^{k+1} = \lambda^k + p_\lambda$. Otherwise find $0 < \alpha_\lambda < 1$ such that $\lambda^{k+1} = \lambda^k + \alpha_\lambda p_\lambda$ minimizes

$$\|\Lambda_{k+1} \Theta_{k+1} e_n - \mu_k e_n\| \tag{44}$$

subject to $\lambda^{(k+1)}$ being in the interval (43). These conditions on the dual step might appear at first glance to be overly restrictive but are actually designed to give maximum flexibility in the choice of λ^{k+1} [21]. In practice we find that both the primal and dual direction vectors are well scaled in MAP reconstructions, and that α_θ and α_λ are typically close to 1. We emphasize that the computational expense of the dual update is quite small in comparison to the forward and back projection costs.

Setting the barrier parameter μ is an important consideration in the primal-dual algorithm. Prior to adjusting μ we must satisfy conditions on complementary slackness and gradient of the Lagrangian:

$$\frac{\left(\lambda^{k+1}\right)^T \theta^{k+1}}{n} \leq \vartheta^C \mu_k \tag{45}$$

and

$$\left\| \nabla f\left(\theta^{k+1}\right) - \lambda^{k+1} \right\| \leq \vartheta^{DF} \mu_k$$

where ϑ^C and ϑ^{DF} are constant parameters. If the above conditions are satisfied, we reduce the barrier parameter according to

$$\mu_{k+1} = \frac{\left(\lambda^{k+1}\right)^T \theta^{k+1}}{n\rho}. \tag{46}$$

From empirical evidence in our computational studies, we have found that a "short-step" algorithm with gradual reduction in μ achieves the fastest convergence to the KKT conditions. Specifically we define $\vartheta^C = 1.9$, $\rho = 2$, and $\vartheta^{DF} = 100$. We have also discovered, however, that a trade-off exists whereby higher values of ρ lead to faster early convergence at the expense of overall convergence. An "rapid early progress" variant of our algorithm thus consists of defining $\vartheta^C = 99$, $\rho = 100$, and $\vartheta^{DF} = 100$. In both variants of the algorithm, subproblems usually consist of only one truncated-Newton step, except for the first subproblem which can often require two steps. Both variants of the algorithm satisfy Conn, Gould, and Toint's "μ-critical" subproblem condition, which is key to their convergence proof [11].

4.2.1 Primal-Dual Initialization and Termination

The initialization strategy of equation (33), being motivated by the perturbed optimality conditions, may also be used in the primal-dual algorithm. We advise against using (46) for initialization however, since the choices of θ^0 and λ^0 are essentially arbitrary and because (46) contains no dependence on the observation data. For both θ^0 and λ^0 we use uniform fields, although other possibilities exist for θ^0.

Final termination of the primal-dual algorithm requires only tests of the gradient of the Lagrangian and complementary slackness. Since the dual line search ensures that $\lambda > 0$ at all times, we need not test that condition. Convexity of the objective function guarantees that the second-order necessary KKT condition will be satisfied. For the test on the gradient of the Lagrangian we use (38) with $\varepsilon_3 = .02$ as we did with the logarithmic barrier method. In the primal-dual method, however, $\ell(\theta, \lambda)$ is computed with the actual dual variables rather than an estimate of the Lagrange multipliers at the optimal solution. For the test on complementary slackness, we drop (39) in favor of

$$\frac{\lambda^T \theta}{n} \leq \varepsilon_5. \tag{47}$$

In our tests we have used $\varepsilon_5 = 1.5 \cdot 10^{-4}$, which yields roughly the same tolerance as using (39) with $\varepsilon_4 = .002$. Athough we use (47) as our primal-dual termination criterion for complementary slackness, in our computational studies we do track $\max(\lambda_i \theta_i)$.

4.2.2 Primal-Dual Stabilization

The stabilization technique of Section 4.1.4 can readily be applied to the computation of the primal direction in the primal-dual algorithm with only minor modifications. The primary difference is that we are now able to use the actual dual variables. Defining $\Lambda_{\mathcal{I}} = \text{diag}(\lambda_i,\ i \in \mathcal{I})$ and $\Lambda_{\mathcal{J}} = \text{diag}(\lambda_i,\ i \notin \mathcal{I})$, the matrix in the left-hand side of (41) may be rewritten as

$$B = G + N \Theta_{\mathcal{I}}^{-1} \Lambda_{\mathcal{I}} N^{\#}$$

where

$$G = \nabla^2 f(\theta) + Z\Theta_{\mathcal{J}}^{-1}\Lambda_{\mathcal{J}} Z^T.$$

As $\mu \to 0$, $\theta_i \to 0$, $i \in \mathcal{I}$, and B becomes increasingly ill conditioned while G avoids the structural ill conditioning. (Note that equations (29) and (30) may also be written in the above form, since $\lambda_i(\mu) = \theta_i(\mu)/\mu$ for the logarithmic barrier method). We thus define a prediction of the active set at the solution as we did for the logarithmic Barrier method. For small μ, say $\mu \leq 2 \cdot 10^{-4}$, we can approximate the Newton direction by $p_\theta \approx p_\theta^1 + \mu p_\theta^2$, where

$$\begin{aligned} p_\theta^1 &= -Z\left(Z^T G Z\right)^{-1} Z^T \nabla F \\ \zeta &= N^{\#}\left(G p_\theta^1 + \nabla F\right) \\ p_\theta^2 &= -\left(N^{\#}\right)^T \Theta_{\mathcal{I}}^{-1}\Lambda_{\mathcal{I}}\zeta, \end{aligned}$$

and $-\nabla F$ is identical to the right-hand side of (41).

In our experience, the above stabilization technique has been quite effective in the last two or so primal-dual subproblems. In the absence of stabilization, near the primal-dual solution, either the primal direction or dual direction can often be poorly scaled, and the conjugate gradient can take more than 10 subiterations. With stabilization, both the primal and dual directions are generally well scaled near the solution, with a further benefit of a reduction in the cost of the conjugate gradient.

4.2.3 Primal-Dual Extrapolation

Since the primal-dual method follows a unique trajectory in μ, one may be able to use the solutions of previous subproblems to predict the next subproblem. Our experience in MAP reconstructions is that the best results are obtained by extrapolating the primal solution $\theta^{new} \leftarrow \theta^{old} + \bar{\alpha}\triangle\theta$, where $\triangle\theta$ is the direction of extrapolation and $\bar{\alpha}$ is used to maintain primal feasibility, and then (in the manner of (42)) find a dual direction vector according to

$$\Delta\lambda = -\lambda^{old} - \Theta_{old}^{-1}\Lambda\left(\bar{\alpha}\triangle\theta\right) + \mu\Theta_{old}^{-1} e_n.$$

The dual vector is then moved according to $\lambda^{new} \leftarrow \lambda^{old} + \alpha_\lambda \triangle\theta$, which requires another dual line search to minimize (44).

Following extrapolation, a gradient evaluation is required to update the vector ∇F. Since the primal-dual algorithm requires between 12 and 25 subproblems to perform a 3-D MAP reconstruction, extrapolation adds that many gradient evaluation operations to the computational cost. So extrapolation is only economical is it reduces the computational burden by at least as much as it adds. Our experience has been that for some data sets, the cost of extrapolation is well worthwhile but for other data sets the cost of extrapolation was slightly greater than the savings. Extrapolation thus appears to serve as somewhat of a safeguard against difficult problems. In an extrapolated primal-dual reconstruction, the convergence measure $\max(\lambda_i\theta_i)$ does not decrease as

monotonically as in a primal dual reconstruction without extrapolation. Certain extrapolated steps seem to cause the algorithm to "get ahead of itself," but this effect is transient. On the studies we've performed, the algorithm does ultimately converge to an accurate solution with extrapolation.

4.3 Sparse Implementation of Interior Point Methods

The largest of the reconstruction problems (the "thin slice" reconstructions of Table 1) pose a special challenge. Notice in Table 1 that a "full" gradient evaluation requires 4 hours whereas a "sparse" gradient evaluation requires only 10 minutes for a 2.5 million count study on 10 IBM RS/6000 SP2 processors. This rather wide disparity is due the sparsity of y. Again referring to Table 1, a data set with 2.5 million counts will have at most 4% of coincidence lines occupied in the thin-slice case. This is significant because most computations do not require a visit to every projection line. For example, the back projection operation may be implemented as follows to exploit sparsity:

$$\nu_i = \sum_{j:y_j \neq 0} \frac{C_{ij} y_j}{\hat{y}_j}.$$

The presence of y_j in the numerator obviates the need for visits to the unoccupied projection lines.

Similar sparse implementations exist for the Hessian diagonal and Hessian-vector product calculations, where in both cases y_j sits in the numerator. Even the forward projection operation can be performed on only those $\hat{y}_j$ for which $y_j \neq 0$. The question is, can we perform an interior-point reconstruction without ever computing those $\hat{y}_j$ in the sparse subspace of y? The answer is yes, but with a few minor modifications to the methods outlined above.

One operation that clearly requires correct values for *all* $\hat{y}_j$ is the objective function evaluation,

$$f(\theta) = \sum_j \left(\hat{y}_j - y_j \log \hat{y}_j\right).$$

Objective function values are not necessary for the purpose of performing a reconstruction. As described in Sections 4.1.7 and 4.2.1, they were used for defining a termination measure for the barrier subproblem. Clearly, the sparse implementation cannot use (36) as a subproblem termination rule. Rather, it uses (35):

$$\left\|\nabla F(\theta, \mu)\right\|_\infty \leq \varepsilon_6$$

We have found that a reasonable value for ε_6 is 0.5.

Another operation that must be modified for thin slice reconstructions is the initialization procedure defined by (33), which requires the full set of projection lines. We can consider alternative ways of finding an initial value of μ that is reasonably close to the point on the barrier trajectory for the initial solution θ^0.

The goal is to find a μ_0 such that

$$\nabla F\left(\theta^0, \mu_0\right) = \nabla f\left(\theta^0\right) - \mu\Theta^{-1}e \approx 0.$$

One such initialization rule that produces results similar to those of (33) is [60]

$$\mu_o = \frac{\left\|\nabla f\left(\theta^0\right)\right\|_2}{\left\|\Theta^{-1}e\right\|_2}.$$

The operation most severely affected by the sparse implementation is the line search. Although the line search does not require a function evaluation, the directional first and second derivatives, defined by (28), require $\hat{y}_j \ \forall j$. In thin slice reconstructions, however, we forward project onto only the dense subspace of y. Thus we must replace (28) with

$$\begin{aligned}
\frac{d}{d\alpha} f\left(\theta^k + \alpha p_\theta\right) &= -p^T \nabla_\theta f_{ML}\left(\theta^k + \alpha p_\theta\right) + \gamma p_\theta^T \nabla_\theta R\left(\theta^k + \alpha p_\theta\right) \\
\frac{d^2}{d\alpha^2} f\left(\theta^k + \alpha p_\theta\right) &= -p_\theta^T \nabla_\theta^2 f_{ML}\left(\theta^k + \alpha p_\theta\right) p_\theta + \gamma p_\theta^T \nabla_\theta^2 R\left(\theta^k + \alpha p_\theta\right) p_\theta.
\end{aligned}$$

One Newton step of the sparse implementation of the line search thus requires one back projection (to compute $\nabla_\theta f_{ML}\left(\theta^k + \alpha p_\theta\right)$ from $\hat{y}^k + \alpha\hat{w}$) and one Hessian-vector product. Clearly this increases the cost of the line search. To reduce the computational burden of the line search, we have raised the Wolfe tolerance of the sparse implementation to $\eta = 0.25$. As we mentioned earlier, the line search converges quickly, and few Newton steps are required in practice.

5 Computational Studies

In order to test the convergence properties of the interior point methods and to compare the computational requirements of the interior point methods against MAP-EM, we have performed a number of reconstructions on data acquired from a small animal scanner and data generated by a Monte Carlo simulation of the same small animal scanner. The scanner consists of two opposed NaI(Tl) continuous-slab detector plates, each $50 \times 50 \times 4$-mm, optically coupled to position sensitive photomultiplier tubes and separated by 128-mm [63]. This scanner can accept axial coincidences of up to 21.9^o. A small animal rotates at least one full revolution on a turntable that is centered between the detector plates. List-mode data from the scanner are binned into "thick-slice" or "thin-slice" projection sinograms as specified in Table 1. Corrections for redundant sampling and radioactive decay have been incorporated into the system matrix; no other physical corrections have been applied.

Reconstructions were performed on six data sets, described below. Four of the data sets are of actual small animal data; two data sets are from Monte Carlo simulated data. The Monte Carlo simulation realistically models positron range and non-colinearity as well as depth of interactions effects and crystal scatter

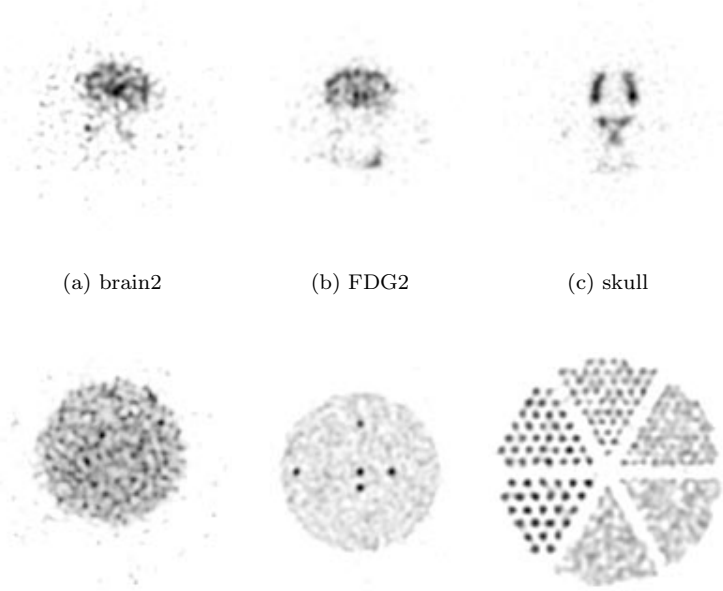
study	x	f^*	*npr*	*nit*	*ncg*	*ngr*	MAP-EM
brain		2,465,770	19	19	110	148	1000
fdg	*	2,397,232	16	16	91	139	> 1000
skull		2,269,180	22	22	126	170	990
uniform	*	2,752,502	14	16	119	165	> 1000
poi/cyl		2,536,110	26	26	131	183	770
Derenzo		3,660,344	24	24	127	175	> 1000

Table 2: Summary of thick slice primal-dual results and comparison with MAP-EM. An asterisk in the column marked x indicates that extrapolation was used; in all cases $\rho = 2$.

but ignores random coincidences, the effects of attenuation and object scatter, and spatial non-uniformities in the scanner's resolution. The resolution model in the system matrix is a simplified uniform 1.2-mm FHWM in all 3 directions (radial, tangential, axial) across the entire field of view of the scanner. The region of support of each reconstruction is a cylinder inscribed in a $(50\text{-mm})^3$ box. Figure 1 depicts central transverse slices of the reconstructed images (from the primal-dual method) of the data sets described below.

- *brain2*- FDG study of a rat brain, 1.5M counts.
- *FDG2* - FDG study of a (different) rat brain, 2.3M counts.
- *skull* - F-18 study of a rat skull. Prominent in the displayed section is the mandible; 1.5M counts.
- *uniform* - Physical uniform cylinder phantom. Observe the relative radial uniformity of the reconstructed image, demonstrating the relative lack of attenuation due to the small size of the object; 1.9M counts.
- *poi/cyl* - Monte Carlo simulation of point sources in a warm cylinder background; 5.1M counts.
- *Derenzo* - Monte Carlo simulation of the Derenzo phantom. The diameters of the capillary tubes, in counter-clockwise order, are 0.6-mm, 0.8-mm, 1.0-mm, 1.2-mm, 1.5-mm, 2.0-mm.

Reconstructions of these six data sets were taken to full convergence. A prior strength of $\gamma = 3 \cdot 10^{-4}$ was used with a 10-pixel neighborhood. This prior strength was determined by finding the "elbow" of the noise/resolution curve from 2-D reconstructions and subsequent adjustment based on visual inspection of the 3-D reconstructions. The primal-dual method terminated upon satisfaction of (38) with $\varepsilon_3 = .02$ and (5) with $\varepsilon_5 = 1.5 \cdot 10^{-4}$. Table 2 compares the computational expense required by the primal-dual method to reach convergence and compares it to that required of DePierro's MAP-EM to reach the same objective function value f^*. Although we generally compare the interior-point



(a) brain2 (b) FDG2 (c) skull

(d) uniform (e) poi/cyl (f) Derenzo

Figure 1: Central transverse sections of the thick-slice primal-dual reconstructions.

results with DePierro's MAP-EM, we have found that DePierro's MAP-EM and OSL produce very similar results, the former being perhaps one iteration ahead of the latter.

We have devised a metric to measure the cost of an interior point reconstruction. Define the number of subproblems to be npr, the number of truncated-Newton iterations nit, the number of conjugate gradient subiterations ncg. The cost of one conjugate gradient iteration (dominated by the Hessian-vector product) is equivalent to the cost of one gradient evaluation or EM iteration. One truncated-Newton iteration requires, in addition to the ncg operations, one diagonal Hessian evaluation plus a forward projection and a back projection. The exact cost of these operations varies depending on the size of the problem and number of counts, but we shall approximate the cost one truncated-Newton iteration to be the equivalent of two EM iterations beyond the cost of the conjugate gradients. Thus in unextrapolated interior-point reconstructions, the total cost of the reconstruction can be measured in units of equivalent number of gradient evaluations (or EM iterations):

$$ngr = 2 \cdot nit + ncg.$$

Extrapolation requires an additional gradient evaluation following the extrapolation in order to update the gradient vector. With extrapolation we modify the formula to

$$ngr = npr + 2 \cdot nit + ncg.$$

In Table 2 the column marked MAP-EM lists the number of MAP-EM iterations required to reach the objective function f^* that required ngr iterations of the primal-dual method. Two of the primal-dual results are extrapolated; for brevity we have not included a complete comparison of extrapolated and unextrapolated primal-dual results but in general extrapolation provides some improvement. Comparing the last two columns of Table 2, the primal-dual method clearly offers significant improvement in convergence to the optimal MAP solution. Table 3 provides a similar comparison between the barrier method and MAP-EM on the same data sets. Extrapolation was used on all barrier studies. While still offering significant convergence acceleration over MAP-EM, the barrier method required half again as much work to reach convergence compared with the primal-dual method. These results are quite consistent across all data sets.

As a sample test case, we consider the progress of the reconstruction of the Derenzo phantom from one subproblem to the next. The most relevant convergence parameters are listed in Tables 4 and 5 for primal-dual and barrier, respectively. In Table 4, the complementary slackness $\lambda^T\theta/n$ can be discerned by μ which was calculated from (46) with $\rho = 2$. In the barrier method we reduced the barrier parameter according to $\mu_{k+1} = \mu_k/\rho$ with $\rho = 10$ for all studies. We have explored other values of ρ for the barrier method as well as other schemes for setting μ but the results were inconclusive. To conserve computation, function evaluations were only performed every 10th MAP-EM iteration. This permitted the use of "sparse" EM iterations (recall from Table

study	f^*	*npr*	*nit*	*ncg*	*ngr*	MAP-EM
brain	2,465,832	5	28	159	218	880
fdg	2,397,199	5	29	198	259	> 1000
skull	2,269,180	5	29	185	246	990
uniform	2,752,499	4	25	207	260	> 1000
poi/cyl	2,536,111	6	40	214	298	780
Derenzo	3,660,351	6	41	214	300	> 1000

Table 3: Summary of thick-slice barrier results and comparison with MAP-EM. Extrapolation was used on all data sets, and in all cases $\rho = 10$.

1 that even in thick-slice reconstructions, full gradient evaluations are more expensive than their sparse equivalents). As such, the number of iterations in the MAP-EM column of Tables 4 and 5 is specified over a range of 10 rather than an exact integer. Being a "long-step" algorithm, the barrier method requires fewer subproblems to reach convergence than primal-dual, but these subproblems require much more computation, as witnessed by a comparison of Tables 4 and 5.

As we noted in Section 4.2, for rapid convergence to the optimal primal-dual solution, the barrier parameter should be reduced gradually. Empirical evidence to that effect motivated our choice of $\rho = 2$ for the above studies. We have also discovered, however, that a more aggressive drop in the barrier parameter using, say, $\rho = 100$ can lead to more rapid early progress at the expense of more overall work to the optimal primal/dual solution. This observation of rapid early progress is in evidence in Table 6 where the Derenzo phantom was reconstructed to a less accurate solution using 9 primal-dual steps and $\rho = 100$. Figure 2 displays central slice reconstructions of the Derenzo phantom reconstructed to various stopping points of various algorithms. The fully converged MAP-EM reconstruction is nearly identical to the fully converged MAP primal-dual reconstruction in Figure 1. The early terminated reconstructions include a positive bias in the regions between the capillary tubes in a wedge.

As a demonstration of the sparse implementation of the primal-dual method, we have performed a thin-slice reconstruction of the FDG2 data set. This data set was selected because of its visual appeal, as evidenced in Figure 3 where transverse, coronal, and sagittal sections of the reconstruction are displayed. This reconstruction reveals activity in cortical and subcortical structures of an intact rat brain. For this reconstruction, we have used an 18-pixel neighborhood and a prior strength of $\gamma = 8 \cdot 10^{-5}$. Although this prior strength is smaller than that used for the thick-slice reconstructions, the overall strength of the regularizer is similar or slightly stronger (due to the greater number of variables and larger neighborhood).

Recall from Table 1 that thin-slice reconstructions involve 1.4M variables and 63M projection lines. For the sparse small animal data, the sparse implementation of the primal-dual method avoids the enormous cost of performing full forward projections. The relevant parameters of this reconstruction are

μ	f	$\|\nabla\ell\|$	$\max(\lambda_i\theta_i)$	*niter*	*ncg*	*ngr*	MAP-EM
10.86	4,429,471	4.044	14.538	1	2	4	1-10
5.088	4,039,659	0.273	8.281	2	4	8	1-10
2.617	3,894,986	0.303	3.307	3	6	12	1-10
1.418	3,773,195	0.0973	1.641	4	9	17	1-10
0.731	3,719,320	0.0792	0.874	5	13	23	1-10
0.374	3,692,283	0.0564	0.690	6	17	29	11-20
0.190	3,677,361	0.0521	0.221	7	21	35	21-30
0.0963	3,668,958	0.0314	0.110	8	27	43	41-50
0.0502	3,665,113	0.0410	0.0730	9	32	50	61-70
0.0256	3,662,775	0.0112	0.0290	10	39	59	101-110
0.0131	3,661,724	0.0338	0.0220	11	44	66	151-160
0.00671	3,661,052	0.0158	0.0156	12	51	75	221-230
0.00455	3,660,830	0.0842	0.0106	13	58	84	271-280
0.00311	3,660,686	0.0443	0.0111	14	65	93	331-340
0.00263	3,660,621	0.0197	0.00907	15	74	104	361-370
0.00207	3,660,568	0.0274	0.00703	16	79	111	401-410
0.00161	3,660,517	0.0260	0.00560	17	85	119	451-460
0.00126	3,660,480	0.0166	0.00472	18	91	127	511-520
9.96E-4	3,660,451	0.0135	0.00351	19	97	135	561-570
7.70E-4	3,660,421	0.00413	0.00201	20	105	145	641-650
5.43E-4	3,660,397	0.0116	0.00108	21	109	151	741-750
3.59E-4	3,660,375	0.00911	5.23E-4	22	115	159	881-890
2.08E-4	3,660,356	0.00472	4.57E-4	23	121	167	> 1000
1.07E-4	3,660,344	0.00966	1.33E-4	24	127	175	> 1000

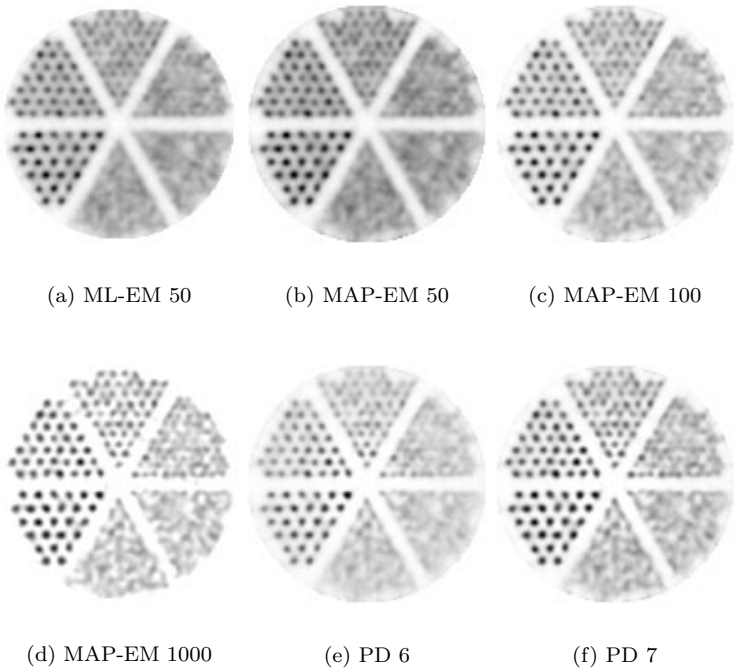
Table 4: Subproblem progress of primal-dual reconstruction of Derenzo phantom, where $\rho = 2$ for convergence to accurate solution.

μ	f	$\|\nabla\ell\|$	$\max(\theta_i\lambda_i)$	*niter*	*ncg*	*ngr*	MAP-EM
10.0	4,765,608	1.66	———	7	15	29	1-10
1.09	3,742,471	0.535	0.254	15	44	74	1-10
0.109	3,669,773	0.109	0.109	21	82	125	41-50
0.0109	3,661,400	0.0114	0.0115	26	118	172	171-180
0.00109	3,660,418	0.00226	0.0228	34	174	245	660-670
1.09E-4	3,660,344	0.00125	0.0105	41	214	300	> 1000

Table 5: Subproblem progress of barrier reconstruction of Derenzo phantom.

μ	f	$\|\nabla\ell\|$	$\max(\lambda_i\theta_i)$	*niter*	*ncg*	*ngr*	MAP-EM
10.862	4,029,471	3.68	13.2	1	2	4	1-10
0.1018	3,778,157	1.08	5.48	2	4	8	1-10
0.0388	3,738,259	0.669	5.61	3	6	12	1-10
0.0314	3,698,611	0.321	3.44	4	9	17	1-10
0.0163	3,678,343	0.211	1.56	5	12	22	21-30
0.00684	3,667,297	0.107	0.906	6	17	29	51-60
0.00340	3,663,134	0.0737	0.296	7	23	37	91-100
0.00114	3,661,518	0.0527	0.246	8	30	46	161-170
8.53E-4	3,660,855	0.0655	0.222	9	35	53	261-270

Table 6: Subproblem progress of primal-dual method with $\rho = 50$ for rapid early progress; Derenzo phantom.



(a) ML-EM 50 (b) MAP-EM 50 (c) MAP-EM 100

(d) MAP-EM 1000 (e) PD 6 (f) PD 7

Figure 2: Central transverse section of Derenzo phantom: (e) and (f) are the solutions at the 6th and 7th subproblem, respectively, of the primal-dual method with $\rho = 50$.

(a) transverse (b) coronal (c) sagittal

Figure 3: Cross sections of the FDG2 thin-slice study.

listed per subproblem in Table 7. The objective function value is not available in accordance with the discussion of Section 4.3. As such, we did not perform a MAP-EM reconstruction for comparison. Recall that in the sparse implementation, every iteration of the line search requires one back projection to compute the directional derivative plus one Hessian-vector product to compute the directional second derivative. Thus we must modify the computational cost formula (extrapolated case) to include the cost of the line searches:

$$ngr = npr + 2 \cdot nit + ncg + 1.5 \cdot nls,$$

where nls represents the number of line search iterations. This reconstruction converged in remarkably few conjugate gradient iterations, in part due to the relative strength of the prior term. Although $\rho = 2$ was used in this reconstruction, some of the subproblems involve a decrease in μ of greater than 2. This is an effect of extrapolation as such behavior is rare in the absence of extrapolation. Another effect of extrapolation in primal-dual reconstructions is the occasional non-monotonicity in the decrease of $\|\nabla \ell\|$ and $\max(\lambda_i \theta_i)$, especially in the early subproblems.

6 Concluding Remarks

In this paper we present a new class of algorithms to the already crowded field of emission tomography reconstruction algorithms. Our algorithms draw heavily from long established principles as well as recent developments in linear and nonlinear programming [60]. Unlike most of the EM method or the methods that have been proposed to date to improve the convergence of EM, the algorithms presented in this paper treat the reconstruction problem as a constrained optimization problem. These interior point methods directly maximize $f_{ML}(\theta)$ or $f_{MAP}(\theta)$ while maintaining feasibility by approximately solving a sequence of subproblems. The solution to each of these subproblems satisfies a perturbed version of the optimality conditions, where in each subproblem the duality gap decreases until the optimal solution is found.

μ	$\|\nabla\ell\|$	$\max(\lambda_i\theta_i)$	*npr*	*niter*	*ncg*	*nls*	*ngr*
2.08	1.68	3.50	1	1	2	1	7
1.64	0.225	5.24	2	2	4	4	16
1.30	0.137	1.39	3	3	6	7	26
0.654	10.3	20.1	4	4	9	11	38
0.271	2.30	15.2	5	5	11	13	46
0.172	0.428	1.48	6	6	13	15	54
0.0877	0.178	3.16	7	7	15	17	62
0.0432	0.0584	0.0801	8	8	19	21	75
0.0184	0.0829	1.50	9	9	22	23	84
0.0117	0.0206	0.277	10	10	28	24	94
0.00638	0.0208	0.319	11	11	34	26	106
0.00263	0.00996	0.489	12	12	40	27	117
0.00196	0.00538	0.0236	13	13	46	28	127
8.54E-4	0.00418	0.0106	14	14	52	29	138
3.54E-4	0.00302	0.00297	15	15	57	31	149
1.90E-4	0.00196	0.00105	16	16	65	32	161
1.29E-4	7.84E-4	1.29E-4	17	17	72	35	176

Table 7: Subproblem progress of thin-slice primal-dual reconstruction of FDG dataset; $\rho = 2$. Extrapolation was performed on subproblem solutions.

Interior-point methods have strong theoretical properties, as do search directions that approximate the Newton direction. The results presented in this paper indicate that interior-point methods converge to the regularized maximum likelihood solution at a much accelerated rate over the EM method for MAP reconstruction. In particular, the primal-dual results are especially favorable. This is consistent with the success of primal-dual techniques in linear programming. Despite having to update twice as many variables, the primal-dual method is actually quite economical. The process of actively estimating the dual variables produces more favorable primal directions and more accurate placement on the primal-dual trajectory than can be achieved by a primal-only method. Another advantage of interior-point methods is that they permit the use of well-defined termination measures based on the KKT optimality conditions. Being a "short-step" algorithm, the primal-dual method gives the user a great deal of flexibility in selecting the desired accuracy of the solution.

The interior point approach is quite different than the EM approach, which defines a complete data space and iteratively maximizes a sequence of separable functions that are defined over the complete data space. However, viewed from an optimization perspective, the EM algorithm is a scaled steepest ascent method with unit step size. The same is true of the generalized EM algorithm for MAP reconstruction that was discussed in Section 2.2. As such, the observed slow convergence of EM methods is not surprising.

Elements of the interior-point methods that we present in this paper have already been proposed in the context of the PET reconstruction problem. Kauf-

man proposed a truncated-Newton approach in conjunction with a bent line search to impose the non-negativity constraint. Mumcuoglu [52] introduced a quadratic penalty term into the objective function to penalize violations of the non-negativity constraint and maximized the resulting unconstrained subproblems with a nonlinear conjugate gradient . The quadratic penalty has the disadvantage of having a discontinuous second derivative; moreover the penalty parameter was kept constant and did not decrease. The nonlinear conjugate gradient method usually offers only incremental improvement in convergence rate over steepest ascent [60]. This is consistent with Mumcuoglu's claims of only modest improvement in convergence. Both Kaufman and Mumcuoglu have contributed significantly to the understanding of the properties of ML and MAP objective functions.

The improved convergence properties of coordinate ascent and grouped coordinate ascent methods have been attributed to suppression of high frequency error [62] and reduced Fisher information of hidden (rather than complete) data spaces [16, 17, 18], respectively. Each grouped coordinate ascent update is a projected steepest ascent direction of the approximating separable quadratic, scaled by the inverse diagonal Hessian of the approximating function. While the results presented in these papers are quite strong, Gauss-Seidel methods are known to be quite sensitive to the relaxation parameter, and have not performed well in many applications. Moreover, the grouped coordinate ascent methods require a parallelization strategy that decomposes the computation in image space. Such a strategy nearly precludes the distribution of the projection-space vectors across the processors. For very large 3-D reconstructions such as the thin-slice reconstructions, these vectors must be partitioned across the processors. While the coordinates can still be updated in parallel groups if the projection space vectors are distributed across the processors, such a strategy can lead to inter-processor communication bottlenecks.

The parallel implementation of the interior point methods presented in this paper can be readily scaled to the largest reconstruction problems. For large but sparse thin-slice problems, most of the columns of the C matrix are never loaded or visited. Ordering the computation primarily in projection space in accordance with the pseudocode of Section 3 permits the processors to simply ignore the columns C_j that correspond to the sparse subspace of y. In contrast, grouped coordinate ascent requires that the forward and back projections be ordered primarily in image space. As such, the $C_{i,j}$ must either be loaded as individual elements (thereby incurring a great deal of I/O latency put permitting the sparse subspace to be ignored) or loaded in entire rows (thereby incurring a great deal of I/O traffic). As such we assert that the projection-space parallel decomposition scheme, which lends itself readily to the interior-point methods, is preferable for very large reconstructions.

This paper illustrates the application of interior-point methodology to image reconstruction problems by selecting the regularized ML objective function. The emphasis of this paper, however, is not in the choice of objective function but rather on the versatility of interior-point methodology. Indeed, the logarithmic barrier and primal-dual methods could readily be adapted to other

constrained objective functions, such as weighted least squares with lower-bound constraints. Much of the computational structure of computing the gradient, etc., applies to that objective function as well [40]. The truncated-Newton method for solving an unconstrained minimization problem can readily be applied to finding the solution of the weighted least squares objective function without lower bound constraints. Interior-point methodology permits modifications to the objective function (including the regularizing term) to be made without requiring a new algorithmic or computational framework.

7 Acknowledgments

Ariela Sofer is partly supported by National Science Foundation grant DMI-9414355. The computational work in this paper was performed on the IBM RS/6000 SP2 parallel computer at the Division of Computer Research and Technology, National Institutes of Health. We are grateful to Jurgen Seidel of the Department of Nuclear Medicine, National Institutes of Health, for kindly providing us with the small animal data and Monte Carlo simulation data.

References

[1] H.H. Barrett, D.W. Wilson, and B.W. Tsui, "Noise properties of the EM algorithm: I. Theory," *Phys. Med. Biol.*, vol. 39 (1994), pp. 833-846.

[2] C. Bouman and K. Sauer, "A unified approach to statistical tomography using coordinate descent optimization," *IEEE Trans. Image Processing*, vol. 5 (1996), pp. 480-492.

[3] J. Browne and A.R. DePierro, "A row-action alternative to the EM algorithm for maximizing likelihoods in emission tomography," *IEEE Trans. Med. Imag.*, vol. 15, no. 4 (1996), pp. 687-699.

[4] C.L. Byrne, "Block-iterative methods for image reconstruction from projections," *IEEE Trans. Imag. Proc.,* vol. 5, no. 5 (1996), pp. 792-794.

[5] Y. Censor, "Row-action methods for huge and sparse systems and their applications," *SIAM Review,* vol. 23, no. 4 (1981), pp. 444-466.

[6] Y. Censor and G. Herman, "On some optimization techniques in image reconstruction from projections," *App. Numer. Math.*, vol. 3 (1987), pp. 365-391.

[7] R.H. Chan and M.K. Ng, "Conjugate gradient methods for Toeplitz systems", *SIAM Review*, vol. 38, no. 3 (1996), pp. 427-482.

[8] Chen, C.M., and Lee, S.Y., "On parallelizing the EM algorithm for PET image reconstruction," *IEEE Trans. Parallel Dist. Sys.*, vol. 5 (1994), pp. 860-873.

[9] G. Chinn and S-C. Huang, "A general class of preconditioners for statistical iterative reconstruction of emission computed tomography," *IEEE Trans. Med. Imag.,* vol. 16, no. 1 (1997), pp. 1-10.

[10] N.H. Clinthorne, T-S. Pan, P-C. Chiao, W.L. Rogers, and J.A. Stamos, "Preconditioning methods for improved convergence rates in iterative reconstructions", *IEEE Trans. Med. Imag.,* vol. 12, no. 2 (1993), pp. 78-83.

[11] A.R. Conn, N. Gould, and Ph.L. Toint, "A primal-dual algorithm for minimizing a nonconvex function subject to bound and linear equality constraints," *IBM T.J. Watson Research Center Technical Report 96/9* (1996).

[12] R.S. Dembo and T. Steihaug, "Truncated-Newton algorithms for large-scale unconstrainted optimization," *Math Programming.*, vol 26 (1983), pp. 190-212.

[13] A.P. Dempster, N.M. Laird, and D.B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *JRSS,* vol. 39 (1977), pp. 1-38.

[14] A.R. DePierro, "A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography," *IEEE Trans Med. Imag.*, vol. 14, no. 1, (1995), pp. 132-137.

[15] A.R. DePierro, "On the convergence of an EM-type algorithm for penalized likelihood estimation in emission tomography," *IEEE Trans. Med. Imag.*, vol. 14, no. 4 (1995), pp. 762-765.

[16] J.A. Fessler, "Penalized maximum-likelihood image reconstruction using space-alternating generalized EM algorithms,". IEEE Trans. Imag. Proc., vol. 4, no. 10 (1995), pp. 1417-1429.

[17] J.A. Fessler, "Hybrid Poisson/polynomial objective functions for tomographic image reconstruction from transmission scans," *IEEE Trans. Imag. Proc.,* vol. 4, no. 10 (1995), pp. 1439-1450.

[18] J.A. Fessler, E.P. Ficaro, N.H. Clinthorne, and K. Lange, "Grouped-coordinate ascent algorithms for penalized-likelihood transmission image reconstruction," *IEEE Trans. Med. Imag.,* vol. 16, no. 2 (1997), pp. 166-175.

[19] A.V. Fiacco and G.P. McCormick, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques,* John Wiley and Sons, New York (1968).

[20] A. Forsgen and P.E. Gill, "Primal-dual interior methods for nonconvex nonlinear programming," Report NA 96-3, Department of Mathematics, University of California, San Diego (1996).

[21] P.E. Gill, W. Murray, D.B. Poncelon, and M.A. Saunders, "Primal-dual methods for linear programming," *Math. Programming*, vol. 70 (1995), pp. 251-277.

[22] D.R. Gilland, R.J. Jaszczak, K.L. Greer, and R.E. Coleman, "Quantitative SPECT reconstruction of iodine-123 data," *J. Nucl. Med.*, vol. 32 (1991), pp. 527-533.

[23] G. Golub and C. Van Loan, *Matrix Computations*, second edition, Johns Hopkins University Press, Baltimore, (1989).

[24] P.J. Green, "On use of the EM algorithm for penalized likelihood estimation," *JRSS* B, vol. 52, no. 3 (1990), pp. 443-452.

[25] P.J. Green, "Bayesian reconstructions from emission tomography data using a modified EM algorithm," *IEEE. Trans. Med. Imag.*, no. 9, vol. 1 (1990), pp. 84-93.

[26] T. Hebert, R. Leahy, and M. Singh, "Fast MLE for SPECT using an intermediate polar representation and a stopping criterion," *IEEE Trans. Nucl. Sci.*, vol. 35, no. 1 (1988), pp. 615-619.

[27] T. Hebert and R. Leahy, "A generalized EM algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs priors," *IEEE Trans. Med. Imag.,* vol. 8, no. 2 (1989), pp. 194-202.

[28] G.T. Herman, *Image Reconstruction from Projections: the Fundamentals of Computerized Tomography,* Academic Press, New York, 1980.

[29] G.T. Herman, D. Odhner, K.D. Toennies, and S.A. Zenios, "A parallelized algorithm for image reconstruction from noisy projections," in *Large Scale Numerical Optimization, Proceedings in Applied Mathematics*, T. Coleman and Y. Li, Eds. Siam: SIAM Publications, 1990, vol. 46, pp. 3-21.

[30] G.T. Herman and D. Odhner, "Performance evaluation of an iterative image reconstruction algorithm for Poisson emission tomography," *IEEE Trans. Med. Imag.*, vol. 10, no. 3 (1991), pp. 336-346.

[31] G.T. Herman and L.B. Meyer, "Algebraic reconstruction techniques can be made computationally efficient," IEEE Trans. Med. Imag., vol. 12, no. 3 (1993), pp. 600-609.

[32] M.R. Hestenes and E. Steifel, "Methods of conjugate gradients for solving linear systems," *J. Research of the National Bureau of Standards*," vol. 49, no. 6 (1952), pp. 409-435.

[33] H.M. Hudson and R.S. Larkin, "Accelerated image reconstruction using ordered subsets of projection data," *IEEE Trans. Med. Imag.*, vol. 13, no. 4 (1994), pp. 601-609.

[34] C.A. Johnson, Y. Yan, R.E. Carson, R.L. Martino, and M.E. Daube-Witherspoon, "A system for the 3D reconstruction of retracted-septa PET data using the EM algorithm," *IEEE Trans. Nucl. Sci.*, vol. 42, no. 4 (1995), pp. 1223-1227.

[35] C.A. Johnson, J. Seidel, R.E. Carson, W.R. Gandler, A. Sofer, M.V. Green, and M.E. Daube-Witherspoon, "Evaluation of 3D reconstruction algorithms for a small animal PET camera," to appear in *IEEE Trans Nucl. Sci.*, vol. 44., no. 3 (1997).

[36] C. Kamphuis, F.J. Beekman, and M.A. Viergever, "Evaluation of OS-EM vs. ML-EM for 1D, 2D, and fully 3D SPECT reconstruction," *IEEE Trans Nucl. Sci.* vol. 43, no. 3 (1996), pp. 2018-2024.

[37] N. Karmarkar, "A new polynomial-time algorithm for linear programming," *Combinatorica*, vol. 4, (1984), pp.49-57.

[38] L. Kaufman, "Implementing and accelerating the EM algorithm for positron emission tomography," *IEEE Trans. Med. Imag.*, vol. 6, no. 1 (1987), pp. 37-51.

[39] L. Kaufman, "Solving emission tomography problems on vector machines," *Ann. Oper. Res.*, vol. 22 (1990), pp. 325-353.

[40] L. Kaufman, "Maximum likelihood, least squares, and penalized least squares for PET," *IEEE Trans. Med. Imag.*, vol. 12, no. 2 (1993), pp. 200-214.

[41] M. Kojima, S. Mizuno, and A. Yoshise, "A primal-dual interior point algorithm for linear programming," in *Progress in Mathematical Programming: Interior-Point and Related Methods*, N, Megiddo (editor), Springer Verlag, New York, pp. 29-47 (1989).

[42] D.S. Lalush and B.M.W. Tsui, "A fast and stable maximum a posteriori conjugate gradient reconstruction algorithm," *Med. Phys.*, vol 22, no. 8 (1995), pp. 1273-1284.

[43] D.S. Lalush and B.M.W. Tsui, "The importance of preconditioners in fast Poisson-based iterative reconstruction algorithms for SPECT," *1995 IEEE Nuclear Science Symposium Conference Record* (1996), pp. 1326-1330.

[44] J. Liow and S. Strother, "Practical tradeoffs between noise, quantitation and number of iterations for maximum likelihood-based reconstructions," *IEEE Trans. Med. Imag.*, vol. 13 (1994), pp. 601-609.

[45] A.A. Lammertsma, S.P. Hume, R. Myers, S. Ashworth, P.M. Bloomfield, S. Rajeswaran, T. Spinks, and T. Jones, "PET scanners for small animals," *J. Nucl. Med.*, vol. 36 (1995), pp. 2391-2392.

[46] K. Lange and R. Carson, "EM Reconstruction Algorithms for Emission and Transmission Tomography," *J. Comp. Assist. Tomogr.*, vol. 8, no. 2 (1984), pp. 306-316.

[47] K. Lange, M. Bahn, and R. Little, "A theoretical study of some maximum likelihood algorithms for emission and transmission tomography," *IEEE Trans. Med. Imag.*, vol. 6, no. 2, (1987), pp. 106-114.

[48] K. Lange, "Convergence of EM image reconstruction algorithms with Gibbs smoothing," *IEEE Trans. Med. Imag.*, vol. 9, no. 4 (1990), pp. 439-446.

[49] E. Levitan and G.T. Herman, "A maximum a posteriori probability expectation maximization algorithm for image reconstruction in emission tomography," *IEEE Trans. Med. Imag.*, vol. 6, no. 3, (1987), pp. 185-192.

[50] J. Llacer and E. Veklerov, "Feasible images and practical stopping rules for iterative algorithms in emission tomography," *IEEE Trans. Med. Imag.*, vol. 186, no. 2 (1989), pp. 186-193.

[51] T. Miller and J. Wallis, "Clinically important characteristics of maximum-likelihood reconstruction," *J. Nucl. Med.*, vol. 33, no. 9 (1992), pp. 1678-1684.

[52] E.U. Mumcuoglu, R.M. Leahy, S.R. Cherry, Z. Zhou, "Fast gradient-based methods for Bayesian reconstruction of transmission and emission PET Images," *IEEE Trans. Med. Imag.*, vol. 13, no. 4 (1994), pp. 687-701.

[53] E.U. Mumcuoglu, R.M. Leahy, and S.R. Cherry, "Bayesian reconstruction of PET images: methodology and performance analysis," *Phys. Med. Biol.,* vol. 41 (1996), 1777-1807.

[54] W. Murray, "Analytic expressions for the eigenvalues and eigenvectors of the Hessian matrices of barrier and penalty functions," *J. Optimization Theory App.*, vol. 7 (1971), pp. 189-196.

[55] S.G. Nash, "Preconditioning of truncated-Newton methods," *SIAM J. Sci. Stat. Comput.*, vol. 6, no. 3 (1985), pp. 599-616.

[56] S.G. Nash and S. Sofer, "Block truncated-Newton methods for parallel optimization," *Math. Programming.*, vol. 45 (1989), pp. 529-546.

[57] S.G. Nash and A. Sofer, "Assessing a search direction within a truncated-Newton method," *Oper. Res. Let.*, vol 9 (1990), pp. 219-221.

[58] S.G. Nash and A. Sofer, "A general-purpose parallel algorithm for unconstrained optimization," *SIAM J. Opt.*, vol. 1, no. 4 (1991), pp. 530-547.

[59] S.G. Nash and A. Sofer, "A barrier method for large-scale constrained optimizaton," *ORSA J. Comp.*, vol. 5, no. 1 (1993), pp. 40-53.

[60] S.G. Nash and A. Sofer, *Linear and Nonlinear Programming*, McGraw-Hill, New York (1996).

[61] J.M. Ollinger and J.A. Fessler, "Positron-Emission Tomography," IEEE Signal Processing Magazine, vol. 14, no. 1 (1997), pp. 43-55.

[62] K. Sauer and C. Bouman, "A local update strategy for iterative reconstruction from projections," *IEEE Trans. Signal Processing*, vol. 41, no, 2 (1996), pp. 534-548.

[63] J. Seidel, W.R. Gandler, and M.V. Green, "A very high resolution single-slice animal PET scanner based on direct detection of coincidence end-points," *J. Nucl. Med.*, vol. 35 (1994), p. 40P.

[64] J. Seidel, W.R. Gandler, and M.V. Green, "Characteristics of a pair of small field-of-view LSO scintillation cameras," *IEEE Trans Nucl. Sci*, vol. 43 (1996), pp. 1968-1973.

[65] J. Seidel, C. Johnson, E. Jagoda, L. Sha, W.C. Eckelman, and M.V. Green, "A high resolution scanner for investigating technical requirements and selected biological applications of continuous 3D detectors in smll animal PET, to appear in *J. Nucl. Med.*

[66] L.A. Shepp and Y. Vardi, "Maximum Likelihood Reconstruction for Emission Tomography," *IEEE Trans Med. Imag.,* vol 1, no. 2 (1982), pp. 113-122.

[67] D.L. Snyder, M.I. Miller, L.J. Thomas, Jr., and D.G. Politte, "Noise and edge artifacts in maximum-likelihood reconstructions for emission tomography," *IEEE Trans. Med. Imag.,* vol MI-6, no. 3 (1987), pp. 228-238.

[68] G.D. Tourassi, C.E. Gloyd, Jr., M.T. Munley, J.E. Bowsher, and R.E. Coleman, "Improved lesion detection in SPECT using MLEM reconstruction," *IEEE Trans. Nucl. Sci.*, vol. 38 (1991), pp. 780-783.

[69] Y. Vardi, L.A. Shepp, and L. Kaufman, "A statistical model for positron emission tomography," *J. Amer. Stat. Assoc.*, vol. 80, no. 389 (1985), pp. 8-37.

[70] E. Veklerov and J. Llacer, "Stopping rule for the MLE algorithm based on statistical hypothesis testing," *IEEE Trans. Med. Imag.*, vol. MI-6, no. 4 (1987), pp. 313-319.

[71] D.W. Wilson and B.M.W. Tsui, "Noise properties of filtered-backprojection and ML-EM reconstructed emission tomography images," *IEEE Trans. Nucl. Sci.*, vol. 40, no. 4 (1993), pp. 1198-1203.

[72] G.L. Zeng, G.T. Gullberg, B.M.W. Tsui, and J.A. Terry, "Three-dimensional iterative reconstruction algorithms with attenuation and geometric point response corrections," *IEEE Trans. Nucl. Sci.*, vol. 38 (1991), pp. 693-702.