Iterative reconstruction techniques in emission computed tomography

2006 Phys. Med. Biol. 51 R541

(http://iopscience.iop.org/0031-9155/51/15/R01)

INSTITUTE OF PHYSICS PUBLISHING PHYSICS IN MEDICINE AND BIOLOGY

Phys. Med. Biol. **51** (2006) R541–R578 doi:10.1088/0031-9155/51/15/R01

TOPICAL REVIEW

Iterative reconstruction techniques in emission computed tomography

Jinyi Qi[1] **and Richard M Leahy**[2]

[1] Department of Biomedical Engineering, University of California, Davis, CA 95616, USA
[2] Signal and Image Processing Institute, Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089-2564, USA

Received 12 January 2006, in final form 26 April 2006
Published 12 July 2006
Online at stacks.iop.org/PMB/51/R541

Abstract

In emission tomography statistically based iterative methods can improve image quality relative to analytic image reconstruction through more accurate physical and statistical modelling of high-energy photon production and detection processes. Continued exponential improvements in computing power, coupled with the development of fast algorithms, have made routine use of iterative techniques practical, resulting in their increasing popularity in both clinical and research environments. Here we review recent progress in developing statistically based iterative techniques for emission computed tomography. We describe the different formulations of the emission image reconstruction problem and their properties. We then describe the numerical algorithms that are used for optimizing these functions and illustrate their behaviour using small scale simulations.

1. Introduction

Emission computed tomography (ECT), encompassing positron emission tomography (PET) and single photon emission computed tomography (SPECT), produces noninvasive measures of a broad range of physiological processes through three-dimensional (3D) imaging of the spatial distribution of radiolabelled biological probes or tracers. The two primary factors limiting image quality are detector resolution, which determines the maximum resolution of the reconstructed images, and photon count density, which determines the minimum noise level that can be achieved at the maximum resolution. To maximize resolution in a particular system, it is necessary to accurately model the photon detection process. In practice, system resolution is often sacrificed to reduce noise in the images. To achieve an optimal trade-off between resolution and noise propagation it is necessary to accurately model the noise processes in the data. The models for the data and noise are determined, either explicitly or implicitly, by the choice of reconstruction algorithm. The choice of reconstruction algorithm therefore plays a crucial role in the overall performance of any ECT system.

In this paper, we will review the recent literature on iterative approaches to PET and SPECT image reconstruction. Our viewpoint is that consideration of the factors limiting noise and resolution leads naturally to the choice of a model-based approach in which the physical and statistical properties of the system are explicitly included. These formulations in turn lead to numerical problems involving large sets of possibly nonlinear equations, which can generally only be solved using iterative methods. We will assume a basic familiarity both with the physical principles underlying PET and SPECT imaging and with the instrumentation used for data acquisition. Excellent reviews of these topics can be found in Cherry *et al* (2004), Wernick and Aarsvold (2004), Cherry and Phelps (1996), Bendriem and Townsend (1998), Rosenthal *et al* (1995).

A huge number of iterative methods have been described for ECT, with a corresponding profusion of acronyms. Here we attempt to describe the commonalities and differences between these methods. In doing this, we keep in mind the following three essential ways in which the algorithms can differ: (i) the choice of cost function—this includes the model for the data, the function of the data to be optimized and, where appropriate, the choice of regularizing or smoothing function on the solution; (ii) the optimization procedure—the numerical algorithm chosen to maximize or minimize the cost function; (iii) computation cost—determined by the convergence rate, which determines the number of iterations required, and the computation cost per iteration.

It is often the choice of cost function that determines the final solution; the route taken to this solution by different algorithms optimizing the same function should not affect the result. Therefore it is important to differentiate between methods optimizing different criteria and methods that take different approaches to minimizing a particular function. However, this situation is complicated in practice by the fact that it is common to terminate many algorithms before convergence, in which case the solution will always depend on the algorithm. For the case of nonconvex smoothing functions, for which multiple local optima may exist, the choice of search algorithm may also affect the final solution since two methods may converge to different local optima. These two issues, selection of an appropriate cost function and the numerical algorithm chosen to optimize this function, are the primary focus of this review.

We organize the paper as follows. We begin with the motivation for using iterative reconstruction methods rather than direct methods based on the line-integral model. In section 3 we give an overview of the various approaches to iterative ECT. We then return and provide a more detailed description of the cost functions typically used in iterative PET and SPECT image reconstruction. We conclude with a detailed review of the different numerical approaches to optimizing these functions.

2. Why use iterative reconstruction methods?

Detection of coincident photon pairs in PET leads naturally to a line-integral model for the data, i.e. the number of coincident photon pairs detected between any pair of detectors is approximately linearly proportional to the integral of the tracer density along a line joining the two detectors. Similarly, data from SPECT systems using parallel, fan beam and cone beam collimators can also be modelled as sets of line integrals of the tracer density along the collimation directions. Consequently, PET and SPECT images can be reconstructed using analytic inversion methods that are based on the relationship between a function and its line integrals.

For two-dimensional (2D) PET systems with septa and SPECT with parallel or fan beam collimators, 2D filtered backprojection (FBP) algorithms are routinely used for reconstruction. For 3D PET systems without septa and cone beam collimated SPECT, the reconstruction

problems are inherently 3D and not separable into a set of independent 2D problems. For cone beam data, tremendous progress has been made in the last decade on analytic inversion methods for a range of sampling geometries as determined by the collimator design and the trajectory of the gamma camera about the patient. For example, exact FBP algorithms have now been developed for helical geometries (Katsevich 2002). The analytic 3D PET reconstruction problem has been solved for some time (Colsher 1980) but is complicated in practice by truncation of the oblique 2D x-ray projections of the source distribution as a result of the finite axial extent of the scanner. This can be solved using a 3D filtered backprojection method in conjunction with a reprojection computation to fill in the missing data (Kinahan and Rogers 1989). An alternative approach to 3D FBP reconstruction is to rebin the data into an equivalent set of stacked 2D sinograms, one per transaxial plane, and reconstruct the images using 2D FBP. These rebinning algorithms range from approximate single-slice (Daube-Witherspoon and Muehllehner 1987) and Fourier rebinning to more computationally expensive but exact Fourier space rebinning methods (Defrise *et al* 1997).

While the analytic approaches typically result in fast reconstruction algorithms, accuracy of the reconstructed images is limited by the approximations implicit in the line-integral model on which the reconstruction formulae are based. In PET the line-integral model is reasonably accurate since depth-dependent variations in detector resolution are relatively small and the effects of attenuation are well modelled as a single multiplicative scaling factor that can be applied to each line integral. Even so, resolution is affected by the combined effects of positron range, noncollinearity of the photon pair, variations in detector-pair sensitivity along the LOR, and the spatially variant response of the detector caused by intercrystal penetration and scatter. In SPECT, line integrals are a far less accurate model. The collimators produce pronounced depth dependence in resolution that is not included in the line-integral model. The single photon detection process results in attenuation that can differ between each voxel and each detector element. In routine SPECT reconstruction protocols, the first of these factors is often ignored while attenuation is often corrected approximately using a post-processing step (Chang 1978).

Methods to compensate for many of these factors that cause data to deviate from the line integral have been applied within the framework of direct image reconstruction. Examples in PET include correction for positron range (Derenzo 1986) and detector response (Huesman *et al* 1989, Liang 1994). Similarly, analytic methods have recently been developed for compensation for depth-dependent resolution (Edholm *et al* 1986, Xia *et al* 1995) and for attenuation (Liang *et al* 1994, Tang *et al* 2005, Huang *et al* 2005) in SPECT. These require analytic models for the factors that cause deviation from the line-integral model. They are often computationally demanding, and are typically not optimized for use with photon-limited data.

In contrast to the analytic methods, an iterative model-based approach can simply model the mapping from source to detector without requiring explicit line-integral or other models. The simplest iterative methods use on-the-fly computation of forward- and backprojection with the same linear-interpolation-based method as is commonly used in FBP. A more flexible approach, which allows more accurate modelling of the physical detection process, is to precompute and store the projection matrices that define this mapping. This can lead to huge computation demands, but methods to reduce costs using sparse matrix structures and efficient factorizations are making this approach increasingly practical. The ability of the model-based approaches to accurately model photon detection leads to improvements in attainable resolution given a sufficient number of counts. A combination of analytic, geometric and Monte Carlo modelling can produce projections of arbitrarily high accuracy (Floyd and Jaszczak 1985, Veklerov and Llacer 1988, Qi *et al* 1998b, Huesman *et al* 2000, Laurette *et al* 2000, Beekman

et al 2002). Alternatively, the projection matrix can be measured directly (Formiconi *et al* 1989, Panin *et al* 2004, Chen *et al* 2004).

Alternative SPECT collimation schemes, such as coded aperture (Paxman *et al* 1985, Koral and Rogers 1979) and electronic collimation (Hebert *et al* 1990), produce measurements that are not simple line integrals and present special challenges for analytic inversion. In contrast, the iterative approaches are essentially identical for all collection geometries differing only in the specifics of the projection matrices.

A second limitation of the analytic approaches is that they do not take account of the statistical variability inherent in photon-limited detection. The resulting noise in the reconstructions is usually controlled, at the expense of resolution, by varying the cut-off frequency of a linear filter applied to the sinogram. Since the noise is signal dependent, this type of filtering is not particularly effective at achieving an optimal bias-variance trade-off. More complex methods, such as Wiener filtering, attempt to achieve an improved trade-off, but even these methods are based on assumptions, such as wide sense stationarity of the noise, that are inconsistent with the Poisson nature of the photon counting process in PET and SPECT. Conversely, a model-based method can explicitly model the statistical variations in the photon detection process through a probabilistic model for the data. If prior information regarding image smoothness or structure is also available then further improvements can be realized through the use of Bayesian estimation schemes.

The combination of improved modelling of the detection process and improved handling of statistical noise when using statistically based methods offers the possibility for enhanced performance of PET and SPECT with both high count data (where model accuracy limits resolution) and low count data (where statistical noise limits resolution). In its simplest form the imaging problem can be cast as one of parameter estimation, where the data are Poisson random variables with mean equal to a linear transformation of the parameters. This formulation is complicated, as we will describe below, by the impact of additional noise and correction terms.

3. An overview of iterative reconstruction methods

The model-based iterative approaches to PET and SPECT reconstruction are based on a linear model that relates the measured projection data $\boldsymbol{y}$ to the unknown source distribution $\boldsymbol{f}$ through a projection matrix $\boldsymbol{P}$. This matrix models the probability of an emission from each voxel element in the source being detected at each sinogram element. The problem is complicated by the presence of scatter, background radiation or randoms, and by the stochastic nature of the photon emission and detection process. The reconstruction problem therefore effectively reduces to approximately solving the set of equations $\boldsymbol{y} = \boldsymbol{P}\boldsymbol{f}$, taking into account the effects of noise and background.

The earliest iterative tomographic methods are the large class of algebraic reconstruction techniques or ART methods (Gordon *et al* 1970). These are a class of numerical procedures for solving large sparse systems of equations. In the event that the data are consistent, i.e., an exact solution to the equations exists, then the methods will converge to such a solution. In the more common case where the data are inconsistent, these methods will typically not converge but instead enter a limit cycle. There are several excellent reviews and tutorials on this class of method (Censor 1983, De Pierro 1990, Censor and Zenios 1997).

Lacking the formulation of the ART methods is a direct consideration of noise in the data, which in the case of uncorrected PET and SPECT data is well modelled as a Poisson process. The earliest attempt to explicitly include a Poisson model in tomographic reconstruction was the maximum-likelihood (ML) method developed by Rockmore and Macovski (1976).

However, it was the work by Shepp and Vardi (1982) on the EM (expectation maximization) algorithm and Lange and Carson (1984) that marked the beginning of a very fruitful period in which a large number of variations on this approach were developed, leading to the now widespread use of statistically based iterative reconstruction methods in clinical and small animal PET and SPECT systems.

The EM algorithm is a general approach to solving ML problems through the introduction of a set of 'complete data' which, if observed, would make the ML problem readily solvable (Dempster *et al* 1977). The algorithm then iterates between computing the mean of the complete data, given the observed data and the current estimate of the image, and maximizing the probability of the complete data over the image space. The choice of complete data in the EM algorithm for the linear Poisson problem leads to an elegant closed form update equation reminiscent of the multiplicative ART algorithms. The EM algorithm has attractive properties including guaranteed non-negativity, convergence and 'count preservation' at each iteration. However it also converges very slowly.

Much of the work in statistically based algorithms over the past 20 years has concentrated on finding alternatives to the EM algorithm that converge faster. In the ordered subsets EM (OSEM) method (Hudson and Larkin 1994), the data are divided into a number of disjoint sets, and the EM algorithm applied sequentially to each of these data sets in turn. This produces remarkable improvements in the initial convergence rate compared to EM. This property, coupled with the relatively simple form of the update equation, has led to the widespread adoption of OSEM as a standard method for iterative reconstruction in PET and SPECT. While the method is practically attractive, it is not a true ML estimator and will not converge in general. There are many variations on this method that we describe in section 6.3. Of particular interest among the variations on OSEM are the convergent OS algorithms (Browne and De Pierro 1996, Ahn and Fessler 2003, Hsiao *et al* 2004). These methods are closely related to the general class of incremental gradient methods in which the cost function is decomposed into a sum of terms and the image updated with respect to the gradient of only one term at each iteration (Bertsekas 1997). Convergence can be guaranteed through an appropriate choice of relaxation parameter.

An alternative approach to ML estimation is to use standard gradient-based optimization procedures such as steepest ascent or the conjugate-gradient method. The EM algorithm itself can be rewritten as a gradient ascent method (Kaufman 1987). In this form, it can be modified to include a line search at each iteration to enhance convergence. Kaufman (1987) also investigated using a standard conjugate-gradient approach in place of the EM algorithm. Interestingly, even in this case, the EM algorithm plays a role since using a preconditioner, derived by analogy to the EM algorithm, results in substantial improvements in the convergence rate. This simple diagonal preconditioner has been widely used by us and others as a means of improving convergence rates, e.g. Mumcuoglu *et al* (1996), Qi *et al* (1998b).

Rather than update all voxel values simultaneously as in the EM and gradient-based methods, we can instead update each voxel in turn as in the iterated coordinate-ascent (ICA) methods (Bouman and Sauer 1996). These methods optimize the cost function with respect to a single variable at each iteration, which can result in a relatively simple update step and surprisingly fast convergence given a good initial estimate. Updating a subset of the voxels at each iteration is referred to as a grouped coordinate-ascent method (Fessler *et al* 1997) and falls naturally between the coordinatewise and the gradient-based methods.

The EM algorithm is in fact a special case of a rather general class of algorithms that are referred to as 'functional substitution' or 'optimization transfer' methods (Lange *et al* 2000). These work as follows. The problem of maximizing likelihood involves the solution

of a set of highly nonlinear equations, which cannot be found in a closed form. Functional substitution methods replace the original cost function at each step with an alternative function, which when maximized, is guaranteed to also increase the value of the original function. By choosing the substituting functions carefully, reductions in computation time and speed-up in convergence can be realized. Among the class of functional substitution methods is the set of SAGE algorithms which generalize the idea of EM by specifying multiple sets of 'hidden data' (analogous to the complete data in EM) with respect to a subset of the voxels (Fessler and Hero 1995). Alternatively, we can combine the functional substitution approach with ICA, and at each iteration replace the function to be updated with a simpler approximation so that the update can be reduced to a closed form. As we will see later, functional substitution methods can also be used to solve penalized ML or Bayesian reconstruction problems in which the presence of a smoothing prior adds further complexity to the cost function.

A problem common to all ML methods in PET and SPECT is ill-conditioning, i.e. the solutions are sensitive to small changes in the data so that the estimates are of high variance. This ill-conditioning is seen in practice through the 'checkerboard' effect of high spatial variance in the ML images for high iteration numbers of the EM or OSEM algorithms. It can be investigated theoretically using the Fisher information matrix, from which lower bounds on the variance of unbiased estimators can be computed. Approaches to solving this instability problem include early termination of the iterations using heuristic or statistically based stopping rules (Veklerov and Llacer 1987, Coakley 1991, Johnson 1994), filtering of the reconstructed images during or after reconstruction (Llacer *et al* 1993, Silverman *et al* 1990), or using Grenander's methods of sieves (Grenander 1981, Snyder and Miller 1985).

An alternative solution to the ill-conditioning problem, which we will emphasize here, is to add a smoothing penalty to the likelihood function and compute a penalized ML solution (Fessler 1996). Equivalently, we can reformulate the problem in a Bayesian framework and include a prior distribution on the image that reflects our expectations of smoothness or other characteristics in the image (Geman and McClure 1985). We then compute an image as a MAP (maximum *a posteriori*) estimate from the posterior density for the image conditioned on the data. The purpose of the penalty term in penalized ML, or the prior in MAP, is then to select among the set of solutions that are essentially equivalent in terms of the data, the one which is most probable in terms of our prior expectations of image structure or smoothness.

The smoothing priors are chosen to reflect the property common to virtually all nuclear medicine images that they are locally smooth. The issue is complicated by the fact that the locally smooth images are often interrupted by sharp changes in intensity, reflecting large tracer density differences between neighbouring tissues of differing types. The large literature that now exists on different priors for nuclear medicine images is largely concerned with the problem of simultaneously satisfying the constraint that the images are locally smooth while still allowing rather sharp transitions in intensity. The simplest form of smoothing priors assumes statistical independence between voxels so that the image probability density function (pdf) is simply a product of univariate densities. The key attraction of these models is that we can often find closed form extensions of the EM algorithm when extended to penalized ML since there is no coupling between pixels introduced through the prior. However, these priors act primarily through the choice of their mean value. In practice, the means are often updated using the results of earlier iterations, which make it difficult to specify the properties of the method or the cost function that is ultimately optimized.

More common in the literature are images drawn from the broad class of Markov random field models whose pdfs are specified through functions defined in terms of interactions between neighbouring voxels. These priors are exponential in form where the negative exponent is the sum of a set of functions defined on the difference between pairs of neighbouring

pixels. By making this function nondecreasing, the prior effectively penalizes images that exhibit large differences between adjacent pixels. The specific choice of this function can greatly affect the appearance of the reconstructed images and they range from a simple Gaussian model (where we take the squared difference between voxels) to highly nonconvex functions that produce more sharply defined boundaries in the image. In section 5 we will describe these priors in more detail, their extensions to higher order interactions, and their use of anatomical priors.

The optimization algorithms for penalized ML or MAP estimation closely parallel those for the standard ML problem. For instance, the EM algorithm is directly extendible to MAP using the generalized EM (GEM) algorithm. The general gradient-based and coordinate-ascent methods extend directly to MAP as do many of the functional substitution methods. For this reason, the optimization algorithms are all reviewed in section 6. A special concern in MAP estimation is the existence of multiple local maxima when using nonconvex priors or priors with discrete variables that are typically used to represent the presence or absence of image boundaries between voxels. In this case, special programming methods, such as simulated annealing, can be used to find a global optimum, although these methods are typically impractical for routine use.

4. Formulation of the maximum-likelihood problem

The ML estimate is the tracer distribution that maximizes the probability of the observed data assuming no prior statistical model on the tracer distribution. In this section, we formulate the likelihood function for emission data and discuss the properties of the ML estimate.

4.1. The linear Poisson model

PET and SPECT data represent the detection of high-energy photons produced through radioactive decay. In the case of SPECT we directly observe gamma-ray emissions from the radiolabelled tracer; for PET we detect photon pairs produced through annihilation of positrons emitted from the tracer. Provided that the detection of each decay event by the system is independent and can be modelled as a Bernoulli process, then the sinogram data are a collection of independent Poisson random variables. Let M be the number of detector bins in the ECT system, the mean of the data, $\bar{\boldsymbol{y}} = \{\bar{y}_i, i = 1, \ldots, M\}$, is related to the unknown tracer distribution $f(\boldsymbol{x})$ through a transform that is defined by the physical characteristics of the ECT system,

$$\bar{y}_i = E[y_i] = \int f(\boldsymbol{x})\, c(i, \boldsymbol{x})\, \mathrm{d}\boldsymbol{x} + r_i + s_i, \qquad i = 1, \ldots, M, \tag{1}$$

where $c(i, \boldsymbol{x})$ is the detection response function for the ith measurement or line of response (LOR), and $\boldsymbol{r}$ and $\boldsymbol{s}$ represent the expectations of random events (in PET) and scattered and background events (in PET and SPECT), respectively. Although the source distribution $f(\boldsymbol{x})$ in (1) is actually a continuous function in the spatial domain, most researchers use a finite set of basis functions to represent $f(\boldsymbol{x})$ in image reconstruction, which converts the continuous–discrete data model in (1) to a discrete–discrete model. Readers are referred to Barrett and Myers (2003) and Lewitt and Matej (2003) for more detailed discussions on different imaging models. The most popular basis element is the 'voxel' in which image intensity is constant over the cubic support of the voxel and zero elsewhere. Alternative basis elements include smooth spherically symmetric 'blobs' (Matej and Lewitt 1996), wavelets (Frese *et al* 2002) and natural pixels (Gullberg *et al* 1996, Bhatia *et al* 1997, Hsieh *et al* 1998). Using a finite set

of basis functions will not result in information loss in the reconstructed images provided they are sufficiently closely spaced. The spatial basis functions are typically centred on a regular 2D or 3D lattice. For the voxel basis, spacing is equal to the voxel size. The image value at each voxel is proportional to the total number of radioactive nuclei contained in the volume spanned by the voxel. We will use a single index to represent the lexicographically ordered voxels, $\boldsymbol{f} = \{f_j, j = 1 \dots N\}$. Let $\phi_j(\boldsymbol{x})$ denote the jth voxel basis function, then

$$f_j = \int f(\boldsymbol{x})\, \phi_j(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x}. \tag{2}$$

The transform in (1) can be approximated as the affine transform,

$$\bar{\boldsymbol{y}} = E[\boldsymbol{y}] = \boldsymbol{P}\boldsymbol{f} + \boldsymbol{r} + \boldsymbol{s}, \tag{3}$$

where $\boldsymbol{P} \in \mathbb{R}^{M \times N}$ is the projection or system matrix. The element, $p_{ij} = \int c(i, \boldsymbol{x})\phi_j(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x}$, is the average probability of detecting an emission from voxel site j at LOR i. As discussed in the previous section, one of the advantages of iterative reconstruction methods is that the system matrix is not limited to a simple line-integral model but can also include factors such as depth-dependent sensitivity and spatially variant detector response.

Let $\boldsymbol{y} = \{y_i, i = 1, \dots, M\}$ denote the measured data. The conditional probability, or likelihood function, for $\boldsymbol{y}$, given the tracer distribution $\boldsymbol{f}$, is given by the independent Poisson distribution:

$$p(\boldsymbol{y}|\boldsymbol{f}) = \prod_i \mathrm{e}^{-\bar{y}_i} \frac{\bar{y}_i^{y_i}}{y_i!}. \tag{4}$$

The ML estimate of $\boldsymbol{f}$ can be computed by maximizing (4), although it is easier to work with the log of the likelihood function:

$$L(\boldsymbol{y}|\boldsymbol{f}) = \sum_{i=1}^{M} y_i \log(\bar{y}_i) - \bar{y}_i - \log(y_i!). \tag{5}$$

Since the log is a monotonic function, maximizing $p(\boldsymbol{y}|\boldsymbol{f})$ and $L(\boldsymbol{y}|\boldsymbol{f})$ will produce identical solutions.

The linear Poisson model does not strictly hold when the emission density is very high since all ECT scanners are limited in the rate at which they can count. This is usually represented by a 'dead-time' calibration factor, which is a measure of the fraction of time that the scanner is unable to record new events because it is occupied with photons that have already arrived. Here we will assume that the count rates are sufficiently low that the system is operating in the linear range. However, methods that account for this nonlinear behaviour are still needed for high count situations (Yu and Fessler 2000).

4.2. Estimation of random and scatter components

Random events in PET are caused by the detection of two photons from two independent annihilation events within the coincidence timing window. In most current scanners the randoms rate is estimated by recording the coincidence events using a delayed timing window of duration equal to that of the coincidence or 'prompt' window. Provided that the randoms rate is invariant between the prompt window and the delayed window, the randoms can be corrected for in the data by subtracting the delayed window events from the prompt window events. When operated in standard mode, PET scanners precorrect data in real time. Randoms subtraction corrects the mean of the data, but increases the variance. Therefore, the data cannot be modelled using the Poisson model: the variance of the prompt and delayed measurements are both equal to their means, since they are Poisson, while the variance of the randoms

corrected data equals the sum of the means of the prompt and delayed measurements. The true distribution of the corrected data has a numerically intractable form. A simple approximation is the shifted Poisson model in which the first two moments of the corrected data are matched by assuming that adding two times the mean randoms rate to the randoms corrected data, i.e. $y_i + 2r_i$, produces a random variable with mean $[\boldsymbol{Pf}]_i + 2r_i + s_i$ whose density has the form of the Poisson distribution. This results in the modified log likelihood (Yavuz and Fessler 1998),

$$L_{\mathrm{SP}}(\boldsymbol{y}|\boldsymbol{f}) = \sum_{i=1}^{M} [y_i + 2r_i]_+ \log([\boldsymbol{Pf}]_i + 2r_i + s_i) - ([\boldsymbol{Pf}]_i + 2r_i + s_i), \qquad (6)$$

where $[x]_+ = x$ if $x > 0$ and is 0 otherwise. Most of the results described later in this paper for the Poisson model can be directly modified to use the shifted Poisson model by replacing the Poisson log likelihood with (6).

The alternative saddle point approximation (Yavuz and Fessler 1998) results in less bias than the shifted Poisson model when applied to low count 3D PET data (Yavuz and Fessler 2000). A third alternative uses a variant on (6) that reduces the systematic bias in the shifted Poisson model by allowing negative sinogram values (Ahn and Fessler 2004). A different approach in which there is renewed interest is to work with uncorrected prompt data and separately estimate the mean of the randoms for use directly in the likelihood function (5). The mean of the randoms can be estimated from the delayed sinogram, or even directly from the individual detector singles rates, using various variance-reduction techniques (Casey and Hoffman 1986, Mumcuoglu *et al* 1996).

The likelihood function (5), which uses the data model (3), also requires an estimate of the mean of the scattered events in the data. A number of different approaches have been applied to this problem, including convolution methods (Bailey and Meikle 1994), dual energy window methods (Grootoonk *et al* 1996), Monte Carlo simulation (Levin *et al* 1995) and analytical methods based on the Klein–Nishima formula (Ollinger 1996, Watson 2000). Among these, the analytic and Monte Carlo are the more accurate, but also more time-consuming since they require a preliminary reconstruction of the emission and attenuation images prior to computing the scatter profile.

An alternative is to model the scatters as part of the system matrix and to estimate them simultaneously with the reconstructed image. This will cause the projection matrix to be non-sparse, which can significantly increase the computational cost. However, this approach has attracted some interest in SPECT reconstruction since the scatter fraction is relatively high and more closely follows the shape of the true events than does the scatter component in PET (Beekman *et al* 2002).

4.3. The list-mode likelihood function

The above Poisson data model is based on histogrammed data, i.e. the data are the number of events detected for each possible LOR in the sinogram. When the total number of detected events in a particular study is far less than the number of possible LORs, the histogram format is not efficient because most of the entries are zero. An alternative approach is to store the data in 'list mode', i.e. recording the detector coordinates for each event. Rather than rebinning the data into histograms, we can instead reconstruct images directly using list-mode maximum likelihood, as first proposed for time-of-flight PET data by (Snyder and Politte 1983). With the development of high-resolution detectors for PET and SPECT, list-mode approaches have recently received renewed interest (Barrett *et al* 1997, Reader *et al* 1998a, 1998b, Parra and Barrett 1998, Byrne 2001, Huesman *et al* 2000, Levkovitz *et al* 2001, Reader *et al* 2002). A second advantage of list-mode acquisition is that the arrival times of each

event can be recorded, offering the possibility of high temporal resolution in dynamic studies. Traditionally, dynamic data are histogrammed into a set of contiguous time intervals with images reconstructed independently for each interval or frame. The preselection of the frame duration limits the achievable temporal resolution. The problem of temporal binning can be avoided entirely by directly using the arrival times in the list-mode data to estimate a dynamic image (Snyder 1984, Nichols *et al* 2002).

The list-mode likelihood function for static data can either be formulated from first principles (Barrett *et al* 1997, Parra and Barrett 1998), or derived from the Poisson distribution (5) of the sinogram data as described in (Huesman *et al* 2000). The only difference lies in the treatment of the total number of detections K, which does not affect the ML estimate of the image. We denote the list-mode data by $\{i_k\}_{k=1}^{K}$, where i_k is the index of the LOR of the kth event. The likelihood function of the list-mode data is

$$p\left(\{i_k\}_{k=1}^{K}|\boldsymbol{f}\right) = \frac{1}{z}\left[\prod_{k=1}^{K}\bar{y}_{i_k}\right]\prod_{j=1}^{N}\mathrm{e}^{-\varepsilon_j f_j}, \qquad \varepsilon_j = \sum_{i=1}^{M} p_{ij}, \tag{7}$$

where z is a normalization factor that makes $p\left(\{i_k\}_{k=1}^{K}|\boldsymbol{f}\right)$ a valid pdf, and ε_j is the overall sensitivity of the system to events emitted from the jth voxel. Note that in this expression, K itself is still a Poisson random variable. The corresponding log-likelihood function is

$$L\left(\{i_k\}_{k=1}^{K}|\boldsymbol{f}\right) = \sum_{k=1}^{K}\ln\left(\bar{y}_{i_k}\right) - \sum_{j=1}^{N}\varepsilon_j f_j. \tag{8}$$

Despite the difference in data format, the likelihood functions for histogrammed and list-mode data contain identical information, and consequently most image reconstruction algorithms developed for histogrammed data can be modified to accommodate list-mode data and retain their convergence properties.

4.4. Properties of the ML estimate

The ML estimate corresponds to the tracer distribution that maximizes the probability of producing the measured data $\boldsymbol{y}$. The requirements for a local maximum of the function are determined by the Kuhn–Tucker conditions described below. A local maximum will also be a global maximum if the function is also concave (Luenberger 1984).

For the linear Poisson model (5), the first- and second-order derivatives of the log-likelihood function are

$$\frac{\partial L(\boldsymbol{y}|\boldsymbol{f})}{\partial f_j} = \sum_i p_{ij}\left(\frac{y_i}{\sum_k p_{ik}f_k + r_i + s_i} - 1\right) \tag{9}$$

$$\frac{\partial^2 L(\boldsymbol{y}|\boldsymbol{f})}{\partial f_j \partial f_l} = -\sum_i \frac{p_{ij}p_{il}y_i}{\left(\sum_k p_{ik}f_k + r_i + s_i\right)^2}. \tag{10}$$

A necessary and sufficient condition for the function to be concave is that the Hessian, i.e. the matrix of second derivatives, is negative semidefinite for all allowed images (Luenberger 1984), i.e.,

$$\sum_{j=1}^{N}\sum_{l=1}^{N} c_j c_l \frac{\partial^2 L(\boldsymbol{y}|\boldsymbol{f})}{\partial f_j \partial f_l} \leqslant 0 \tag{11}$$

for any vector $\boldsymbol{c} = \{c_1, c_2 \ldots c_N\}$ and any non-negative image $\boldsymbol{f} \geqslant 0$. By substituting (9) in (11) it is straightforward to show that this inequality is satisfied and hence the log-likelihood

function is concave. Similarly, one can also show that the shifted Poisson log-likelihood function (6) is concave.

Necessary and sufficient conditions for an image $\boldsymbol{f}^*$ to be a maximizer of $L(\boldsymbol{y}|\boldsymbol{f})$ subject to $\boldsymbol{f} \geqslant 0$ are given by the Kuhn–Tucker conditions:

$$\frac{\partial}{\partial f_j} L(\boldsymbol{y}|\boldsymbol{f})\bigg|_{\boldsymbol{f}=\boldsymbol{f}^*} \begin{cases} =0, & f_j^* > 0 \\ \geqslant 0, & f_j^* = 0 \end{cases} \quad j = 1, \ldots, N. \tag{12}$$

Direct solution of (12) is intractable. Instead, iterative algorithms are used that converge towards an image satisfying these conditions, as we describe in section 6. It is interesting to note that in the special case that $\boldsymbol{y} = \boldsymbol{P}\boldsymbol{f} + \boldsymbol{r} + \boldsymbol{s}$ has an exact solution, that solution is an ML estimate. In general, the equations are inconsistent, but an ML solution will still exist.

Maximum-likelihood estimators have attractive properties including asymptotic unbiasedness and efficiency, i.e. achieving the lowest variance among all unbiased estimators. The variance of an unbiased ML estimator is bounded by the Cramer–Rao inequality, which states that the variance of any unbiased estimator cannot be lower than the diagonal entries of the inverse of the Fisher information matrix, $\boldsymbol{F}$ (Sorenson 1980). The Fisher information matrix is defined as the inverse of minus the expected value of the Hessian matrix, given in (10) for the ECT Poisson model. Rewriting (10) in matrix form, it then follows that the variance of the ML estimate of the jth voxel is bounded by

$$\mathrm{var}_j \geqslant [\boldsymbol{F}^{-1}]_{jj} = \left[\boldsymbol{P}' \mathrm{diag}\left[\frac{1}{\bar{y}_i}\right] \boldsymbol{P}\right]^{-1}_{jj} \tag{13}$$

where the subscript jj denotes the jth diagonal element of the matrix.

The attractive properties of the ML estimator are asymptotic ones, and for finite data the ML estimator may not be unbiased and may exhibit high variance. This is often the case in ECT where unregularized ML image estimates are often noisy. The Cramer–Rao inequality provides a tool to investigate variance, and a clue to the noisy nature of ML images is contained in the ill-conditioning or even singularity of the Fisher information matrix. Using a sequence of approximations (Fessler 1996, Qi and Leahy 1999), the Cramer–Rao bound in (13) can be approximated by

$$\mathrm{var}_j \approx \frac{\sum_i p_{ij}^2}{N \sum_i p_{ij}^2 / \bar{y}_i} \sum_{i=0}^{N-1} \frac{1}{\lambda_i} \tag{14}$$

where λ_i are the Fourier coefficients of the jth column of the system response $\boldsymbol{P}'\boldsymbol{P}$. The operator $\boldsymbol{P}'\boldsymbol{P}$ corresponds to forward projection followed by backprojection and its jth column is the point spread function of the combined forward- and backprojection operators centred at the jth voxel. From Radon transform theory we know that in 2D, the combination of forward and backprojection produces a point spread function whose Fourier transform has the form $1/r$, where r is the radial frequency. With finite detector resolution in real imaging systems, the λ_i drop off faster than a $1/r$ rate in 2D. As the number of image voxels increase, $\min_i \lambda_i \to 0$, and hence $\mathrm{var}_j \to \infty$.

In practice, the variance of an ML estimator is often controlled by starting with a smooth initial estimate and terminating an ML search before convergence (Veklerov and Llacer 1987, Coakley 1991, Johnson 1994) or post-smoothing of the reconstruction (Llacer *et al* 1993, Silverman *et al* 1990). An alternative approach is to regularize the likelihood function as we describe below. In that case, we can again derive approximations to the variance of the estimator analogous to (14) but which are now stabilized so that the variance is finite (Fessler 1996, Qi and Leahy 1999).

4.5. Relation to other objective functions

It is interesting to note that the log-likelihood function (5) can be justified as a cost criterion without consideration of the Poisson model. For instance, the Kullback–Leibler cross entropy distance (Byrne 1993, 1995, 1997) is given by

$$KL(\boldsymbol{y}, \hat{\boldsymbol{y}}) = \sum_{i=1}^{M} y_i \log(y_i/\hat{y}_i) + \hat{y}_i - y_i, \tag{15}$$

where $\hat{y}_i = \sum_k p_{kj} x_j + s_i + r_i$. This expression differs from the negative log likelihood (5) by an additive constant independent of the image $\boldsymbol{f}$. Consequently, minimizing the cross entropy (15) is equivalent to maximizing the Poisson likelihood. Equation (15) is also referred to as the I-divergence in (Csiszar 1991). Formulation with respect to general entropy considerations rather than use of the Poisson likelihood function serves as a justification for using the methods reviewed later in this paper for a wider class of problems in which the data are not Poisson distributed.

An alternative to the Poisson log likelihood is the weighted least-squares (WLS) criterion

$$L_{WLS}(\boldsymbol{y}|\boldsymbol{f}) = -\frac{1}{2} \sum_{i=1}^{M} \frac{(\bar{y}_i - y_i)^2}{\hat{\sigma}_i^2} \tag{16}$$

where $\hat{\sigma}_i$ is an estimate of the variance of each measurement. This function was proposed for use with randoms corrected data in PET (Fessler 1994), where the variance should equal $\bar{y}_i + 2r_i + s_i$. Since the true mean is unknown, the variance can be approximated using the observed data as $y_i + 2\hat{r}_i + \hat{s}_i$, where $\hat{r}_i$ and $\hat{s}_i$ are estimates of the mean of the randoms and scatters, respectively. This quadratic approximation of the Poisson likelihood function has advantages in reducing computation cost in coordinate-ascent algorithms (see section 6).

A variation of weighted least-squares is model-weighted least-squares (Anderson *et al* 1997)

$$-\frac{1}{2} \sum_{i=1}^{M} \frac{(\bar{y}_i - y_i)^2}{\bar{y}_i}. \tag{17}$$

This function is applicable to Poisson data where the variance equals the mean $\bar{y}_i$ yet avoids the log function in the true likelihood. Unlike the WLS function (16), (17) is not quadratic; however it is convex. Although minimizing this function may result in some computational savings relative to maximum likelihood, the result will be negatively biased and some form of compensation is required to alleviate this problem (Anderson *et al* 1997).

As discussed earlier in the paper, the cost of iterative reconstruction is dominated by forward- and backprojection. A clever approach to reducing the computational cost in fully 3D PET while still using an iterative algorithm is to first rebin the data into a set of stacked 2D sinograms and then reconstruct each 2D slice separately (Comtat *et al* 1998). In order to apply Fourier rebinning, the data must first be corrected for attenuation and detector sensitivity. As a result, the stacked 2D sinograms will have quite different statistical noise properties than the original data. Approaches to dealing with this issue include use of a weighted least-squares function of the form (16) or reweighting of the stacked 2D data using the attenuation factors combined with use of the Poisson likelihood function (Comtat *et al* 1998, Lee *et al* 2004, Alessio *et al* 2003).

Each of the objectives described above is a measure of the fit of the estimated image to the observed data. Consequently, without regularization they will exhibit ill-conditioning, and hence high variance behaviour, similar to that of maximization of the Poisson likelihood.

5. Bayesian formulations and image priors

Here we consider explicit regularization procedures in which a prior distribution on the image, $p(\boldsymbol{f})$, is introduced through a Bayesian reformulation of the problem to resolve the ill-conditioning in the likelihood function.

The likelihood function and image prior are combined through the Bayes rule to form the posterior density

$$p(\boldsymbol{f}|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{f})p(\boldsymbol{f})}{p(\boldsymbol{y})}. \tag{18}$$

Bayesian formulations of the PET problem are usually reduced to computing a maximum *a posteriori* (MAP) estimate of the image as the maximizer of the posterior. Taking the log of the posterior and dropping constants, we have the basic form of the MAP objective function

$$\Phi(\boldsymbol{f}) = L(\boldsymbol{y}|\boldsymbol{f}) + \ln p(\boldsymbol{f}). \tag{19}$$

Thus the MAP estimator maximizes the sum of the log likelihood and the log of the prior. Some authors prefer to view this estimator as a penalized ML method, but the differences between MAP and penalized ML are largely semantic, except in the case where the penalty functions are explicit functions of the data, e.g. Fessler and Rogers (1996), and hence cannot be viewed as true priors. A third viewpoint is that of Tikhonov regularization, in which the objective function consists of the sum of a data fitting term (the log likelihood in (19)) and a regularization term (the log prior). All three points of view can lead to the same objective function and hence the same solution, but may differ in their interpretation of that solution and in the manner in which any free parameters in the prior are selected.

5.1. Spatially independent priors

The simplest form of prior assumes statistical independence between voxels. The independent Gaussian model (Levitan and Herman 1987, Huesman *et al* 2000) results in a simple quadratic form with the log prior equivalent to the L_2 norm on the image when zero mean is assumed. The gamma prior (Lange *et al* 1987) is conjugate for the Poisson likelihood function[3] so that the posterior density is also gamma. Furthermore, the gamma distribution is in some ways a more natural model for images since the prior allows only non-negative image values. A third class of independent priors is based on the use of the maximum entropy criterion (Gullberg 1995, Liang *et al* 1989, Nunez and Llacer 1990, Lalush *et al* 2000). In this case the prior has the form

$$p(\boldsymbol{f}) = \frac{1}{Z}\prod_{i=1}^{N} \mathrm{e}^{-f_i \ln f_i} \tag{20}$$

where Z is a normalizing constant. Taking the log of (20) and substituting in (19) results in an entropy-regularized cost function. Since entropy is correctly computed with respect to probabilities rather than actual image intensities, we can replace f_i in (20) with the normalized quantity $f_i \big/ \sum_j f_j$, although in this case the normalization produces coupling between voxels so that the prior is no longer spatially independent. The function is also readily modified to allow use of cross entropy.

The key attraction of the independent models is that the prior does not introduce any coupling between voxels. Hence we can often find closed form extensions of the EM ML

[3] A prior distribution is said to be *conjugate* to a likelihood if the resulting posterior distribution belongs to the same functional family as the prior.

algorithm (see the next section). However, these priors require specification of the mean image value, which can have a substantial effect on the estimated image and cause significant bias. To reduce this effect, a number of researchers have investigated estimating the image mean during reconstruction using either the median (Alenius and Ruotsalainen 1997, Alenius *et al* 1998, Alenius and Ruotsalainen 2002, Hsiao *et al* 2001a) or the mean (Hsiao *et al* 2001b) of neighbouring voxels. However, in this case the priors are not truly independent. In (Hsiao *et al* 2001a, 2001b) the coupling problem is solved by jointly estimating the image and the median or mean field.

5.2. Spatial interaction models and Gibbs priors

Rather than simply using the prior to separately control deviation of each voxel from its mean with the spatially independent models described above, we can instead capture the locally structured properties of images using Markov random field models, also known as Gibbs distributions. The Markovian properties of these distributions make them theoretically attractive as a formalism for describing empirical local image properties, as well as computationally appealing since the local nature of their density functions results in computationally efficient update strategies (Besag 1974).

Let $S = \{1, 2, \ldots, N\}$ denote the ordered set of image voxel indices. The Gibbs distribution is defined on a neighbourhood system which associates a set of sites $\mathcal{N}_j \subset S$ with each site j. The neighbourhood system must satisfy the property that $i \in \mathcal{N}_j$ if and only if $j \in \mathcal{N}_i$. The sites in $\mathcal{N}_j$ are typically the collection of voxels closest, up to some maximum Euclidean distance, to site j. The Gibbs distribution has the general form

$$p(\boldsymbol{f}|\beta) = \frac{1}{Z} \mathrm{e}^{-\beta U(\boldsymbol{f})} \tag{21}$$

where $U(\boldsymbol{f})$ is the Gibbs energy function defined as a sum of potentials, each of which is a function of a subset or clique $c_k \subset S$. The cliques for a particular neighbourhood system must satisfy the condition that each pair of sites in each clique c_k are mutual neighbours (Geman and Geman 1984).

The form of Gibbs distributions most commonly used in image processing is that for which the energy function $U(\boldsymbol{f})$ contains potentials defined on pairwise cliques of neighbouring voxels:

$$U(\boldsymbol{f}) = \sum_{j=1}^{N} \sum_{k \in \mathcal{N}_j, k > j} \phi_{jk}(f_j - f_k). \tag{22}$$

For a 3D problem, the neighbours of an internal voxel would be the nearest 6 voxels for a first-order model, or the nearest 26 voxels for a second-order model (with appropriate modifications for the boundaries of the lattice).

The potential functions $\phi_{jk}(f_j - f_k)$ are chosen to attempt to reflect the locally smooth characteristic of tracer distributions within a specific organ or type of tissue while still allowing abrupt changes as we move between different organs or tissue types. A wide range of functions have been studied in the literature that attempt to produce local smoothing while not removing or blurring true boundaries or edges in the image. All have the basic property that they are monotonic nondecreasing functions of the absolute intensity difference $|(f_j - f_k)|$. Taking the square of this function leads to a Gauss–Markov prior which produces smooth images with very low probability of sharp transitions in intensity. In an attempt to increase the probability of these sharp transitions, Bouman and Sauer (1996) propose using the generalized p-Gaussian model where $\phi(f) = |f|^p$, $1 < p < 2$. An alternative function with similar behaviour, which

derives from the literature on robust estimation, is the Huber prior in which the continuous function $\phi(f)$ switches from quadratic to linear at a user-specified transition point (Huber 1981). The function $\phi(f) = \log\cosh(cf)$, where c is a user-specified parameter, has similar shape to that of the Huber prior but has the advantage that its derivatives are continuous (Green 1990). The potential functions mentioned so far are all convex and so is the corresponding Gibbs energy function.

In an attempt to produce sharper intensity transitions, several nonconvex functions have also been proposed. For example, Geman and McClure (1985), who were also the first to specifically use Gibbs distributions in emission tomography, proposed the function $\phi(f) = \frac{f^2}{f^2+\delta^2}$. This and other nonconvex potentials have the property that the derivative of the function decreases once the intensity difference exceeds some threshold. The limiting case of this approach is found in the weak membrane model which is quadratically increasing up to some threshold and then remains constant beyond this (Gindi *et al* 1993).

Higher order neighbourhoods are able to capture a more complex correlation structure than the simple pairwise models (Chan *et al* 1995). Unfortunately, the problem of choosing and justifying such a model becomes increasingly difficult with the size of the neighbourhood. One example of a higher order neighbourhood is the thin plate spline model of (Lee *et al* 1995). This model uses a discrete approximation to the bending energy of a thin plate as the Gibbs energy function. Since the thin plate energy involves second-order derivatives, higher order cliques must be used in the model.

Rather than implicitly modelling region boundaries as in the examples above, the compound MRFs include a second coupled random field, defined on a dual lattice of 'line sites', which explicitly represents the boundaries in the image (Geman and Geman 1984). The prior energy becomes

$$U(\boldsymbol{f}) = \sum_{j=1}^{N} \sum_{k\in\mathcal{N}_j, k>j} [(1 - l_{jk})(f_j - f_k)^2] + \psi(\boldsymbol{l}), \tag{23}$$

where the line sites l_{jk} are placed between voxels j and k in the image lattice and are set to unity if there is an image boundary between that pair of voxels, otherwise they are set to zero, and $\psi(\boldsymbol{l})$ encourages the formation of connected boundaries (Geman and Geman 1984, Johnson *et al* 1991, Leahy and Yan 1991, Yu and Fessler 2002, Feng *et al* 2003). For the case where $\psi(\boldsymbol{l})$ is the L_2 norm of $\boldsymbol{l}$ the compound MRF is equivalent to the weak membrane model (Lee *et al* 1995). It is interesting to note that most non-quadratic potential functions can be converted into the form of (23) by allowing l_{jk} to take continuous values between 0 and 1 (Geman and Reynolds 1992, Geman and Yang 1995, Idier 2001).

One of the primary attractions of the compound MRFs is that they serve as a natural framework for incorporating anatomical information into the reconstruction process (Leahy and Yan 1991, Gindi *et al* 1991, 1993, Sastry and Carson 1997, Wang *et al* 2004). Since different anatomical structures have different physiological functions, we can expect to see differences in tracer uptake between structures. This general observation is borne out in high-resolution autoradiographic images in which functional images also clearly reveal the morphology of the underlying structures (Gindi *et al* 1993). Because the anatomical modalities, such MR and CT, have superior resolution to PET, accurate estimates of anatomical boundaries can be formed from them. These can then be used to influence the formation of the boundary process in the PET image estimation procedure.

In Bowsher *et al* (1996) a hierarchical MRF model for emission tomography is described in which the image consists of regions of activity, each of which has an intensity that is assumed to vary smoothly about its mean. The reconstruction procedure involves estimating

the support of each of these regions and their mean intensities. These models can also be extended to include anatomical prior information (Lipinski *et al* 1997, Rangarajan *et al* 2000, Nuyts *et al* 2005).

5.3. The maximum a posteriori estimate

Substituting the log prior of the Gibbs distribution in (21) into the MAP objective function (19) produces the objective:

$$\Phi(\boldsymbol{f}) = L(\boldsymbol{y}|\boldsymbol{f}) - \beta U(\boldsymbol{f}). \tag{24}$$

The log-likelihood functions in section 4 are concave so that $\Phi(\boldsymbol{f})$ will also be concave if the Gibbs energy $U(\boldsymbol{f})$ is convex. In this case, a global maximum can be found using standard nonlinear optimization methods.

The Kuhn–Tucker necessary condition for $\boldsymbol{f}$ to be a maximizer of (24) is similar to (12) with $\frac{\partial}{\partial f_j} L(\boldsymbol{y}|\boldsymbol{f})$ being replaced by $\frac{\partial}{\partial f_j} L(\boldsymbol{y}|\boldsymbol{f}) - \beta \frac{\partial}{\partial f_j} U(\boldsymbol{f})$. This condition is also sufficient for convex priors.

If MRFs with nonconvex potential functions are used then local maxima will exist for the MAP objective function. Global search techniques such as simulated annealing are typically impractical for PET reconstruction because of the number of voxels in a typical image. Furthermore, the non-local property of the forward and backward projection results in a fully coupled posterior, i.e. the posterior density does not share the local neighbourhood system of the prior. In practice, local searches for a stationary point of the objective function are typically used to find a local maximum of $\Phi(\boldsymbol{f})$. In addition to the difficulty in computing MAP estimates for nonconvex priors, it should be noted that the solutions are also inherently discontinuous with respect to the data. This can result in high variance estimates. When using compound MRFs, the reconstruction problem is further complicated by the need to estimate the additional set of discrete variables that represent image boundaries.

A broader class of Bayesian estimators can be developed that minimize an expected loss, computed over the posterior density, that is designed to optimize performance for a specific task. One example of this broader view of Bayesian inference in ECT can be found in Bowsher *et al* (1996). Several other researchers, e.g. Green (1996), Weir (1997), Higdon *et al* (1997), have proposed alternative Markov chain Monte Carlo methods for image estimation. These methods are attractive since they are better suited to more complex priors and the use of hyperprior densities on unknown parameters; however, the high computational burden is a major obstacle preventing their widespread use.

6. Optimization methods

To compute an ML or MAP image estimate we maximize the appropriate objective function or, equivalently, solve the set of coupled nonlinear equations given by the Kuhn–Tucker conditions. Since closed form solutions do not generally exist, iterative estimation algorithms are used. The underlying principle of iterative optimization algorithms is as follows: starting from an initial estimate, $\boldsymbol{f}^0$, find a sequence of images, $\boldsymbol{f}^k, k = 1, 2, \ldots$, that converges to $\boldsymbol{f}^*$, a maximizer of the objective function, as $k \to \infty$ (assuming the sequence converges). All algorithms can be written in the following update form,

$$\boldsymbol{f}^{k+1} = \boldsymbol{f}^k - \alpha^k \boldsymbol{a}^k, \tag{25}$$

where $\boldsymbol{a}^k$ is a vector of the same dimension as the image and represents a search direction and α^k is a scalar step size. Different approaches to finding $\boldsymbol{a}^k$ and α^k give rise to different algorithms, as we review below.

6.1. Classical gradient-based and coordinatewise optimization methods

Gradient-based algorithms. Standard gradient-based algorithms have been used to solve emission tomography problems since the early 1970s (Goitein 1972, Budinger and Gullberg 1974). As the name implies, the search direction is calculated from the gradient of the objective function at the current iterate. The simplest method is that of steepest ascent, which sets $\boldsymbol{a}^k$ equal to the gradient and chooses the step size α^k using a one-dimensional (1D) line search algorithm, such as the Newton–Raphson method or the Armijo rule (Luenberger 1984). Convergence of steepest ascent is slow but can be enhanced using positive-definite preconditioning matrices, which are used to modify the search direction (Lewitt and Muehllehner 1986, Kaufman 1987, Rajeevan *et al* 1992). Preconditioners alone have limited effectiveness, but when combined with the conjugate-gradient algorithm, rather than steepest ascent, they produce substantial gains in convergence rates (Kaufman 1993, Mumcuoglu *et al* 1994, 1996, Qi *et al* 1998a, 1998b). A general form of the preconditioned conjugate-gradient algorithm is

$$\begin{aligned}
\boldsymbol{f}^{(k+1)} &= \boldsymbol{f}^{(k)} + \alpha^{(k)}\boldsymbol{a}^{(k)} : \ \alpha \text{ from line search} \\
\boldsymbol{a}^{(k)} &= \boldsymbol{d}^{(k)} + \beta^{(k-1)}\boldsymbol{a}^{(k-1)} \\
\boldsymbol{d}^{(k)} &= \boldsymbol{C}^{(k)}\boldsymbol{g}^{(k)} \\
\beta^{(k-1)} &= \frac{(\boldsymbol{g}^{(k)} - \boldsymbol{g}^{(k-1)})'\boldsymbol{d}^{(k)}}{\boldsymbol{g}^{(k-1)'}\boldsymbol{d}^{(k-1)}}
\end{aligned} \tag{26}$$

where $\boldsymbol{g}^{(k)}$ is the gradient vector of the objective function at $\boldsymbol{f} = \boldsymbol{f}^{(k)}$, $\boldsymbol{C}^{(k)}$ is the preconditioning matrix, and $\alpha^{(k)}$ is the step size found using a line search. For non-quadratic functions, $\boldsymbol{a}^k$ is not guaranteed to be an ascent direction so that it is necessary to check whether this is the case at each iteration, i.e. $\boldsymbol{g}^{(k)'}\boldsymbol{a}^{(k)} > 0$. We reset $\boldsymbol{a}^k$ to $\boldsymbol{d}^{(k)}$ if the condition is violated (Luenberger 1984).

Ideally, the preconditioner would be the inverse of the Hessian of the objective function. Exact computation of the inverse of the Hessian is impractical and approximations are used instead. A simple but effective diagonal preconditioner, inspired by the form of the EM ML algorithm described below, is (Kaufman 1993)

$$\boldsymbol{C}^{(k)} = \operatorname{diag}\left\{\frac{f_j^{(k)} + \delta}{\sum_i p_{ij}}\right\} \tag{27}$$

where δ is a small positive number to ensure that $\boldsymbol{C}^{(k)}$ is positive definite. An alternative diagonal preconditioner uses the inverse of the diagonal elements of the Hessian matrix (Johnson *et al* 2000). More complicated approaches to approximating the inverse of the Hessian have also been investigated and include Fourier-based methods (Clinthorne *et al* 1993, Fessler and Booth 1999) and matrix factorization (Chinn and Huang 1997).

One of the major challenges in the use of gradient-based methods in emission tomography is inclusion of the non-negativity constraint. The simplest approach is to restrict the step size in (25) so that all $f_j^{(k+1)}$ are non-negative, but this restriction slows down or even prevents convergence (Kaufman 1987). The bent line search overcomes the problem by allowing a larger step size so that voxel values can take negative values; a truncation operator is then applied to remove negative values before moving to the next iteration. A refinement of the bent line search uses a second line search in the direction formed as the difference between the truncated update and the image from the previous iteration (Kaufman 1993). A similar method is the active set approach in which an unconstrained search is performed over the subset of voxels identified as not belonging to the 'active set' of zero-valued voxels. The active set is then updated based on the gradient at the current estimate (Mumcuoglu *et al* 1996).

Penalty functions can be used to convert the constrained optimization to an unconstrained one (Luenberger 1984) but care must be taken to modify the preconditioner to account for the effect of active penalties on the Hessian (Mumcuoglu *et al* 1994). Primal-dual interior point methods can also be used to handle the positivity constraint (Johnson *et al* 2000).

Coordinate-accent (CA) algorithms. Seeking easier means to deal with the non-negativity constraint, researchers have also turned to coordinatewise algorithms. While there are a number of variations on this basic theme (Fessler 1994, Bouman and Sauer 1996, Sauer and Bouman 1993, Zheng *et al* 2000), the essence of these methods is to update each voxel in turn so as to maximize the objective function with respect to that voxel. Given the current estimate $\boldsymbol{f}^{(k)}$, the update for voxel $j = 1 + (k - 1) \bmod N$ is

$$\begin{aligned} f_j^{(k+1)} &= \arg\max_{f \geqslant 0} \Phi\left(f_1^{(k)}, f_2^{(k)}, \ldots, f_{j-1}^{(k)}, f, f_{j+1}^{(k)}, \ldots, f_N^{(k)}\right), \\ f_l^{(k+1)} &= f_l^{(k)}, \qquad l \neq j. \end{aligned} \tag{28}$$

The sequential update changes the multi-dimensional optimization problem into a series of one-dimensional optimizations and makes the imposition of non-negativity constraints straightforward. Although many algorithms can be used to solve the 1D maximization problem, exact solution of (28) requires repeated computation of the derivative of the objective function, which in turn involves forward- and backprojection and would greatly increase the per iteration computational cost. To solve this problem, Gaussian approximations of the log-likelihood function have been used (Sauer and Bouman 1993, Fessler 1994). Bouman and Sauer (1996) approximate the log-likelihood function by a quadratic function at each iteration using a Newton–Raphson approach. However, the Newton–Raphson update does not guarantee monotonicity and convergence. A method to restore these desired properties is to use functional substitution (Zheng *et al* 2000) (see section 6.2).

Coordinate-ascent algorithms can achieve fast convergence rates if given a good initial image (e.g., a filtered backprojection reconstruction). However, when starting from a uniform image, initial convergence can be very slow. This is because in CA algorithms high-frequency components converge much faster than low-frequency components (Sauer and Bouman 1993, Fessler 1994). In addition, if a CA algorithm updates voxels in a raster-scan fashion, then the algorithm will exhibit a faster convergence rate in the scan direction than in the orthogonal direction. Therefore, it is preferable to update the image voxels using either four different raster-scan orderings or a random ordering. We should note that besides simultaneous update and sequential update algorithms, there exist algorithms that update a group of voxels at a time (Fessler *et al* 1997, Zheng *et al* 2000). These methods can potentially reduce reconstruction time by grouping the voxels properly, but are more difficult to implement.

6.2. *Functional substitution methods*

As an alternative to the classical-gradient and coordinate-ascent methods described above, many investigators have studied alternative approaches that exploit specific aspects of the emission tomography problem such as the sparse nature of the projection matrix $\boldsymbol{P}$ or the Poisson form of the likelihood function. Many of these approaches belong to a class referred to as 'functional substitution' (FS) or 'optimization transfer' methods, which replace the original cost function at each step with an alternative or surrogate function, which when maximized, is guaranteed to also increase the value of the original function. By choosing the substituting functions carefully, reductions in computation time and speed-up in convergence

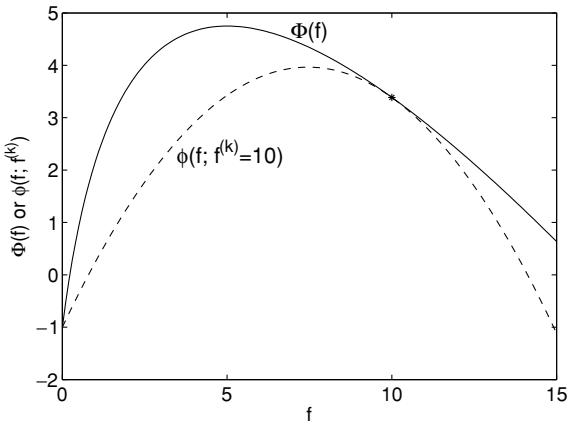

Figure 1. A one-dimensional example of functional substitution where a non-quadratic function $\Phi(f)$ is substituted by a quadratic function $\phi(f; f^{(k)})$ at $f^{(k)} = 10$ (denoted by '*').

can be realized. Two conditions are generally required for the substituting function $\phi(\boldsymbol{f}; \boldsymbol{f}^{(k)})$:

$$\Phi(\boldsymbol{f}) - \Phi(\boldsymbol{f}^{(k)}) \geqslant \phi(\boldsymbol{f}; \boldsymbol{f}^{(k)}) - \phi(\boldsymbol{f}^{(k)}; \boldsymbol{f}^{(k)}), \tag{29}$$

$$\nabla\Phi(\boldsymbol{f})|_{\boldsymbol{f}=\boldsymbol{f}^{(k)}} = \nabla_{\boldsymbol{f}}\phi(\boldsymbol{f}; \boldsymbol{f}^{(k)})|_{\boldsymbol{f}=\boldsymbol{f}^{(k)}}. \tag{30}$$

Condition (29) guarantees that any increase in the substituting function $\phi(\boldsymbol{f}; \boldsymbol{f}^{(k)})$ will result in at least the same amount of increase in the original objective function $\Phi(\boldsymbol{f})$; condition (30) guarantees that $\boldsymbol{f}^{(k)}$ will not be a maximizer of the substituting function $\phi(\boldsymbol{f}; \boldsymbol{f}^{(k)})$ if it is not also a maximizer of the original objective function $\Phi(\boldsymbol{f})$. These two conditions guarantee that the original objective function can be monotonically increased by maximizing the substituting function at each iteration

$$\boldsymbol{f}^{(k+1)} = \arg\max_{\boldsymbol{f}} \phi(\boldsymbol{f}; \boldsymbol{f}^{(k)}). \tag{31}$$

Figure 1 shows a one-dimensional example of functional substitution where a non-quadratic function $\Phi(f)$ is substituted by a quadratic function $\phi(f; f^{(k)})$ at $f^{(k)} = 10$. While the maximizer of the substituting function $\phi(f; f^{(k)})$ does not coincide with that of the original function $\Phi(f)$, maximizing $\phi(f; f^{(k)})$ will guarantee an increase in $\Phi(f)$ and applying the FS scheme iteratively will converge to the maximizer of $\Phi(f)$. A proof of convergence of FS methods can be found in Fessler and Hero (1995). Readers can also find a nice review in Lange *et al* (2000). A closely related method is the widely used Newton–Raphson algorithm where a quadratic function is used to approximate the original function locally at each iteration. However, the Newton–Raphson algorithm does not guarantee monotonic increase of the objective function, so, strictly speaking, it is not a FS method as discussed here.

The EM and GEM algorithms. The best known example of a functional substitution method in ECT is the EM algorithm of Dempster *et al* (1977), which was first applied to maximum-likelihood PET reconstruction by Shepp and Vardi (1982) and Lange and Carson (1984). The EM algorithm is a general framework for computing maximum-likelihood estimates through specification of a 'complete' but unobserved data set. Each iteration consists of an E-step—computing the conditional expectation of the log likelihood of the complete data, and an M-step—maximization of this conditional expectation with respect to the image. While the

EM method of (Dempster *et al* 1977) was originally developed from the statistical relationship between the probability densities for the complete and incomplete data, in the context of ECT the method can be more easily viewed as a functional substitution method in which the conditional expectation of the E-step forms the surrogate function at each iteration.

The complete data for the ECT reconstruction problem is the set of unobservable random variables $\boldsymbol{w} = \left\{\{w_{ij}\}_{j=1}^{N}\right\}_{i=1}^{M}$ with each w_{ij} representing the total emissions from voxel j detected at LOR i (Lange and Carson 1984, Vardi *et al* 1985). The complete data were selected so that if observed, the maximum-likelihood problem is readily solvable. The two steps of the EM algorithm are then

$$\text{E-step}: \quad \phi(\boldsymbol{f}; \boldsymbol{f}^{(k)}) = E[L(\boldsymbol{w}|\boldsymbol{f})|\boldsymbol{y}; \boldsymbol{f}^{(k)}]$$
$$= \sum_j \left[f_j^{(k)} \sum_i \frac{p_{ij} y_i}{\left(\sum_l p_{il} f_l^{(k)} + r_i + s_i\right)} \log(p_{ij} f_j) - f_j \sum_i p_{ij} \right], \tag{32}$$

$$\text{M-step}: \quad \boldsymbol{f}^{(k+1)} = \arg\max_{\boldsymbol{f}} \phi(\boldsymbol{f}; \boldsymbol{f}^{(k)}) = f_j^{(k)} e_j(\boldsymbol{f}^{(k)}) \Big/ \sum_i p_{ij},$$

where

$$e_j(\boldsymbol{f}^{(k)}) = \sum_i \frac{p_{ij} y_i}{\left(\sum_l p_{il} f_l^{(k)} + r_i + s_i\right)}. \tag{33}$$

The E-step calculates a surrogate function $\phi(\boldsymbol{f}; \boldsymbol{f}^{(k)})$ which can be shown to satisfy the two conditions in (29) and (30). Because the surrogate function in the EM algorithm is separable, the M-step can be solved in a closed form, as shown in (32). It is interesting to note that $\phi(\boldsymbol{f}; \boldsymbol{f}^{(k)})$ can also be derived directly, using the concavity of the log-likelihood function and Jensen's inequality, without involving the complete data concept (De Pierro 1993).

The EM algorithm is shown above for the Poisson data model that includes randoms, r_i, and scatter, s_i. Setting these to zero in (33) results in the original update equation derived in Shepp and Vardi (1982). Vardi *et al* (1985) showed that the EM algorithm is equivalent to the alternating I-divergence minimization of Csiszar and Tusnady (1984). It is also interesting to note that the same update equation (32) (without r_i and s_i) was earlier developed for restoration of astronomy images without explicit recourse to either the EM algorithm or functional substitution methodology (Richardson 1972, Lucy 1974).

The EM algorithm (32) converges monotonically to a global maximizer of the likelihood function (Vardi *et al* 1985) and guarantees non-negativity. When initialized with a positive image, the following relations hold at each iteration k:

$$\text{Non-negativity:} \qquad f^{(k)} \geqslant 0, \qquad j = 1, \ldots, N. \tag{34}$$

$$\text{Monotonicity:} \quad L(\boldsymbol{y}|\boldsymbol{f}^{(k+1)}) > L(\boldsymbol{y}|\boldsymbol{f}^{(k)}), \text{ unless } \boldsymbol{f}^{(k+1)} = \boldsymbol{f}^{(k)}. \tag{35}$$

$$\text{Count preservation } (\boldsymbol{r} = \boldsymbol{s} = 0): \quad \sum_i \sum_j p_{ij} f_j^{(k)} = \sum_i y_i. \tag{36}$$

The simple update equation (32) together with these properties make the EM algorithm an attractive approach. Unfortunately its converge rate is very slow; an issue we return to below.

The EM algorithm can be directly extended to MAP reconstruction where it is generally refereed to as generalized EM (GEM). Using the same set of complete data, the E-step remains the same. However, the M-step now has to maximize the expected value of the log posterior density function:

$$\boldsymbol{f}^{(k+1)} = \arg\max_{\boldsymbol{f} \geqslant 0} \sum_j \left(f_j^{(k)} \sum_i \frac{p_{ij} y_i \log(p_{ij} f_j)}{\left(\sum_l p_{il} f_l^{(k)} + r_i + s_i\right)} - f_j \sum_i p_{ij} \right) - \beta U(\boldsymbol{f}). \tag{37}$$

The difficulty of solving (37) is highly dependent on the form of the prior energy function $U(\boldsymbol{f})$. In the case of spatially independent priors, such as the gamma (Lange *et al* 1987, Wang and Gindi 1997) or independent Gaussian prior (Levitan and Herman 1987, Huesman *et al* 2000), the objective function in (37) is again separable and the M-step can be solved directly. However, for the coupled MRFs described in section 5 the surrogate function is not separable and direct solution is generally not possible. Instead one or more iterations of gradient or coordinate ascent are performed during the M-step (Hebert and Leahy 1989).

One approach to achieving a direct solution to the M-step in (37) is the 'one-step-late' method (Green 1990) where the partial derivatives of $U(\boldsymbol{f})$ are evaluated at the current estimate $f^{(k)}$, resulting in the closed form update equation

$$f_j^{(k+1)} = \frac{f_j^{(k)}}{\sum_i p_{ij} + \beta \frac{\partial}{\partial f_j} U(\boldsymbol{f})|_{\boldsymbol{f}=\boldsymbol{f}^{(k)}}} \sum_i \frac{p_{ij} y_i}{\sum_l p_{il} f_l^{(k)} + r_i + s_i}. \tag{38}$$

Unfortunately this method does not share the nice properties of the EM algorithm: in general it does not converge to a MAP solution and the non-negativity constraint needs special handling. Lange (1990) modified OSL by introducing a one-dimensional line search so that it has better convergence properties; however, in that case, a closed form update equation no longer exists.

An alternative approach is to apply the functional substitution methodology to the prior term $U(\boldsymbol{f})$ as well as the log likelihood (De Pierro 1995). For example, for the pairwise MRF model,

$$U(\boldsymbol{f}) = \sum_{j=1}^{N} \frac{1}{2} \sum_{l \in \mathcal{N}_j} \frac{1}{2} \kappa_{jl} (f_j - f_l)^2, \tag{39}$$

an appropriate surrogate function is

$$U_s(\boldsymbol{f}; \boldsymbol{f}^{(k)}) = \frac{1}{2} \sum_{j=1}^{N} \sum_{l \in \mathcal{N}_j} \frac{1}{2} \kappa_{jl} \left(2 f_j - f_j^{(k)} - f_l^{(k)}\right)^2. \tag{40}$$

If we replace $U(\boldsymbol{f})$ in (37) with $U_s(\boldsymbol{f}; \boldsymbol{f}^{(k)})$, the M-step can now be solved directly because the cost function is again separable. This algorithm retains some of the nice properties of the ML EM algorithm, such as guaranteeing non-negativity, and monotonically increases the objective function. Interestingly, a closed form update equation is possible even for non-quadratic Gibbs energies (Chang *et al* 2004).

SAGE algorithms. To accelerate the EM and GEM algorithms, and to avoid the difficulty in solving the M-step for the GEM algorithm, the space-alternating generalized EM (SAGE) algorithms (Fessler and Hero 1995) update the parameters sequentially using a sequence of smaller 'hidden' data spaces. The hidden data spaces introduced are less informative than the complete data space of the EM algorithm, resulting in faster convergence.

The SAGE surrogate function for the Poisson log likelihood is

$$\phi(f_j; \boldsymbol{f}^{(k)}) = \left(f_j^{(k)} + z_j^{(k)}\right) e_j(\boldsymbol{f}^{(k)}) \log\left(f_j + z_j^{(k)}\right) - \left(f_j + z_j^{(k)}\right) \sum_i p_{ij} \tag{41}$$

where $\phi(f_j; \boldsymbol{f}^{(k)})$ is a one-dimensional function of f_j. Introduction of the variable $z_j^{(k)}$ reduces the curvature of $\phi(f_j; \boldsymbol{f}^{(k)})$ while ensuring that $\phi(f_j; \boldsymbol{f}^{(k)})$ satisfies conditions (29) and (30). The optimal value of $z_j^{(k)}$ is (Fessler and Hero 1995)

$$z_j^{(k)} = \min_{i: p_{ij} \neq 0} \left\{ \left(r_i + s_i + \sum_{l \neq j} p_{il} f_l^{(k)} \right) \Big/ p_{ij} \right\}. \tag{42}$$

The SAGE algorithm uses sequential update, which makes the M-step a one-dimensional optimization problem and imposition of non-negativity constraints trivial. For an ML estimate, the update equation can be found by setting the derivative of (41) to zero for the closed form solution,

$$f_j^{(k+1)} = \left[\left(f_j^{(k)} + z_j^{(k)} \right) e_j(\boldsymbol{f}^{(k)}) \Big/ \sum_i p_{ij} - z_j^{(k)} \right]_+, \tag{43}$$

where $[x]_+ \equiv \max\{0, x\}$. For MAP reconstruction with a log-quadratic prior of the form in (39), the SAGE update again has a closed form,

$$f_j^{(k+1)} = \left[-B_j + \sqrt{B_j^2 + \left(f_j^{(k)} + z_j^{(k)} \right) e_j(\boldsymbol{f}^{(k)}) \Big/ \left(\beta \sum_{l \in N_j} \kappa_{jl} \right)} - z_j^{(k)} \right]_+, \tag{44}$$

where

$$B_i = \left[\sum_i p_{ij} - \beta \sum_{l \in N_j} \kappa_{jl} \left(f_l^{(k)} + z_j^{(k)} \right) \right] \Big/ \left(2\beta \sum_{l \in N_j} \kappa_{jl} \right). \tag{45}$$

Paraboloidal surrogate methods. The EM and SAGE algorithms construct the surrogate function based on statistical arguments. The surrogate functions can also be derived purely in terms of function approximation, which can result in faster algorithms because of the greater degree of freedom in choosing the surrogates. An early example of such an FS algorithm is the MAP estimator of De Pierro (1995), where a separable function was used to approximate the prior energy function in solving the M-step of the GEM algorithm. The paraboloidal surrogate (PS) method (Fessler and Erdogan 1998) approximates the log-likelihood function around $\boldsymbol{f}^{(k)}$ using a non-separable quadratic surrogate $\phi(\boldsymbol{f}; \boldsymbol{f}^k)$,

$$\phi(\boldsymbol{f}; \boldsymbol{f}^k) = \sum_i h_i\left(l_i^k\right) + \dot{h}_i\left(l_i^k\right)\left(l - l_i^k\right) - \tfrac{1}{2} n_i\left(l_i^k\right)\left(l - l_i^k\right)^2, \tag{46}$$

where

$$h_i(l) = y_i \log(l + r_i + s_i) - (l + r_i + s_i). \tag{47}$$

$$n_i(l) = \begin{cases} \dfrac{2}{l^2}[h_i(l) - h(0) - l\dot{h}_i(l)], & l > 0 \\ -\ddot{h}_i(l), & l = 0, \end{cases} \tag{48}$$

$l_i = (\boldsymbol{P}\boldsymbol{f})_i$ and $l_i^k = (\boldsymbol{P}\boldsymbol{f}^k)_i$. It is straightforward to verify that the surrogate function in (46) satisfies conditions (29) and (30). Note that $h_i(l)$ exist only when $r_i + s_i > 0$. Unlike the EM surrogate, (46) is not separable, which means maximizing (46) (with an additional prior energy for the MAP case) requires the use of gradient or coordinate-accent algorithms. However, the non-separable substituting function can potentially approximate the original cost function more closely and accordingly produce faster convergence rates. Combining the paraboloidal surrogate function (46) with a coordinate-ascent algorithm produces a rapidly converging algorithm that alternates between the two steps: (i) computing the surrogate function and (ii) maximizing the new objective function (sum of the surrogate function and prior term) (Fessler and Erdogan 1998).

Functional substitution in coordinate accent. In the algorithms described above, functional substitution is used to decouple interaction between voxels through the selection of separable surrogate functions (EM and GEM methods) or to find a quadratic approximation of the log-likelihood function (paraboloidal surrogate method). Another application of functional

substitution is to replace the one-dimensional Newton–Raphson algorithm, which is often used in coordinate-ascent algorithms, to achieve monotonicity or convergence. For example, in (Zheng *et al* 2000), a quadratic substitute functional $\phi(f_j; \boldsymbol{f}^k)$, which satisfies the conditions in (29) and (30), was constructed to approximate the log-likelihood function around $\boldsymbol{f}^k$ with respect to f_j. Compared to the Newton–Raphson quadratic approximation in (Bouman and Sauer 1996), this FS method guarantees monotonic convergence to a MAP solution. In addition, FS can also be used to derive group coordinate-accent algorithms (Zheng *et al* 2000), i.e., to make coordinate-ascent algorithms parallelizable.

6.3. Acceleration using ordered subsets

All of the methods described so far update the image at each iteration using either the original or a surrogate cost function which is computed from the entire data set. Methods based on the use of subsets of the data can produce further increases in the convergence rate, particularly at the earlier iterations. The best known among these is the ordered subsets EM (OSEM) algorithm, which is now widely used in clinical applications. While the original OSEM algorithm (Hudson and Larkin 1994) and earlier methods employing subsets of the data (Hebert *et al* 1990, Holte *et al* 1990) are not convergent, convergent forms have since been developed as we describe below.

Ordered subsets (OS) methods divide the objective function into a sum of sub-objective functions as

$$\Phi(\boldsymbol{f}) = \sum_{q=1}^{N_s} \Phi_q(\boldsymbol{f}), \tag{49}$$

where each sub-objective function $\Phi_q(\boldsymbol{f})$ is a function of a subset of the data. We use $\{S_q\} \subseteq \{1, 2, \ldots, M\}$ to denote the collection of the indices of the projections in the qth subset of the data. The subsets are defined so that $\{1, 2, \ldots, M\} = \bigcup_{q=1}^{N_s} S_i$, where N_s is the number of subsets, and $S_i \cap S_j = \emptyset$ for $i \neq j$. At each iteration, only one $\Phi_q(\boldsymbol{f})$ is involved in the calculation. Let k denote the index for a complete cycle and i the index for a sub-iteration, and define $f^{(k,0)} = f^{(k-1)}$, $f^{(k,N_s)} = f^{(k)}$.

The OSEM algorithm applies the EM algorithm to each sub-objective function $\Phi_q(\boldsymbol{f})$ in turn. The update equation is given by

$$f_j^{(k,q)} = \frac{f_j^{(k,q-1)}}{\sum_{i\in S_q} p_{ij}} \sum_{i\in S_q} \frac{p_{ij} y_i}{\sum_l p_{il} f_l^{(k,q-1)} + r_i + s_i}, \quad \text{for} \quad j = 1, \ldots, N, \quad q = 1, \ldots, N_s. \tag{50}$$

To avoid directional artefacts, the projections are usually chosen to have maximum separation in angle in each subset. Hudson and Larkin (1994) further recommend that the subsets are chosen so that an emission from each pixel has equal probability of being detected in each of the subsets (subset balance), i.e. $\sum_{i\in S_q} p_{ij}$ is independent on q. In practice, this can be difficult to achieve due to spatially varying attenuation and detector sensitivities.

The requirement of subset balance can be alleviated by reformulating the OSEM algorithm in a gradient-ascent form and replacing $\sum_{i\in S_q} p_{ij}$ by a voxel-independent scaling factor $t_q = \max_j\{\sum_{i\in S_q} p_{ij}\}$. The resulting rescaled block iterative (RBI) EM algorithm (Byrne 1998) has update equation

$$f_j^{(k,q)} = f_j^{(k,q-1)} + t_q^{-1} f_j^{(k,q-1)} \sum_{i\in S_q} \left(\frac{p_{ij} y_i}{\sum_l p_{il} f_l^{(k,q-1)} + r_i + s_i} - 1 \right), \quad \text{for} \quad j = 1, \ldots, N, \quad q = 1, \ldots, N_s. \tag{51}$$

The difference between (51) and (50) is that the scaling factor t_q in the RBI-EM algorithm depends only on the subset index q rather than the voxel index j. This rescaling makes the RBI-EM algorithm less susceptible to unbalanced subsets.

OSEM produces remarkable improvements in convergence rates in the early iterations. For a reasonable number of subsets (e.g., 32 in Hudson and Larkin (1994)), almost linear speed-up with respect to the number of subsets can be achieved in the early iterations. For this reason, OSEM is by far the most widely used iterative ECT algorithm in clinical and research PET and SPECT studies. In the case of 3D PET, the data are usually first Fourier rebinned into stacked 2D sinograms before application of a 2D OSEM algorithm (Comtat *et al* 1998, Lee *et al* 2004).

The major drawback of the OSEM and RBI-EM algorithms is that they generally do not converge if the number of subsets remains greater than 1. Reducing the number of sets as the iterations proceed will resolve this problem but will result in the asymptotic convergence rate of the original EM algorithm. In the special case of consistent data, i.e. when there exists an image whose forward projection exactly matches the measurements, the OSEM algorithm under subset balance and the RBI-EM algorithm have both been shown to converge to the ML solution. For inconsistent data, OSEM has been observed to enter a limit cycle condition when the number of subsets remains greater than 1 (Byrne 1997), but currently there has been no proof of this behaviour. Both OSEM and RBI-EM can be extended to speed up the GEM algorithms (Hudson and Larkin 1994, Lalush and Tsui 1998) but again, in general these methods will not converge.

6.4. Making ordered subset methods convergent

Although OSEM does not converge, it is currently the most widely used iterative method for statistically based reconstruction in emission tomography. This is primarily due to the resulting improvements in image quality, compared to standard filtered backprojection methods, coupled with its relatively low computational cost. This has spurred interest in searching for convergent ordered subsets methods. These convergent methods fall into two categories: relaxed OS methods and incremental OS methods.

Relaxed OS methods. This group of methods uses a subset-independent step size and gradually decreases the step size to zero. One example of this approach is the row-action maximum-likelihood algorithm (RAMLA) of Browne and De Pierro (1996), which was motivated by the row-action ART algorithms developed in the 1970s (Censor 1983). The update equation is

$$f_j^{(k,q)} = f_j^{(k,q-1)} + \eta_k f_j^{(k,q-1)} \sum_{i \in S_q} \left(\frac{p_{ij} y_i}{\sum_l p_{il} f_l^{(k,q-1)} + r_i + s_i} - 1 \right), \quad \text{for} \quad j = 1, \ldots, N, \quad q = 1, \ldots, N_s \tag{52}$$

where η_k is a sequence of positive relaxation parameters such that

$$\lim_{k \to \infty} \eta_k = 0, \qquad \sum_{k=0}^{\infty} \eta_k = +\infty. \tag{53}$$

Note that in (52), η_k is fixed throughout a complete cycle through the subsets. This subset-independent step size is critical for convergence, although it has been shown recently that convergence can also be achieved with step sizes that are asymptotically subset independent (Helou Neto and De Pierro 2005). Browne and De Pierro (1996) proved that unlike the OSEM algorithm, RAMLA in (52) converges to an ML solution if the log-likelihood function is *strictly* concave. However, this guarantee of convergence is achieved at the price of slower

convergence due to underrelaxation. In fact, its final convergence rate can be even slower than that of the EM algorithm.

Recently, De Pierro and Yamagishi (2001) extended the RAMLA method to MAP reconstruction. The method is called block sequential regularized expectation maximization (BSREM). The prior term is treated as another subset of the objective function so that one iteration of the algorithm consists of an ML RAMLA cycle using (52) followed by an update in the gradient direction of the prior. The treatment of the prior as an additional subset added to the likelihood function makes the proof of convergence of ML RAMLA directly extendible to BSREM. However, an assumption that the sequence $\{\boldsymbol{f}^k\}$ is positive and bounded is required because of the addition of the prior term. This assumption is not automatically ensured in the BSREM update. This problem was solved by Ahn and Fessler in (2001, 2003) in which a projection operator is used to impose an upper bound and the likelihood function slightly modified for $\bar{y}_i$ close to zero, so that the sequence is guaranteed to satisfy non-negativity and boundedness.

A general form for the relaxed OS algorithm can be written as (Ahn and Fessler 2003)

$$f_j^{(k,q)} = f_j^{(k,q-1)} + \eta_k f_j^{(k,q-1)} \nabla \Phi_q(\boldsymbol{f}^{(k,q-1)}), \quad \text{for} \quad j = 1, \ldots, N, \quad q = 1, \ldots, N_s \tag{54}$$

where $\Phi_q(\boldsymbol{f})$ is the objective function for the qth subset and η_k satisfies (53). An open problem in the use of relaxed OS algorithms is to find a sequence η_k that maximizes the convergence rate while satisfying the convergence condition (53).

Incremental OS methods. Incremental OS methods achieve convergence through the introduction of an augmented cost function (Ahn 2004). Consider a surrogate $\phi_q(\boldsymbol{f}; \bar{\boldsymbol{f}})$ for the sub-objective function $\Phi_q(\boldsymbol{f})$ that satisfies

$$\phi_q(\boldsymbol{f}; \boldsymbol{f}) = \Phi_q(\boldsymbol{f}), \qquad \forall \boldsymbol{f} \tag{55}$$

$$\phi_q(\boldsymbol{f}; \bar{\boldsymbol{f}}) \leqslant \Phi_q(\boldsymbol{f}), \qquad \forall \boldsymbol{f}, \bar{\boldsymbol{f}}. \tag{56}$$

Then for all $\bar{\boldsymbol{f}}$, $\max_{\boldsymbol{f}} \phi_q(\boldsymbol{f}; \bar{\boldsymbol{f}}) = \phi_q(\boldsymbol{f}; \boldsymbol{f}) = \Phi_q(\boldsymbol{f})$. It follows that if we construct an augmented cost function as

$$F(\boldsymbol{f}; \bar{\boldsymbol{f}}_1, \ldots, \bar{\boldsymbol{f}}_{N_s}) = \sum_{q=1}^{N_s} \phi_q(\boldsymbol{f}; \bar{\boldsymbol{f}}_q) \tag{57}$$

then the maximizer of $\Phi(\boldsymbol{f})$ can be found as

$$\hat{\boldsymbol{f}} = \arg\max_{\boldsymbol{f}} \Phi(\boldsymbol{f}) = \arg\max_{\boldsymbol{f}} \max_{\bar{\boldsymbol{f}}_1, \ldots, \bar{\boldsymbol{f}}_{N_s}} F\left(\boldsymbol{f}; \bar{\boldsymbol{f}}_1, \ldots, \bar{\boldsymbol{f}}_{N_s}\right). \tag{58}$$

Incremental OS methods alternately update $\boldsymbol{f}$ and one $\bar{\boldsymbol{f}}_q$ at each step, i.e.,

$$\boldsymbol{f}^{(k+1,q)} = \arg\max_{\boldsymbol{f}} F\left(\boldsymbol{f}; \bar{\boldsymbol{f}}_1^{(k+1)}, \ldots, \bar{\boldsymbol{f}}_{q-1}^{(k+1)}, \bar{\boldsymbol{f}}_q^{(k)}, \ldots, \bar{\boldsymbol{f}}_{N_s}^{(k)}\right) \tag{59}$$

$$\begin{aligned} &\bar{\boldsymbol{f}}_q^{(k+1)} = \boldsymbol{f}^{(k+1,q)} \\ &\text{for } q = 1, \ldots, N_s. \end{aligned} \tag{60}$$

Using the EM surrogate function in (32) for each subset of the data, i.e.,

$$\phi_q(\boldsymbol{f}; \bar{\boldsymbol{f}}) = \sum_j \left(\bar{f}_j \sum_{i \in S_q} \frac{p_{ij} y_i}{\sum_l p_{il} \bar{f}_l + r_i + s_i} \log(p_{ij} f_j) - f_j \sum_{i \in S_q} p_{ij} \right), \tag{61}$$

we obtain the update equation of the convergent OSEM algorithm (COSEM) (Hsiao *et al* 2004, Khurd *et al* 2004),

$$f_j^{(k+1,q)} = \frac{1}{\sum_i p_{ij}} \sum_{q'=1}^{N_s} e_{q',j}, \tag{62}$$

where

$$e_{q',j} = \begin{cases} \sum_{i \in S_q} \dfrac{p_{ij} y_i}{\sum_l p_{il} f_l^{(k+1,q')} + r_i + s_i}, & q' < q \\ \sum_{i \in S_q} \dfrac{p_{ij} y_i}{\sum_l p_{il} f_l^{(k,q')} + r_i + s_i}, & q' \geqslant q. \end{cases} \tag{63}$$

Note that in emission tomography, the dominant cost of forward- and backprojection is associated with the update of the surrogate functions $\phi_q(\boldsymbol{f}; \bar{\boldsymbol{f}}_q)$ as a result of changes in $\bar{\boldsymbol{f}}_q$. In COSEM this corresponds to the evaluation of $e_{q',j}$; however, this quantity remains unchanged at sub-iteration q for all $q' \neq q$. Therefore, by updating only one sub-objective function $\phi_q(\boldsymbol{f}; \bar{\boldsymbol{f}}_q)$ at a time, incremental OS methods reduce the computation cost by a factor similar to conventional OS methods. Unfortunately, incremental OS methods, while faster than their non-OS counterparts, are slower than the conventional OS methods at early iterations. Faster convergence can therefore be achieved by starting with a conventional OS algorithm, but switching to a convergent incremental OS algorithm at later iterations (Hsiao *et al* 2004, Ahn 2004). This can be taken a step further by noting that, asymptotically, preconditioned conjugate-gradient (PCG) methods converge faster than the OS methods and consequently one could switch from incremental OS to PCG at later iterations (Li *et al* 2005).

As a final note on these algorithms, speed-up using OS method is most significant at early iterations. Asymptotic convergence is only guaranteed for certain relaxed or incremental OS algorithms. Even then, none of the existing OS algorithms guarantees monotonic increase of the objective function.

6.5. *Methods for nonconvex priors*

For nonconvex priors, the log posterior may have multiple local maxima, which makes the optimization far more difficult. All the algorithms described above converge, at best, to a local maximum. When applied to nonconvex cost functions, the result is dependent on the initial image and the choice of optimization algorithm. Stochastic searches employing simulated annealing can guarantee convergence but are impractical for routine use in image reconstruction. A compromise that may help to avoid local maxima, but which in general does not guarantee convergence to a global maximum, is to use a deterministic annealing scheme. These methods, which are also referred to as graduated nonconvexity approaches, approximate the nonconvex prior with a sequence of smooth functions that converges uniformly to the original function. One or more iterations of one of the algorithms described above can be applied to each function in this sequence with the goal of avoiding local maxima of the original function. Although, in general, no global convergence properties are known, favourable behaviour has been observed by several authors (Geiger and Girosi 1991, Gindi *et al* 1993, Mumcuoglu *et al* 1994, Yu and Fessler 2002).

7. Numerical examples

We conclude with a few illustrations based on a small scale problem in which we estimate the intensities of two voxels from three measurements. By restricting the image space to

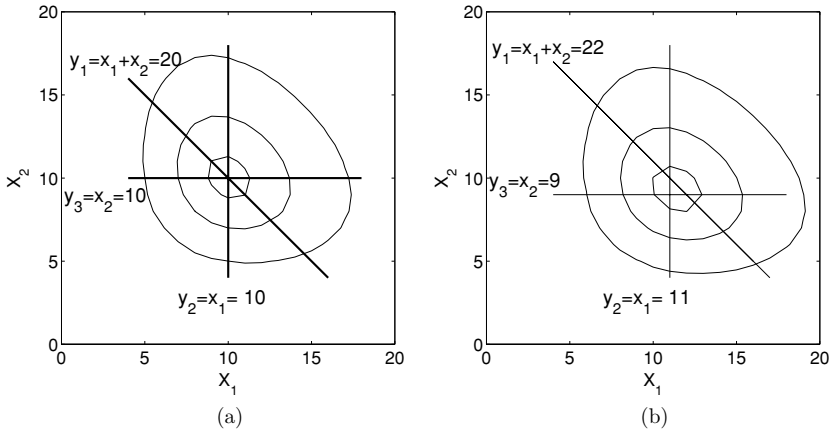

Figure 2. Illustration of a well-conditioned problem with system matrix $\boldsymbol{P}$ given in (64) with source $\boldsymbol{x} = [10, 10]^T$. (a) The noiseless case: $\boldsymbol{y} = E(\boldsymbol{y}) = [10, 10, 10]^T$ where the ML solution is $[10, 10]^T$; (b) the noisy case: $\boldsymbol{y} = [11, 11, 9]^T$ and the ML solution is $[11.4, 9.3]^T$. Shown in each case are isocontours of the Poisson log-likelihood function $L(\mathbf{y}|\boldsymbol{x})$ and lines corresponding to solutions of each of the three equations $y_i = P_i\boldsymbol{x}$, $i = 1, 2, 3$.

two dimensions, we can illustrate the shape of the cost function for well- and ill-conditioned problems and plot the trajectories of the different algorithms through this space.

Figure 2 illustrates a well-conditioned problem where the projection matrix is

$$\boldsymbol{P} = \begin{bmatrix} 0.5 & 0.5 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}, \tag{64}$$

and the mean emission rates of the two source voxels are $\boldsymbol{x} = [10, 10]^T$. Shown in each figure are the lines corresponding to solutions of each of the three equations $y_i = P_i\boldsymbol{x}$ $i = 1, 2, 3$, where P_i represents the ith row of $\boldsymbol{P}$, and the isocontours of the Poisson log-likelihood function $L(\boldsymbol{y}|\boldsymbol{x})$. The well-conditioning is reflected in the near circular isocontours of the likelihood function. In the noiseless case (figure 2(a)) the equations are consistent so that the three lines intersect, and the maximum-likelihood solution coincides with the exact solution to the three equations $\boldsymbol{y} = \boldsymbol{P}\boldsymbol{x}$. In practice, the photon-limited data are modelled as Poisson random variables and an exact solution of $\boldsymbol{y} = \boldsymbol{P}\boldsymbol{x}$ will not exist in general. This case is illustrated on the right of figure 2, where now the three lines $y_i = P_i\boldsymbol{x}, i = 1, 2, 3$, do not intersect. However, the log likelihood still has a unique maximum, which corresponds to the ML solution and lies close to the three points of pairwise intersection of the three equations.

Figure 3 shows an ill-conditioned problem, which is often the case in emission tomography. The projection matrix $\boldsymbol{P}$ is

$$\boldsymbol{P} = \begin{bmatrix} 0.5 & 0.5 \\ 0.6 & 0.4 \\ 0.4 & 0.6 \end{bmatrix}, \tag{65}$$

and the source image is again $\boldsymbol{x} = [10, 10]^T$. In this case the ill-conditioning of $\boldsymbol{P}$ is reflected in the small angles between each of the lines $y_i = P_i\boldsymbol{x}$. Also note that the isocontours of the likelihood function are elongated, indicating high variance in the ML solution in the direction of the major axis of these approximate ellipsoids; i.e., small changes in the data can produce large changes in the solution along this direction. This is illustrated in the right of the figure

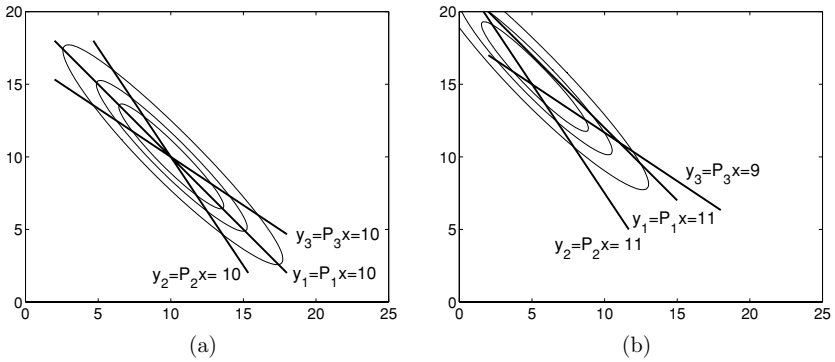

Figure 3. Illustration of the case where $\boldsymbol{P}$ is the ill-conditioned system matrix in (65). (a) The noiseless case: $\boldsymbol{y} = E(\boldsymbol{y}) = [10, 10, 10]^T$ and again the ML solution is $[10, 10]^T$; (b) the noisy case: $\boldsymbol{y} = [11, 11, 9]^T$ and the ML solution is $[5.2, 15.5]^T$.

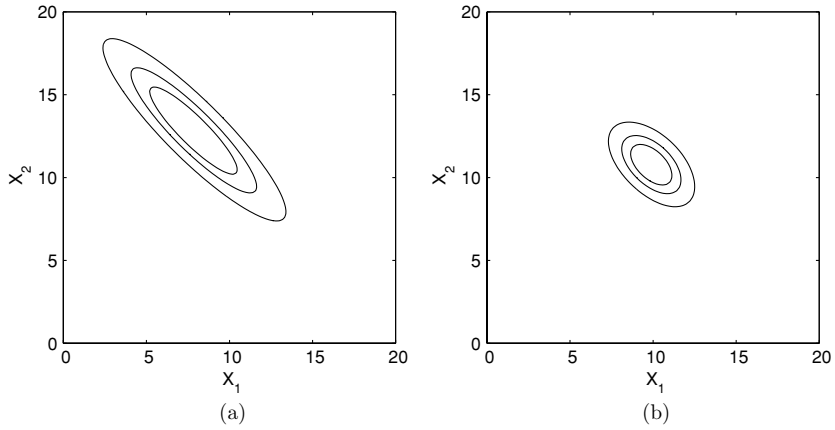

Figure 4. Illustration of the effect on the cost function of adding the prior smoothing term $\beta(x_1 - x_2)^2$ to the log-likelihood function for the ill-conditioned problem shown in figure 3(b). Shown are the isocontours of $\Phi(\boldsymbol{y}, \boldsymbol{x})$ (equation (24)) for the cases (a) $\beta = 0.001$ and (b) $\beta = 0.1$.

where noisy data result in a substantial shift of the maximum of the likelihood function relative to the noiseless case, where the maximum-likelihood solution coincides with the exact solution to the three consistent equations. Also shown in the right figure are the set of solutions to each of the equations $y_i = P_i \boldsymbol{x}$. Note that the distances between points in the set of pairwise intersections of these lines is larger than that for the well-conditioned case, again indicating the increased sensitivity to noise of the ill-conditioned system.

In figure 4 we illustrate the effect of adding the smoothing term $\beta(x_1 - x_2)^2$ to the log-likelihood function of the ill-conditioned problem in figure 3. This smoothing term has the form of the Gibbs energy function defined in (22) with a quadratic potential function. The regularizing effect of this prior is manifested through the reduced eccentricity of the ellipsoidal isocontours. As the influence of the prior is increased, through increasing β, the isocontours become increasingly circular, resulting in improved stability of the estimate formed by maximizing the posterior density. The ML solution for this ill-conditioned example gives the estimate $[5.2, 15.5]^T$ while the MAP solutions are $= [7.8, 12.8]^T$ for $\beta = 0.001$

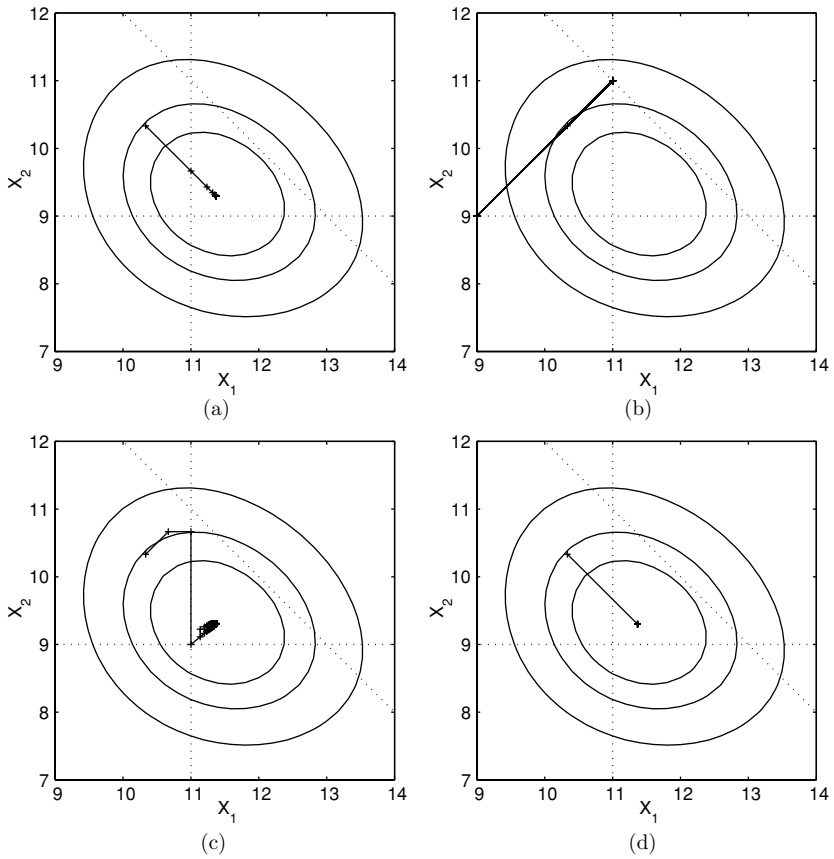

Figure 5. Trajectories representing the sequence of estimates of the image for (a) EM (b) OSEM (c) RAMLA (d) PCG (e) ICA for the well-conditioned ML problem in figure 2. (f)–(g) The convergence behaviour showing log likelihood versus iteration number for each algorithm. Note that OSEM with three subsets (one sample per subset) quickly enters a limit cycle ($[11, 11]^T \rightarrow [11, 11]^T \rightarrow [9, 9]^T$) and does not converge. EM converges, but slowly. RAMLA initially converges rapidly but slows down due to underrelaxation.

and $= [9.9, 10.8]^T$ for $\beta = 0.1$. Since the true source values are $[10, 10]$ we see that the MAP solution does indeed produce an improved estimate compared to ML for this case. A Monte Carlo simulation would also show that the variance of the ML estimate is considerably higher than that of the MAP solution. However, it is important to note that this prior uses the assumption that the two voxel values are similar. In cases where there is actually a large difference between the two values, this prior would tend to introduce a substantial bias into the estimate.

We now turn to the behaviour of the algorithms themselves. The convergence behaviour of the algorithms is largely determined by the shape of the objective function, so that for those algorithms that can be applied to both ML and MAP estimation, we expect similar behaviour for well-conditioned ML estimates as for MAP estimates. Consequently, we illustrate the behaviour of all methods for ML only, but for both the well-conditioned and ill-conditioned cases. MAP algorithms are applied to these data by setting $\beta = 0$.

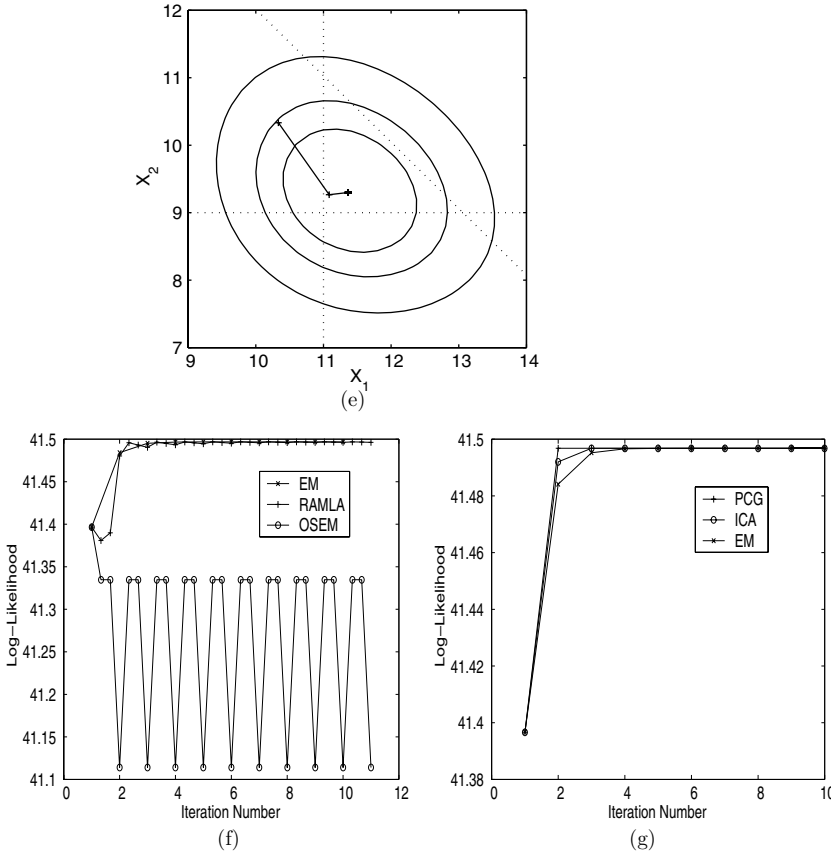

Figure 5. (Continued.)

In figure 5 we show the convergence behaviours of several algorithms for the well-conditioned problem of figure 2(b). To better visualize this behaviour we show a smaller region of the search space than in figure 2(b). All algorithms were initialized with a uniform estimate $[10.3, 10.3]^T$. Because the objective function is well-conditioned, EM, PCG, and ICA all effectively converge within a few iterations. OSEM with three subsets enters a limit cycle after the first iteration as can be seen in the likelihood versus iteration plotted in figure 5(f). This particularly poor performance of OSEM is partly because the subsets are unbalanced (they cannot be balanced for this example). RAMLA, which uses a subset-independent scaling and underrelaxation (here we use $\lambda_k = 1/\sqrt{k}$), avoids the need for subset balance and does exhibit convergent behaviour, but also shows the non-monotonic behaviour typical of OS methods. The relative performance of PCG, ICA and EM shown in figure 5(g) is illustrative of larger scale problems. Typically both PCG and ICA produce substantially faster convergence than EM; relative convergence rates of PCG and ICA are highly dependent on the specific implementation and choice of preconditioner.

In figure 6 we show the behaviour of the same algorithms for an ill-conditioned problem. To make the problem more realistic, and challenging, we also include an active non-negativity constraint; i.e., the unconstrained maximum of the log likelihood contains a negative value for voxel x_2. The non-negativity constraint requires that we find the maximum of the log likelihood

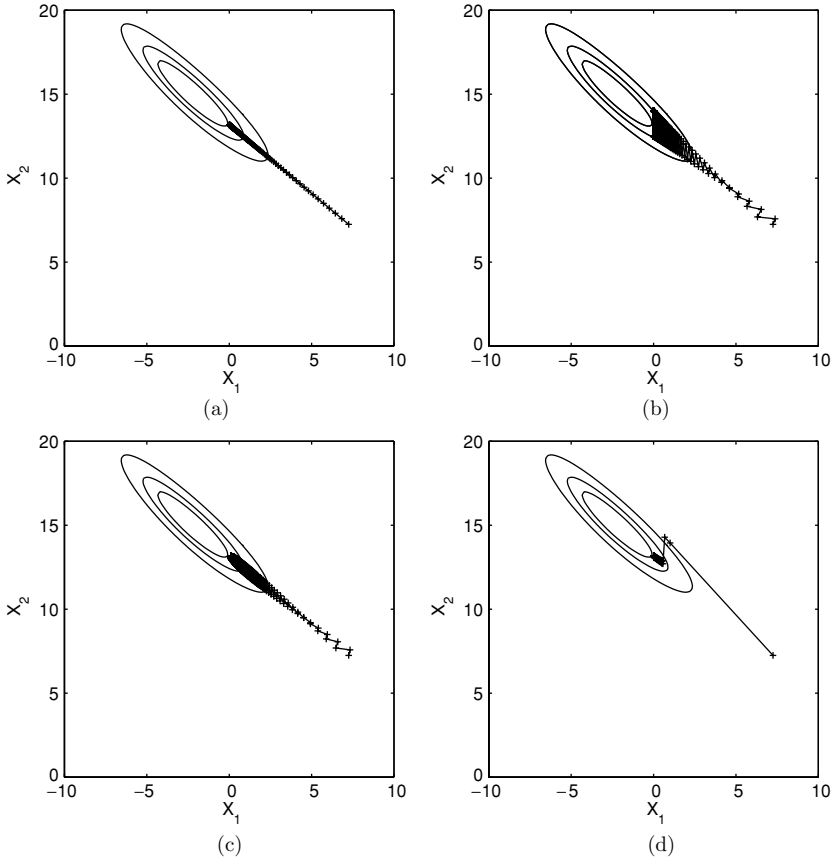

Figure 6. Trajectories representing the sequence of estimates of the image for (a) EM (b) OSEM (c) RAMLA (d) PCG (e) ICA for an ill-conditioned ML problem with an active non-negativity constraint in voxel x_2. (f)–(g) The convergence behaviour showing log likelihood versus iteration number for each algorithm. In this case OSEM and RAMLA use two balanced subsets. Again, OSEM reaches a limit cycle while the convergence rate of RAMLA slows as the relaxation parameter decreases. The convergence rate of ICA may be improved using a better initial estimate.

within the set of non-negative voxels only. The projection matrix $\boldsymbol{P}$ for this example is

$$\boldsymbol{P} = \begin{bmatrix} 0.5 & 0.5 \\ 0.6 & 0.4 \\ 0.4 & 0.6 \\ 0.3 & 0.7 \end{bmatrix}, \tag{66}$$

and the noisy data are $\boldsymbol{y} = [7, 10, 4, 8]^T$. One observation common to all methods is that because the objective function is ill-conditioned, they take many more iterations to convergence (or to enter a limit cycle). In this case EM converges slowly as does ICA. The PCG converges rapidly, although this is in part because there are only two unknowns and conjugate-gradient methods converge rapidly for low-dimensional problems. For the OS algorithms, we used two balanced subsets. Both OSEM and RAMLA show their characteristic faster convergence compared to EM in the earlier iterations. However, OSEM eventually enters a limit cycle while RAMLA does converge, but increasingly slowly due to underrelaxation. We note that situation here may not represent the best performance of ICA since when applied to real ECT

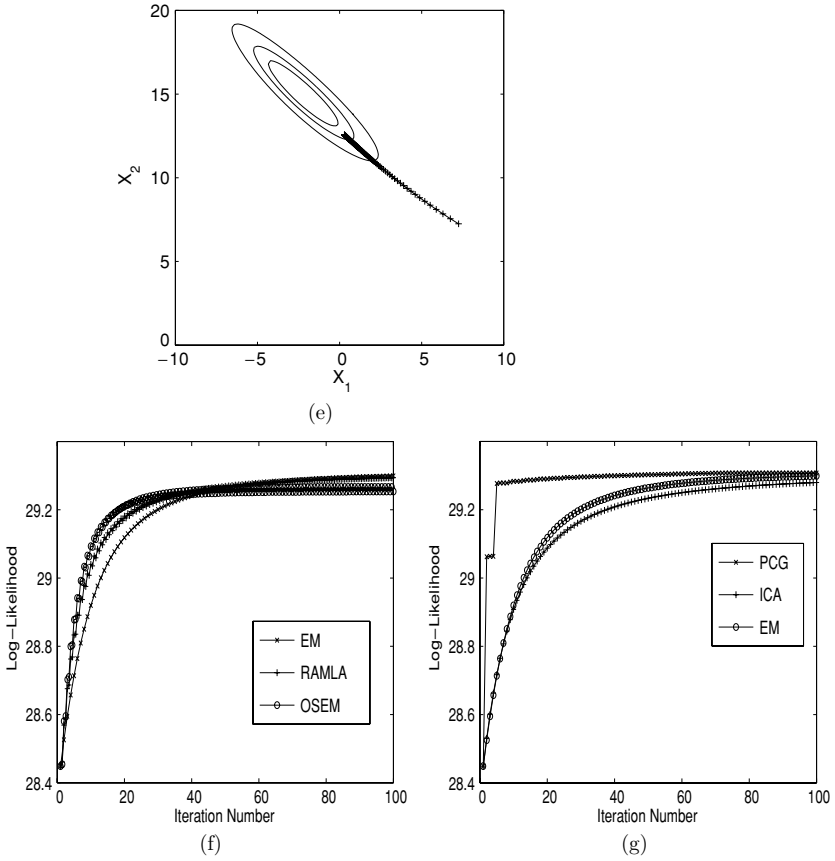

Figure 6. (Continued.)

problems, ICA is typically initialized with a FBP image while here we start with a uniform image. Empirically, the convergence rates of PCG and ICA algorithms are comparable and both are significantly faster than EM. The convergence rate of RAMLA (and other convergent OS methods) heavily depends on the underrelaxation sequence. As noted above, selection of optimal relaxation schedules is an open problem.

8. Conclusions

We have provided a review of the many approaches to iterative image reconstruction for emission tomography. These methods differ in two distinct ways: the objective function used and the trajectory through the image space taken by the algorithm to maximize the objective function. While our preference is towards the use of fast convergent algorithms for MAP or penalized ML estimation, it is clear that although the OSEM ML method is nonconvergent and unregularized, it still produces satisfactory results in practice. It is also the case that there are far bigger differences between analytic reconstructions based on filtered backprojection and those based on ML or MAP principles than there are between the different ML and MAP solutions. However, as this field continues to mature and the cost of convergent regularized methods approaches that of OSEM, the differences between these methods may warrant the broader adoption of convergent and regularized methods in practical applications.

Acknowledgments

This work was supported in part by the National Institutes of Health under grants nos R01 EB00393 and R01 EB00194.

References

Ahn S 2004 Convergent algorithms for statistical image reconstruction in emission tomography *PhD Thesis* University of Michigan

Ahn S and Fessler J A 2001 Globally convergent ordered subsets algorithms: application to tomography *Proc. IEEE NSS-MIC* pp 1064–8

Ahn S and Fessler J A 2003 Globally convergent image reconstruction for emission tomography using relaxed ordered subsets algorithms *IEEE Trans. Med. Imaging* **22** 613–26

Ahn S and Fessler J A 2004 Emission image reconstruction for randoms-precorrected pet allowing negative sinogram values *IEEE Trans. Med. Imaging* **23** 591–601

Anderson J M M, Mair B A, Rao M and Wu C-H 1997 Weighted least-squares reconstruction methods for positron emission tomography *IEEE Trans. Med. Imaging* **16** 159–65

Alenius S and Ruotsalainen U 1997 Bayesian image reconstruction for emission tomography based on median root prior *Eur. J. Nucl. Med.* **24** 258–65

Alenius S and Ruotsalainen U 2002 Generalization of median root prior reconstruction *IEEE Trans. Med. Imaging* **21** 1413–20

Alenius S, Ruotsalainen U and Astola J 1998 Using local median as the location of the prior distribution in iterative emission tomography image reconstruction *IEEE Trans. Nucl. Sci.* **45** 3097–104

Alessio A, Sauer K and Bouman C A 2003 MAP reconstruction from spatially correlated PET data *IEEE Trans. Nucl. Sci.* **50** 1445–51

Bailey D L and Meikle S R 1994 A convolution-subtraction scatter correction method for 3D PET *Phys. Med. Biol.* **39** 411–24

Barrett H H and Myers K J 2003 *Foundations of Image Science* (New York: Wiley)

Barrett H H, White T and Parra L 1997 List-mode likelihood *J. Opt. Soc. Am.* A **14** 2914–23

Beekman F J, de Jong H W A M and van Geloven S 2002 Efficient fully 3-D iterative SPECT reconstruction with Monte Carlo-based scatter compensation *IEEE Trans. Med. Imaging* **21** 867–77

Bendriem B and Townsend D W (ed) 1998 *The Theory and Practice of 3D PET* (Dordrecht: Kluwer)

Bertsekas D P 1997 A hybrid incremental gradient method for least squares *SIAM J. Optim.* **7** 913–26

Besag J 1974 Statistical analysis of lattice systems *J. R. Stat. Soc.* B **36** 192–236

Bhatia M, Karl W C and Willsky A S 1997 Tomographic reconstruction and estimation based on multiscale natural-pixel bases *IEEE Trans. Image Process.* **6** 463–78

Bouman C and Sauer K 1996 A unified approach to statistical tomography using coordinate descent optimization *IEEE Trans. Image Process.* **5** 480–92

Bowsher J E, Johnson V E, Turkington T G, Jaszczak R J, Floyd C E and Coleman R E 1996 Bayesian reconstruction and use of anatomical a priori information for emission tomography *IEEE Trans. Med. Imaging* **15** 673–86

Browne J and De Pierro A R 1996 A row-action alternative to the EM algorithm for maximizing likelihoods in emission tomography *IEEE Trans. Med. Imaging* **15** 687–99

Budinger T F and Gullberg G T 1974 Three-dimensional reconstruction in nuclear medicine emission imaging *IEEE Trans. Nucl. Sci.* **21** 2–20

Byrne C 1993 Iterative image reconstruction algorithms based on cross-entropy minimization *IEEE Trans. Image Process.* **2** 96–103

Byrne C 1995 Erratum and addendum to 'iterative image reconstruction algorithms based on cross-entropy minimization' *IEEE Trans. Image Process.* **4** 225–6

Byrne C 2001 Likelihood maximization for list-mode emission tomographic image reconstruction *IEEE Trans. Med. Imaging* **20** 1084–92

Byrne C L 1997 Convergent block-iterative algorithms for image reconstruction from inconsistent data *IEEE Trans. Image Process.* **6** 1296–304

Byrne C L 1998 Accelerating the EMML algorithm and related iterative algorithms by rescaled block-iterative (RBI) methods *IEEE Trans. Image Process.* **7** 100–9

Casey M E and Hoffman E J 1986 Quantitation in positron emission computed tomography: 7. a technique to reduce noise in accidental coincidence measurements and coincidence efficiency calibration *J. Comput. Assist. Tomogr.* **10** 845–50

Censor Y 1983 Finite series-expansion reconstruction methods *IEEE Proc.* **71** 409–18
Censor Y and Zenios S A 1997 *Parallel Optimization: Theory, Algorithms, and Applications* (Oxford: Oxford University Press)
Chan M, Herman G T and Levitan E 1995 Bayesian image reconstruction using a high-order interacting MRF model *Proc. 8th Int. Conf. on Image Analysis and Processing (San Remo, Italy)* (Berlin: Springer) pp 609–14
Chang L T 1978 A method for attenuation correction in radionuclide computed tomography *IEEE Trans. Nucl. Sci.* **25** 638–43
Chang J H, Anderson J M and Votaw J R 2004 Regularized image reconstruction algorithms for positron emission tomography *IEEE Trans. Med. Imaging* **23** 1165–75
Chen Y, Furenlid L R, Wilson D W and Barrett H H 2004 Measurement and interpolation of the system matrix for pinhole SPECT: comparison between MLEM and ART reconstructions *Proc. IEEE NSS-MIC* M5-306
Cherry S R and Dahlbom M 2004 PET: physics, instrumentation and scanners *PET: Molecular Imaging and Its Biological Applications* 1st edn, ed M E Phelps (Berlin: Springer) pp 1–124
Cherry S R and Phelps M E 1996 Imaging brain function with positron emission tomography *Brain Mapping: The Methods* ed A W Toga and J C Mazziotta (New York: Academic) chapter 8
Chinn G and Huang S C 1997 A general class of preconditioners for statistical iterative reconstruction of emission computed tomography *IEEE Trans. Med. Imaging* **16** 1–10
Clinthorne N H *et al* 1993 Preconditioning methods for improved convergence rates in iterative reconstruction *IEEE Trans. Med. Imaging* **12** 78–83
Coakley K J 1991 A cross-validation procedure for stopping the EM algorithm and deconvolution of neutron depth profiling spectra *IEEE Trans. Nucl. Sci.* **38** 9–15
Colsher J G 1980 Fully three-dimensional positron emission tomography *Phys. Med. Biol.* **25** 103–15
Comtat C, Kinahan P E, Defrise M, Michel C and Townsend D W 1998 Fast reconstruction of 3D PET data with accurate statistical modeling *IEEE Trans. Nucl. Sci.* **45** 1083–9
Csiszar I 1991 Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems *Ann. Stat.* **19** 2032–66
Csiszar I and Tusnady G 1984 Information geometry and alternating minimization procedures *Stat. Decis. Suppl.* 205–37
Daube-Witherspoon M E and Muehllehner G 1987 Treatment of axial data in three-dimensional PET *J. Nucl. Med.* **28** 1717–24
Defrise M, Kinahan P E, Townsend D W, Michel C, Sibomana M and Newport D F 1997 Exact and approximate rebinning algorithms for 3-D PET data *IEEE Trans. Med. Imaging* **16** 145–58
Dempster A P, Laird N M and Rubin D B 1977 Maximum likelihood from incomplete data via the EM algorithm *J. R. Stat. Soc. Ser.* B **39** 1–38
De Pierro A R 1990 Multiplicative iterative methods in computed tomography *Mathematical Methods in Tomography* ed G T Herman, A K Louis and F Natterer (Berlin: Springer) pp 167–86
De Pierro A R 1993 On the relation between the ISRA and the EM algorithm for positron emission tomography *IEEE Trans. Med. Imaging* **12** 328–33
Derenzo S E 1986 Mathematical removal of positron range blurring in high resolution tomography *IEEE Trans. Nucl. Sci.* **33** 565–9
De Pierro A R 1995 A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography *IEEE Trans. Med. Imaging* **14** 132–7
De Pierro A R and Yamagishi M E B 2001 Fast EM-like methods for maximum a posteriori estimates in emission tomography *IEEE Trans. Med. Imaging* **20** 280–8
Edholm P R, Lewitt R M and Lindholm B 1986 Novel properties of the Fourier decomposition of the sinogram *Proc. SPIE* **671** 8–18
Feng H H, Karl W C and Castanon D A 2003 A curve evolution approach to object-based tomographic reconstruction *IEEE Trans. Image Process.* **12** 44–57
Fessler J A 1994 Penalized weighted least-squares image reconstruction for PET *IEEE Trans. Med. Imaging* **13** 290–300
Fessler J A 1996 Mean and variance of implicitly defined biased estimators (such as penalized maximum likelihood): applications to tomography *IEEE Trans. Image Process.* **5** 493–506
Fessler J A and Booth S D 1999 Conjugate-gradient preconditioning methods for shift-variant image reconstruction *IEEE Trans. Image Process.* **8** 688–99
Fessler J A and Erdogan H 1998 A paraboloidal surrogates algorithm for convergent penalized-likelihood emission image reconstruction *Proc. IEEE NSS-MIC (Toronto, ON, Canada, Nov.)* vol 2 (Piscataway, NJ: IEEE) pp 1132–35

Fessler J A, Ficaro E P, Clinthorne N H and Lange K 1997 Grouped-coordinate ascent algorithms for penalized-likelihood transmission image reconstruction *IEEE Trans. Med. Imaging* **16** 166–75

Fessler J A and Hero A O 1995 Penalized maximum-likelihood image reconstruction using space-alternating generalized EM algorithms *IEEE Trans. Image Process.* **4** 1417–29

Fessler J A and Rogers W L 1996 Spatial resolution properties of penalized-likelihood image reconstruction: spatial-invariant tomographs *IEEE Trans. Image Process.* **9** 1346–58

Floyd C S and Jaszczak R J 1985 Inverse Monte Carlo: A unified reconstruction algorithm for SPECT *IEEE Trans. Nucl. Sci.* **32** 779–85

Formiconi A R, Pupi A and Passeri A 1989 Compensation of spatial system response in SPECT with conjugate gradient reconstruction technique *Phys. Med. Biol.* **34** 69–84

Frese T, Bouman C A and Sauer K 2002 Adaptive wavelet graph model for Bayesian tomographic reconstruction *IEEE Trans. Image Process.* **11** 756–70

Geiger D and Girosi F 1991 Parallel and deterministic algorithms from MRF's: surface reconstruction *IEEE Trans. Pattern Anal. Mach. Intell.* **13** 401–12

Geman S and Geman D 1984 Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images *IEEE Trans. Pattern Anal. Mach. Intell.* **6** 721–41

Geman S and McClure D E 1985 Bayesian image analysis: an application to single photon emission tomography *Proc. Statistical Computing Section of the American Statistical Association* pp 12–8

Geman D and Reynolds G 1992 Constrained restoration and the recovery of discontinuities *IEEE Trans. Pattern Anal. Mach. Intell.* **14** 367–83

Geman D and Yang C D 1995 Nonlinear image recovery with half-quadratic regularization *IEEE Trans. Image Process.* **4** 932–46

Gindi G, Lee M, Rangarajan A and Zubal I G 1991 Bayesian reconstruction of functional images using registered anatomical images as priors *Information Processing in Medical Imaging, 12th IPMI Int. Conf. (Wye, UK)* pp 121–31

Gindi G, Lee M, Rangarajan A and Zubal I G 1993 Bayesian reconstruction of functional images using anatomical information as priors *IEEE Trans. Med. Imaging* **12** 670–80

Gindi G, Rangarajan A, Lee M and Zubal I G 1993 Bayesian reconstruction for emission tomography via deterministic annealing *Information Processing in Medical Imaging* ed H Barrett and A Gmitro (Berlin: Springer) pp 322–38

Goitein M 1972 Three-dimensional density reconstruction from a series of two-dimensional projections *Nucl. Instrum. Methods* **101** 509–18

Gordon R, Bender R and Herman G T 1970 Algebraic reconstruction techniques for three-dimensional electron microscopy and X-ray photography *J. Theor. Biol.* **29** 471–81

Green P J 1990 Bayesian reconstructions from emission tomography data using a modified EM algorithm *IEEE Trans. Med. Imaging* **9** 84–93

Green P J 1996 MCMC in image analysis *Markov Chain Monte Carlo in Practice* ed W R Gilks, S Richardson and D J Spiegelhalter (London: Chapman and Hall) chapter 21, pp 381–99

Grenander U 1981 *Abstract Inference* (New York: Wiley)

Grootoonk S, Spinks T J, Sashin D, Spyrou N M and Jones T 1996 Correction for scatter in 3D brain PET using a dual energy window method *Phys. Med. Biol.* **41** 2757–74

Gullberg G T 1995 Entropy and transverse section reconstruction *Proc. 4th Int. Conf. on Information Processsing in Scintigraphy* ed C Raynaud and A Todd-Pokropek, pp 325–32, 2975

Gullberg G T, Hsieh Y L and Zeng G L 1996 An SVD reconstruction algorithm using a natural pixel representation of the attenuated Radon transform *IEEE Trans. Nucl. Sci.* **43** 295–303

Hebert T and Leahy R 1989 A generalized EM algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs priors *IEEE Trans. Med. Imaging* **8** 194–202

Hebert T, Leahy R and Singh M 1990 3D ML reconstruction for a prototype SPECT system *J. Opt. Soc. Am.* A **7** 1305–13

Helou Neto E S and De Pierro A R 2005 Convergence results for scaled gradient algorithms in positron emission tomography *Inverse Problems* **21** 1905–14

Higdon D M, Bowsher J E, Johnson V E, Turkington T G, Gilland D R and Jaszczak R J 1997 Fully Bayesian estimation of Gibbs hyperparameters for emission computed tomography data *IEEE Trans. Med. Imaging* **16** 516–26

Holte S, Schmidlin P, Linden A and Rosenqvist G 1990 Iterative image reconstruction for positron emission tomography: a study of convergence and quantitation problems *IEEE Trans. Nucl. Sci.* **37** 629–35

Hsiao I T, Rangarajan A and Gindi G 2001a A median prior for tomographic reconstruction, *Proc. IEEE NSS-MIC* pp 1779–82

Hsiao I T, Rangarajan A and Gindi G 2001b A smoothing prior with embedded positivity constraint for tomographic reconstruction *Int. Meeting on Fully 3D Image Reconstruction in Radiology and Nuclear Medicine* pp 81–4

Hsiao I-T, Rangarajan A, Khurd P and Gindi G 2004 An accelerated convergent ordered subsets algorithm for emission tomography *Phys. Med. Biol.* **49** 2145–56

Hsieh Y L, Zeng G L and Gullberg G T 1998 Projection space image reconstruction using strip functions to calculate pixels more natural for modeling the geometric response of the SPECT collimator *IEEE Trans. Med. Imaging* **17** 24–44

Huang Q, Zeng G L, You J and Gullberg G T 2005 An FDK-like cone-beam SPECT reconstruction algorithm for non-uniform attenuated projections acquired using a circular trajectory *Phys. Med. Biol.* **50** 2329–39

Huber P J 1981 *Robust Statistics* (New York: Wiley)

Hudson H M and Larkin R S 1994 Accelerated image reconstruction using ordered subsets of projection data *IEEE Trans. Med. Imaging* **13** 601–9

Huesman R H, Klein G J, Moses W W, J. Qi, Reutter B W and Virador P R G 2000 List mode maximum likelihood reconstruction applied to positron emission mammography with irregular sampling *IEEE Trans. Med. Imaging* **19** 532–7

Huesman R, Salmeron E and Baker J 1989 Compensation for crystal penetration in high resolution positron tomography *IEEE Trans. Nucl. Sci.* **36** 1100–7

Idier J 2001 Convex half-quadratic criteria and interacting auxiliary variables for image restoration *IEEE Trans. Image Process.* **10** 1001–9

Johnson C A, Seidel J and Sofer A 2000 Interior-point methodology for 3-D PET reconstruction *IEEE Trans. Med. Imaging* **19** 271–85

Johnson V 1994 A note on stopping rules in EM-ML reconstructions of ECT images *IEEE Trans. Med. Imaging* **13** 569–71

Johnson V E, Wong W H, Hu X and Chen C T 1991 Bayesian restoration of PET images using Gibbs priors *Proc. Int. Conf. on Information Processing in Medical Imaging* ed A Colchester and D Hawkes (Wilmington, DE: Wiley-Liss) pp 15–28

Katsevich A 2002 Theoretically exact FBP-type inversion algorithm for spiral CT *SIAM J. Appl. Math.* **62** 2012–26

Kaufman L 1987 Implementing and accelerating the EM algorithm for positron emission tomography *IEEE Trans. Med. Imaging* **6** 37–51

Kaufman L 1993 Maximum likelihood, least squares, and penalized least squares for PET *IEEE Trans. Med. Imaging* **12** 200–14

Khurd P, Hsiao I T, Rangarajan A and Gindi G 2004 A globally convergent regularized ordered-subset EM algorithm for list-mode reconstruction *IEEE Trans. Nucl. Sci.* **51** 719–25

Kinahan P E and Rogers J G 1989 Analytic 3D image reconstruction using all detected events *IEEE Trans. Nucl. Sci.* **36** 964–8

Koral K F and Rogers W L 1979 Application of ART to time-coded emission tomography *Phys. Med. Biol.* **24** 879–94

Lalush D S, Frey E C and Tsui B M W 2000 Fast maximum entropy approximation in SPECT using the RBI-MAP algorithm *IEEE Trans. Med. Imaging* **19** 286–94

Lalush D S and Tsui B M W 1998 Block-iterative techniques for fast 4D reconstruction using a priori motion models in gated cardiac SPECT *Phys. Med. Biol.* **43** 875–86

Lange K 1990 Convergence of EM image reconstruction algorithms with Gibbs smoothing *IEEE Trans. Med. Imaging* **9** 439–46

Lange K 1991 Convergence of EM image reconstruction algorithms with Gibbs smoothing *IEEE Trans. Med. Imaging* **10** 228 (erratum)

Lange K, Bahn M and Little R 1987 A theoretical study of some maximum likelihood algorithms for emission and transmission tomography *IEEE Trans. Med. Imaging* **6** 106–14

Lange K and Carson R 1984 EM reconstruction algorithms for emission and transmission tomography *J. Comput. Assist. Tomogr.* **8** 306–16

Lange K, Hunter D R and Yang I 2000 Optimization transfer using surrogate objective functions *J. Comput. Graph. Stat.* **9** 1–20

Laurette I, Zeng G L, Welch A, Christian P E and Gullberg G T 2000 A three-dimensional ray-driven attenuation, scatter and geometric response correction technique for SPECT in inhomogeneous media *Phys. Med. Biol.* **45** 3459–80

Leahy R and Yan X H 1991 Incorporation of anatomical MR data for improved functional imaging with PET *Information Processing in Medical Imaging* ed A Colchester and D Hawkes (Wilmington, DE: Wiley-Liss) pp 105–20

Lee K, Kinahan P E, Fessler J A, Miyaoka R S, Janes M and Lewellen T K 2004 Pragmatic fully 3D image reconstruction for the MiCES mouse imaging PET scanner *Phys. Med. Biol.* **49** 4563–78

Lee S-J, Rangarajan A and Gindi G 1995 Bayesian image reconstruction in SPECT using higher order mechanical models as priors *IEEE Trans. Med. Imaging* **14** 669–80

Levin C S, Dahlbom M and Hoffman E J 1995 A Monte Carlo correction for the effect of Compton scattering in 3-D PET brain imaging *IEEE Trans. Nucl. Sci.* **42** 1181–5

Levkovitz R, Falikman D, Zibulevsky M, Ben-Tal A and Nemirovski A 2001 The design and implementation of COSEM, an iterative algorithm for fully 3-D listmode data *IEEE Trans. Med. Imaging* **20** 633–42

Levitan E and Herman G T 1987 A maximum a posteriori probability expectation maximization algorithm for image reconstruction in emission tomography*IEEE Trans. Med. Imaging* **6** 185–92

Lewitt R M and Matej S 2003 Overview of methods for image reconstruction from projections in emission computed tomography *Proc. IEEE* **91** 1588–611

Lewitt R M and Muehllehner G 1986 Accelerated iterative reconstruction for positron emission tomography based on the EM algorithm for maximum likelihood estimation *IEEE Trans. Med. Imaging* **5** 16–22

Li Q, Ahn S and Leahy R 2005 Fast hybrid algorithms for PET image reconstruction *Proc. IEEE NSS-MIC (Puerto Rico)* pp 1933–7

Liang Z 1994 Detector response restoration in image reconstruction of high resolution positron emission tomography *IEEE Trans. Med. Imaging* **10** 314–21

Liang Z, Jaszczak R and Greer K 1989 On Bayesian image reconstruction from projections: uniform and nonuniform a priori source information *IEEE Trans. Med. Imaging* **8** 227–35

Liang Z, Ye J and Harrington D P 1994 An analytical approach to quantitative reconstruction of non-uniform attenuated brain SPECT *Phys. Med. Biol.* **39** 2023–41

Lipinski B, Herzog H, Rota Kops E, Oberschelp W and Muller-Gartner H W 1997 Expectation maximization reconstruction of positron emission tomography images using anatomical magnetic resonance information *IEEE Trans. Med. Imaging* **16** 129–36

Llacer J, Veklerov E, Coakley K, Hoffman E and Nunez J 1993 Statistical analysis of maximum likelihood estimator images of human brain FDG studies *IEEE Trans. Med. Imaging* **12** 215–31

Lucy L B 1974 An iterative technique for the rectification of observed distributions. (statistical astronomy) *Astron. J.* **79** 745–54

Luenberger D G 1984 *Linear and Nonlinear Programming* 2nd edn (Reading, MA: Addison-Wesley)

Matej S and Lewitt R 1996 Practical considerations for 3-D image reconstruction using spherically symmetric volume elements *IEEE Trans. Med. Imaging* **15** 68–78

Mumcuoglu E, Leahy R and Cherry S 1996 Bayesian reconstruction of PET images: methodology and performance analysis *Phys. Med. Biol.* **41** 1777–807

Mumcuoglu E U, Leahy R, Cherry S R and Zhou Z 1994 Fast gradient-based methods for Bayesian reconstruction of transmission and emission PET images *IEEE Trans. Med. Imaging* **13** 687–701

Nichols T E, J Qi, Asma E and Leahy R M 2002 Spatiotemporal reconstruction of list mode PET data *IEEE Trans. Med. Imaging* **21** 396–404

Nunez J and Llacer J 1990 A fast Bayesian reconstruction algorithm for emission tomography with entropy prior converging to feasible images *IEEE Trans. Med. Imaging* **9** 159–71

Nuyts J, Baete K, Beque D and Dupont P 2005 Comparison between map and postprocessed ml for image reconstruction in emission tomography when anatomical knowledge is available *IEEE Trans. Med. Imaging* **24** 667–75

Ollinger J M 1996 Model-based scatter correction for fully 3D PET *Phys. Med. Biol.* **41** 153–76

Panin V Y, Kehren F, Rothfuss H, Hu D, Michel C and Casey M E 2004 PET reconstruction with measured system matrix *Proc. IEEE NSS-MIC* pp 2483–7

Parra L and Barrett H H 1998 List-mode likelihood: EM algorithm and image quality estimation demonstrated on 2-D PET *IEEE Trans. Med. Imaging* 228–35

Paxman R G, Barrett H H, Smith W E and Milster T D 1985 Image reconstruction from coded data: II. Code design *J. Opt. Soc. Am.* A **2** 501–9

Qi J and Leahy R M 1999 A theoretical study of the contrast recovery and variance of MAP reconstructions from PET data *IEEE Trans. Med. Imaging* **18** 293–305

Qi J, Leahy R M, Cherry S R, Chatziioannou A and Farquhar T H 1998a High resolution 3D Bayesian image reconstruction using the microPET small animal scanner *Phys. Med. Biol.* **43** 1001–13

Qi J, Leahy R M, Hsu C, Farquhar T H and Cherry S R 1998b Fully 3D Bayesian image reconstruction for ECAT EXACT HR+ *IEEE Trans. Nucl. Sci.* **45** 1096–103

Rajeevan N, Rajgopal K and Krishna G 1992 Vector-extrapolated fast maximum likelihood estimation algorithms for emission tomography *IEEE Trans. Med. Imaging* **11** 9–20

Rangarajan A, Hsiao I T and Gindi G 2000 A Bayesian joint mixture framework for the integration of anatomical information in functional image reconstruction *J. Math. Imaging Vis.* **12** 199–217

Reader A J, Ally S, Bakatselos F, Manavaki R, Walledge R J, Jeavons A P, Julyan P J, Sha Zhao, Hastings D L and Zweit J 2002 One-pass list-mode EM algorithm for high-resolution 3-D PET image reconstruction into large arrays *IEEE Trans. Nucl. Sci.* **49** 693–9

Reader A J, Erlandsson K, Flower M A and Ott R J 1998a Fast accurate iterative reconstruction for low-statistics positron volume imaging *Phys. Med. Biol.* **43** 835–46

Reader A J, Erlandsson K, Flower M A and Ott R J 1998b Fast accurate iterative three-dimensional Bayesian reconstruction for low-statistics positron volume imaging *IEEE Trans. Nucl. Sci.* **45** 1090–5

Richardson W H 1972 Bayesian-based iterative method of image restoration *J. Opt. Soc. Am.* **62** 55–9

Rockmore A and Macovski A 1976 A maximum likelihood approach to emission image reconstruction from projections *IEEE Trans. Nucl. Sci.* **23** 1428–32

Rosenthal M S, Cullom J, Hawkins W, Moore S C, Tsui B M W and Yester M 1995 Quantitative SPECT imaging—a review and recommendations by the focus committee of the society-of-nuclear-medicine computer and instrumentation council *J. Nucl. Med.* **36** 1489–513

Sastry S and Carson R E 1997 Multimodality Bayesian algorithm for image reconstruction in positron emission tomography: a tissue composition model *IEEE Trans. Med. Imaging* **16** 750–61

Sauer K and Bouman C 1993 A local update strategy for iterative reconstruction from projections *IEEE Trans. Signal Process.* **41** 534–48

Shepp L A and Vardi Y 1982 Maximum likelihood reconstruction for emission tomography *IEEE Trans. Med. Imaging* **1** 113–22

Silverman B W, Jones M C, Wilson J D and Nychka D W 1990 A smoothed EM approach to indirect estimation problems, with particular reference to stereology and emission tomography *J. R. Stat. Soc. Ser.* B **52** 271–324

Snyder D L 1984 Utilizing side information in emission tomography *IEEE Trans. Nucl. Sci.* **31** 533–7

Snyder D L and Miller M 1985 The use of sieves to stabilize images produced with the EM algorithm for emission tomography *IEEE Trans. Nucl. Sci.* **32** 3864–72

Snyder D L and Politte D G 1983 Image reconstruction from list-mode data in an emission tomography system having time-of-flight measurements *IEEE Trans. Nucl. Sci.* **20** 1843–8

Sorenson H 1980 *Parameter Estimation: Principles and Problems* (New York: Dekker)

Tang Q, Zeng G L and Gullberg G T 2005 Analytical fan-beam and cone-beam reconstruction algorithms with uniform attenuation correction for SPECT *Phys. Med. Biol.* **50** 3153–70

Vardi Y, Shepp L A and Kaufman L 1985 A statistical model for positron emission tomography *J. Am. Stat. Assoc.* **80** 8–37

Veklerov E and Llacer J 1987 Stopping rule for the MLE algorithm based on statistical hypothesis testing *IEEE Trans. Med. Imaging* **6** 313–9

Veklerov E and Llacer J 1988 MLE reconstruction of a brain phantom using a Monte Carlo transition matrix and a statistical stopping rule *IEEE Trans. Nucl. Sci.* **35** 603–7

Wang C H, Chen J C and Liu R S 2004 Development and evaluation of MRI based Bayesian image reconstruction methods for PET *Comput. Med. Imaging Graph.* **28** 177–84

Wang W and Gindi G 1997 Noise analysis of MAP-EM algorithms for emission tomography *Phys. Med. Biol.* **42** 2215–32

Watson C C 2000 New, faster, image-based scatter correction for 3D PET *IEEE Trans. Nucl. Sci.* **47** 1587–94

Weir I S 1997 Fully Bayesian reconstructions from single-photon emission computed tomography data *J. Am. Stat. Assoc.* **92** 49–60

Wernick M N and Aarsvold J N 2004 *Emission Tomography : The Fundamentals of PET and SPECT* (New York: Academic)

Xia W, Lewitt R M and Edholm P R 1995 Fourier correction for spatially variant collimator blurring in SPECT *IEEE Trans. Med. Imaging* **14** 100–15

Yavuz M and Fessler J A 1998 Statistical image reconstruction methods for randoms-precorrected PET scans *Med. Image Anal.* **2** 369–78

Yavuz M and Fessler J A 2000 Maximum likelihood emission image reconstruction for random-precorrected PET scans *Proc. IEEE NSS-MIC* pp 15/229–15/233

Yu D F and Fessler J A 2000 Mean and variance of singles photon counting with deadtime *Phys. Med. Biol.* **45** 2043–56

Yu D F and Fessler J A 2002 Edge-preserving tomographic reconstruction with nonlocal regularization *IEEE Trans. Med. Imaging* **21** 159–73

Zheng J, Saquib S S, Sauer K and Bouman C A 2000 Parallelizable Bayesian tomography algorithms with rapid, guaranteed convergence *IEEE Trans. Image Process.* **9** 1745–59