Chapter 18

Emission ML Image Reconstruction

Contents

18.1	Introduction (s,eml,intro)	18.2
18.2	Problem statement (s,eml,state)	18.3
	18.2.1 Gradients	18.4
	18.2.2 Nonnegativity	18.5
	18.2.3 Uniqueness?	18.5
18.3	ML estimates are bounded (s,eml,bound)	18.5
18.4	E-ML-EM and E-ML-EM-3 algorithms (s,eml,em,intro)	18.6
	18.4.1 E-ML-EM-3 iterates are well defined (s,eml,em,def)	18.7
	18.4.2 E-ML-EM-3 iterates are bounded (s,eml,em,bound)	18.8
	18.4.3 Relationship to Kullback-Leibler divergence (s,eml,em,kl)	18.9
18.5	E-ML-EM-3 algorithm derivations (s,eml,em,derive)	18.9
	18.5.1 Fixed-point "derivation" (s,eml,em,fix)	18.10
	18.5.2 Optimization transfer derivation (s,eml,em,adp)	18.10
	18.5.3 Expectation-Maximization (EM) derivation (s,eml,em,em)	18.11
	18.5.4 Posinomial derivation (s,eml,em,pos)	18.12
	18.5.5 Alternating minimization derivation (s,eml,em,amp)	18.13
18.6	E-ML-EM acceleration	18.15
	18.6.1 Slow convergence (s,eml,em,slow)	18.15
	18.6.2 Acceleration methods (s,eml,em,accel)	18.15
	18.6.2.1 Algorithm variations	18.15
	18.6.2.1.1 Grid refinement	18.15
	18.6.2.1.2 Raised powers	18.15
	18.6.2.1.3 Over-relaxation	18.15
	18.6.2.1.4 Line search	18.15
	18.6.2.2 Implementation tricks	18.15
	18.6.2.2.1 System matrix	18.16
	18.6.2.2.2 Factoring attenuation	18.16
	18.6.2.2.3 Hardware acceleration	18.16
	18.6.2.2.4 Removing low-intensity pixels	18.16
	18.6.2.2.5 Mismatched back-projectors	18.16
18.7	Block iterative variations	18.17
	18.7.1 Ordered-subsets methods (s,eml,os,intro)	18.17
	18.7.1.1 E-ML-OSEM (s,eml,os)	18.17
	18.7.1.2 RBI	18.18
	18.7.1.3 RAMLA	18.18
	18.7.1.4 BSREM	18.18
	18.7.2 Incremental EM methods (s,eml,em,inc)	18.18

18.8 Alternative algorithms (s,eml,alg) 18.20					
18.8.1 General-purpose optimization methods	18.20				
18.8.2 Special-purpose methods	18.21				
18.8.3 Nested or cascade EM algorithms (s,eml,nested)	18.21				
18.9 E-ML-EM modifications	18.21				
18.9.1 Stopping rules (s,eml,em,stop)	18.21				
18.9.2 Smoothing and EMS (s,eml,ems)	18.21				
18.10E-ML-EM convergence proof (s,eml,em,proof) 18.22					
18.11Problems (s,eml,prob) 18.22					
18.12Bibliography 18.22					

ntro 18.1 Introduction (s,eml,intro)

This *chapter* summarizes maximum-likelihood (ML) algorithms for reconstructing emission images from Poisson measurements. The algorithms are applicable to PET, SPECT, and applications like photon-limited optical image restoration [1], as well as in astronomical problems [2], electron microscopy [3], and nuclear imaging [4]. We focus primarily on algorithms that monotonically increase the likelihood, are globally convergent, and that naturally accommodate the constraint that emission distributions are nonnegative¹.

In the context of PET and SPECT, statistical methods for image reconstruction have been used routinely in clinical practice since the mid 1990s. Not only do statistical methods yield images that "look better" than FBP images, but also, with proper physical modeling, statistical methods lead to improved performance in tasks such as lesion detection, *e.g.*, [8], particularly with proper system models [9]. For many years, the only likelihood-based methods available commercially for human scans were the unregularized ML methods of the kinds described in this *chapter*. Regularized methods of the kind discussed in Chapter 19 became available commercially in human scanners in about 2012 [10, 11], although such methods have long been available commercially for animal PET scanners [12].

Rockmore and Macovski first proposed a statistical approach to this problem in 1976 [13]. However, in deriving a solution they disregarded the nonnegativity constraint, leading circuitously to an unweighted least-squares estimator, noted to be erroneous by Vermeulen in 1982 [14, 15]. The publication of an expectation-maximization (EM) algorithm for ML estimation in emission tomography that same year by Shepp and Vardi [16] sparked a torrent of research on statistical image reconstruction (SIR) for PET and SPECT. Within two years of the initial publication of that E-ML-EM algorithm, the method had been extended to list-mode acquisitions [17] and to dynamic studies [18], and methods for using anatomical side information were described [19]. The first results with real PET data also appeared [20, 21]. A year later the SPECT problem with nonuniform attenuation and depth-dependent PSF was also addressed [22]. These early papers charted the course of numerous subsequent investigations.

On a historical note, Chen and Metz observed in [23] that the **Richardson-Lucy** deconvolution procedure [24, 25], which is algorithmically equivalent to E-ML-EM, had been proposed in the early 1970's for scintigraphy, citing [26] (an unpublished IPMI proceedings paper) and [27]. In a retrospective, Pizer [28] reiterates this statement. A similar iteration was proposed as a "generalized iterative scaling" method in the statistics literature [29]. Historians will have trouble sorting out "who was first" because Kosarev [30, eqn. (51)] cites a 1969 *preprint* (in Russian!) for an iteration that is the special case of E-ML-EM when the columns of the system matrix sum to unity. The iteration has been rediscovered more than once, *e.g.*, [31, 32].

There have been several surveys of iterative methods for tomographic image reconstruction [33–38]. Here we focus primarily on E-ML-EM type methods.

In the emission tomography literature, the kind of algorithm discussed here is usually called "the" ML-EM algorithm. That terminology overlooks the fact that there are many EM algorithms in general, and many EM variants even for emission tomography. This chapter uses the (non-standard) acronym E-ML-EM as reminder that it focuses on the *emission* problem.

This chapter is organized as follows. First the problem is defined and analyzed generally. Then several derivations of (a slight generalization of) the E-ML-EM algorithm are given, followed by a convergence proof. Acceleration methods are described, emphasizing the ordered-subsets or block-iterative approach. Then several alternative algorithms are discussed.

¹Modified algorithms have been proposed for problems like radar imaging where negative values can occur [5], or even in PET where absence of attenuation correction leads to inconsistent data that can induce negative values [6, 7].

s.eml.state

18.2 Problem statement (s,eml,state)

As described in Chapter 8, the goal in **emission tomography** is to reconstruct an emission distribution $\lambda(\vec{x})$ from recordings of emitted photons. We parameterize the emission distribution using a finite series expansion as in Chapter 10:

$$\lambda(\vec{x}) = \sum_{j=1}^{n_{\rm p}} x_j \, b_j(\vec{x}), \tag{18.2.1}$$

where x_j denotes the *j*th basis coefficient (*e.g.*, the mean number of emissions from the *j*th voxel for the pixel basis). The goal is to estimate the vector $\boldsymbol{x} = (x_1, \ldots, x_{n_p})$ from a realization $\boldsymbol{y} = (y_1, \ldots, y_{n_d})$ of the projection measurement random vector $\boldsymbol{Y} = (Y_1, \ldots, Y_{n_d})$.

Throughout this *chapter*, we use the usual Poisson statistical model² (as derived in Chapter 8) for the emission measurements:

$$Y_i \sim \mathsf{Poisson}\{\bar{y}_i(\boldsymbol{x}_{\mathrm{true}})\}, \qquad i = 1, \dots, n_{\mathrm{d}}, \tag{18.2.2}$$

where x_{true} denotes the "true" unknown value of x and, following (8.4.9), we model the measurement means by:

$$\bar{y}_i(\boldsymbol{x}) = \mathsf{E}[Y_i] = \sum_{j=1}^{n_{\rm p}} a_{ij} x_j + \bar{r}_i = [\boldsymbol{A}\boldsymbol{x}]_i + \bar{r}_i.$$
(18.2.3)

The vector $\bar{\boldsymbol{r}} = (\bar{r}_1, \dots, \bar{r}_{n_d})$ denotes (the mean of) background events such as random coincidences [40] and scatter [41]. The matrix $\boldsymbol{A} = \{a_{ij}\}$ represents the system model, including ray-dependent factors such as attenuation and detector efficiency, where a_{ij} is proportional to the probability that an emission from the *j*th voxel (or *j*th basis component more generally) is recorded by the *i*th detector. We assume the \bar{r}_i and a_{ij} values are known nonnegative constants. Numerous papers have shown that accurate models for the a_{ij} values can lead to significant improvements in image spatial resolution and accuracy, *e.g.*, [12, 42].

In some cases we assume $\bar{r} > 0$, which is reasonable and realistic because any real PET scan will have nonzero mean randoms. Similarly, any real SPECT scan will be contaminated by a nonzero mean scattered component and by a nonzero (but possibly very small) mean component from background radiation. The assumption that $\bar{r} > 0$, when used, is thus more physically realistic, and it also turns out to simplify the derivation of some algorithms.

To ensure that the problem is well-defined mathematically, we make the following assumption for the remainder of this *chapter*.

Assumption 18.2.1

$$a_j \triangleq \sum_{i=1}^{n_d} a_{ij} > 0, \qquad j = 1, \dots, n_p.$$
 (18.2.4)

The a_j values are often called **sensitivity factors** because they are related to the system sensitivity pattern (8.5.13). If any $a_j = 0$, then the *j*th voxel does not contribute to any of the measurements and x_j cannot be estimated meaningfully (in the unregularized case considered here). Such voxels are *invisible* to the imaging system [43]. In other words, all voxels in locations where the system sensitivity is zero must be excluded from the model and should not be estimated. In contrast, regularized methods can form estimates of x_j even if $a_j = 0$; see Chapter 19.

For ML image reconstruction, we seek a vector \hat{x} that maximizes the log-likelihood:

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x} \succeq \boldsymbol{0}}{\arg \max} \mathsf{L}(\boldsymbol{x}), \tag{18.2.5}$$

where $x \succeq 0$ imposes the nonnegativity constraint $x_j \ge 0$. It follows from the Poisson model (18.2.2) and (18.2.3) that the log-likelihood is³:

$$\mathsf{L}(\boldsymbol{x}) \stackrel{\mathrm{c}}{=} \sum_{i=1}^{n_{\mathrm{d}}} y_i \log \bar{y}_i(\boldsymbol{x}) - \bar{y}_i(\boldsymbol{x}), \tag{18.2.6}$$

where " $\stackrel{c}{=}$ " indicates that we neglect constants independent of x.

-,---,_--

e eml state lam

e,eml,a-

 $^{^{2}}$ For data that is not Poisson, the algorithms herein may still be useful provided one adjusts the data and the model so that the mean closely matches the variance [39].

³To be completely precise, we should write (18.2.6) as $L(x) \stackrel{c}{=} \sum_{i \in \mathcal{I}_+} y_i \log \bar{y}_i(x) - \sum_{i=1}^{n_d} \bar{y}_i(x)$, where the first sum (but not the second!) is restricted to measurements where $y_i > 0$. To avoid that notational inconvenience, we adopt the convention that $0 \log 0 = 0$ in such expressions.

In the *absence of noise* one could simply eliminate the y_i values that are zero and set the corresponding x_j values (for which $a_{ij} > 0$) to zero [44, p. 572], but this reasoning does *not* apply to the Poisson likelihood for noisy data in general.

We take the domain of the log-likelihood L(x) to be the set

$$\mathcal{G}_{+} \triangleq \left\{ \boldsymbol{x} \succeq \boldsymbol{0} : \left[\boldsymbol{A} \boldsymbol{x} \right]_{i} + \bar{r}_{i} > 0, \, \forall i \in \mathcal{I}_{+} \right\}, \tag{18.2.7}$$

where

$$\mathcal{I}_{+} \triangleq \{ i = 1, \dots, n_{\rm d} : y_i > 0 \}, \tag{18.2.8}$$

and $L : \mathcal{G}_+ \to (-\infty, \sum_{i=1}^{n_d} y_i \log y_i - y_i]$. The restriction $x \succeq 0$ in (18.2.7) is needed only on *physical* grounds, rather than *mathematical*.

The ML problem is to find a maximizer \hat{x} using an algorithm that converges rapidly and that requires as little CPU time per iteration as possible. Usually these are conflicting requirements, because often one can modify an algorithm to converge in fewer iterations, but at the expense of more work per iteration.

To proceed, we rewrite the log-likelihood (18.2.6) as follows:

$$\mathsf{L}(\boldsymbol{x}) = \sum_{i=1}^{n_{\rm d}} \mathsf{h}_i([\boldsymbol{A}\boldsymbol{x}]_i), \tag{18.2.9}$$

where

$$\mathsf{h}_{i}(l) \triangleq \begin{cases} y_{i} \log(l + \bar{r}_{i}) - (l + \bar{r}_{i}), & y_{i} > 0, \quad l + \bar{r}_{i} > 0 \\ -(l + \bar{r}_{i}), & y_{i} = 0, \quad l + \bar{r}_{i} \ge 0 \\ -\infty, & \text{otherwise.} \end{cases}$$
(18.2.10)

This h_i function is concave for $l \in (-\bar{r}_i, \infty)$, and is strictly concave over that interval if $y_i > 0$. Because \bar{y}_i is linearly related to the x_j values (in contrast to the nonlinear relationship in in the transmission case in Chapter 9), the emission reconstruction problem is considerably easier than the transmission problem. Many of the algorithms described in Chapter 11 and Chapter 14 apply to both the emission problem and to other inverse problems having log-likelihood functions of the general form (18.2.9). This *chapter* describes several algorithms for maximizing the emission log-likelihood L(x). Extensions to the regularized problem are described in Chapter 19.

18.2.1 Gradients

One could attempt to find a maximizer \hat{x} of the log-likelihood L(x) by examining the following Karush-Kuhn-Tucker (KKT) conditions (see §29.11) that correspond to the nonnegativity constraints:

$$\frac{\partial}{\partial x_j} \mathsf{L}(\hat{\boldsymbol{x}}) \begin{cases} = 0, \quad \hat{x}_j > 0\\ \le 0, \quad \hat{x}_j = 0, \end{cases}$$
(18.2.11)

where the partial derivatives of the log-likelihood (18.2.6) are:

$$\begin{split} \frac{\partial}{\partial x_j} \mathsf{L}(\boldsymbol{x}) &= \sum_{i=1}^{n_{\mathrm{d}}} \left(\frac{y_i}{\bar{y}_i(\boldsymbol{x})} - 1 \right) \frac{\partial}{\partial x_j} \, \bar{y}_i(\boldsymbol{x}) = \sum_{i=1}^{n_{\mathrm{d}}} \left(\frac{y_i}{\bar{y}_i(\boldsymbol{x})} - 1 \right) a_{ij} \\ &= \sum_{i=1}^{n_{\mathrm{d}}} a_{ij} \left(\frac{y_i - [\boldsymbol{A}\boldsymbol{x}]_i - \bar{r}_i}{\bar{y}_i(\boldsymbol{x})} \right) = \sum_{i=1}^{n_{\mathrm{d}}} a_{ij} \frac{1}{\bar{y}_i(\boldsymbol{x})} \left(y_i - \bar{r}_i - [\boldsymbol{A}\boldsymbol{x}]_i \right). \end{split}$$

The corresponding column gradient is

$$\nabla \mathsf{L}(\boldsymbol{x}) = \boldsymbol{A}' \operatorname{Diag}\{1/\bar{y}_i(\boldsymbol{x})\}(\boldsymbol{y}-\bar{\boldsymbol{r}}) - \boldsymbol{A}' \operatorname{Diag}\{1/\bar{y}_i(\boldsymbol{x})\} \boldsymbol{A}\boldsymbol{x}.$$
(18.2.12)

Lemma 18.5.1 combines this equality with (18.2.11) for one of several derivations.

18.2.2 Nonnegativity

One can find in the literature, e.g., [2], incorrect expressions for the ML estimate of the form

$$\hat{\boldsymbol{x}} \stackrel{\prime}{=} [\boldsymbol{A}' \operatorname{Diag}\{1/\bar{y}_i(\hat{\boldsymbol{x}})\} \boldsymbol{A}]^{-1} \boldsymbol{A}' \operatorname{Diag}\{1/\bar{y}_i(\hat{\boldsymbol{x}})\}(\boldsymbol{y}-\bar{\boldsymbol{r}}).$$

These are "derived" by zeroing the gradient of the log-likelihood in (18.2.12). However, such a "solution" \hat{x} is rarely nonnegative, and it is even possible that $\bar{y}_i(\hat{x})$ could be negative. As noted in [45], such derivations ignore the **nonnegativity constraint** associated with \hat{x} , or, at a minimum, the requirement that $\bar{y}_i \ge 0$ (with equality possible only if $y_i = 0$). Furthermore, even if the above expression were correct, the right-hand side depends on \hat{x} so it is not a direct solution. Thus, iterative algorithms are needed to find \hat{x} , and such algorithms must consider both parts of the KKT conditions (18.2.11).

e,eml,Gr

e.eml.L.hi

e.eml.hi

e.eml.kkt

18.2.3 Uniqueness?

One can verify that the (negative) Hessian of the log-likelihood is

$$-\nabla^2 \operatorname{L}({\boldsymbol x}) = {\boldsymbol A}'\operatorname{Diag}\!\left\{\frac{y_i}{\bar y_i^2({\boldsymbol x})}\right\}{\boldsymbol A},$$

which is a positive-semidefinite matrix (Because $y_i \ge 0$). Thus the log-likelihood is a concave function (cf. §29.9), but it is not strictly concave in general. There may be many maximizers of L(x), all of which are equally "satisfactory" in terms of their log-likelihood. When multiple maximizers exist, it is an open problem whether any of them are preferable in terms of image quality.

Example 18.2.2 Consider the case where $\mathbf{A} = [1 \ 1]$, and $y_1 > r_1 \ge 0$. Then there are multiple (nonnegative) ML solutions of the form $\{(x_1, x_2) : x_1 \in [0, y_1 - r_1], x_2 = y_1 - \bar{r}_1 - x_1\}$.

Byrne [46] shows that *if* the system matrix A has full rank, then when $n_p \ge n_d$ the ML estimate is unique in the case where there are no nonnegative solutions to the system of equations $y - \bar{r} = Ax$. In practice this full rank property seems quite difficult to verify, so one must be prepared for the ML estimate to be non-unique. See [47].

Philosophically, this non-uniqueness seems rather undesirable, and is addressed easily by including a regularizing penalty function because most reasonable choices easily ensure uniqueness of \hat{x} [48] (see §19.4). In the absence of such regularization, each convergent ML iterative algorithm could converge to a different ML solution, a state of affairs that seems difficult to advocate. Furthermore, practitioners of unregularized ML often stop iterating before convergence, and the properties of the intermediate images are also highly algorithm dependent.

18.3 ML estimates are bounded (s,eml,bound)

To aid in proving convergence of some iterative algorithms, it is helpful to establish that the ML estimates \hat{x} are bounded. The following theorem *constructs* an easily computable upper bound U = U(y, A) for which $\hat{x}_j \leq U, \forall j$. The physical intuition behind this bound is that the largest possible values for \hat{x}_j correspond to the case where all of the counts recorded by the *i*th detector originated in the lowest-sensitivity pixel that contributes to the *i*th ray.

Theorem 18.3.1 Under the assumptions given above, the ML estimates in (18.2.5) satisfy the upper bound $\hat{x}_j \leq U, \forall j$, where

$$U \triangleq U(\boldsymbol{y}, \boldsymbol{A}) = \max_{i} \left\{ \frac{y_i}{\min_{a_{ij} \neq 0} a_{ij}} \right\}.$$

Proof:

s.eml.bound

Suppose x is a vector for which the set of "too large" elements $\mathcal{J} = \{j = 1, ..., n_p : x_j > U\}$ is nonempty. Define the "clipped" estimate $\tilde{x}_j = \min\{x_j, U\}, j = 1, ..., n_p$. To establish that \hat{x} cannot have elements larger than U, it suffices to show that $L(x) < L(\tilde{x}) \le L(\hat{x})$.

First, note that $h_i(l)$ in (18.2.10) is monotone decreasing when $l + \bar{r}_i \ge y_i$. Second, note that

$$\boldsymbol{x} = \tilde{\boldsymbol{x}} + \sum_{j \in \mathcal{J}} \delta_j \boldsymbol{e}_j,$$

where $\delta_j = x_j - U > 0$ by construction. Now define $\mathcal{I} = \{i \in \{1, \dots, n_d\} : \exists j \in \mathcal{J} \text{ such that } a_{ij} > 0\}$. For $i \in \mathcal{I}$ we have $[\mathbf{A}\tilde{\mathbf{x}}]_i \ge a_{ij}\tilde{x}_j = a_{ij}U \ge y_i$, and $\delta_i \triangleq [\mathbf{A}\tilde{\mathbf{x}}]_i - [\mathbf{A}\mathbf{x}]_i = \sum_{j \in \mathcal{J}} a_{ij}\delta_j > 0$. Thus for $i \in \mathcal{I}$,

$$\mathsf{h}_i([\boldsymbol{A}\boldsymbol{x}]_i) = \mathsf{h}_i([\boldsymbol{A}\tilde{\boldsymbol{x}}]_i + \delta_i) < \mathsf{h}_i([\boldsymbol{A}\tilde{\boldsymbol{x}}]_i) \,.$$

For any $i \notin \mathcal{I}$, we have $h_i([Ax]_i) = h_i([A\tilde{x}]_i)$. By the assumptions that \mathcal{J} is nonempty and that the sensitivity factors $\sum_{i=1}^{n_d} a_{ij}$ are nonzero, we have that \mathcal{I} is nonempty. Thus we have strict inequality of at least one of the h_i functions, so

$$\mathsf{L}(\boldsymbol{x}) = \sum_{i \in \mathcal{I}} \mathsf{h}_i([\boldsymbol{A}\boldsymbol{x}]_i) + \sum_{i \notin \mathcal{I}} \mathsf{h}_i([\boldsymbol{A}\boldsymbol{x}]_i) < \sum_{i \in \mathcal{I}} \mathsf{h}_i([\boldsymbol{A}\tilde{\boldsymbol{x}}]_i) + \sum_{i \notin \mathcal{I}} \mathsf{h}_i([\boldsymbol{A}\tilde{\boldsymbol{x}}]_i) = \mathsf{L}(\tilde{\boldsymbol{x}}) \,.$$

So "clamping" all elements of x greater than U will always increase the likelihood.

For the case of a regularized cost function with a penalty function that is based on differences of neighboring pixels with a potential function $\psi(t)$ that is nondecreasing in |t|, one can also show that "clipping" all elements of x greater than U will always decrease the roughness penalty. So the same bound U applies to the regularized case for one family of cost functions. Ahn [49, App. A.2] has established an upper bound U that applies to an even broader family of penalty functions.

18.4 E-ML-EM and E-ML-EM-3 algorithms (s,eml,em,intro)

We now turn to algorithms for finding a ML estimate. The classical **expectation-maximization** (**EM**) algorithm for solving the ML estimation problem (18.2.5) has the following equivalent expressions.

$$\begin{array}{|c|c|c|c|c|c|} \hline \textbf{E-ML-EM Algorithm} \\ x_{j}^{(n+1)} &= \frac{x_{j}^{(n)}}{a_{j}} \sum_{i \in \mathcal{I}_{+}} a_{ij} \frac{y_{i}}{\bar{y}_{i}(\boldsymbol{x}^{(n)})} & (18.4.1) \\ &= x_{j}^{(n)} + \frac{x_{j}^{(n)}}{a_{j}} \sum_{i \in \mathcal{I}_{+}} a_{ij} \left(\frac{y_{i}}{\bar{y}_{i}(\boldsymbol{x}^{(n)})} - 1 \right) \\ &= x_{j}^{(n)} + \frac{x_{j}^{(n)}}{a_{j}} \frac{\partial}{\partial x_{j}} L(\boldsymbol{x}^{(n)}), \qquad j = 1, \dots, n_{\mathrm{p}}. \end{array}$$

This *chapter* presents several distinct derivations of the following modestly generalized form of this algorithm, originally called **E-ML-EM-3** in [50].

E-ML-EM-3 Algorithm

$$e_j(\boldsymbol{x}) \triangleq \sum_{i \in \mathcal{I}_+} a_{ij} y_i / \bar{y}_i(\boldsymbol{x}), \qquad j = 1, \dots, n_{\rm p}$$
(18.4.3)

$$x_{j}^{(n+1)} = \left[\frac{x_{j}^{(n)} + \gamma_{j}}{a_{j}} e_{j}(\boldsymbol{x}^{(n)}) - \gamma_{j}\right]_{+}, \qquad j = 1, \dots, n_{p}$$
(18.4)

$$= \left[x_j^{(n)} + \frac{x_j^{(n)} + \gamma_j}{a_j} \frac{\partial}{\partial x_j} \mathsf{L}(\boldsymbol{x}^{(n)}) \right]_+.$$
(18.4.5)

As explained in subsequent derivations, the γ_j values are any (user-selected) nonnegative constants that satisfy the following constraints

$$\sum_{j=1}^{n_{\rm p}} a_{ij} \gamma_j \le \bar{r}_i, \qquad i = 1, \dots, n_{\rm d}.$$
(18.4.6)

One way to specify the γ_j values is as follows:

$$\gamma_j = \min_{a_i. \neq 0} \frac{\bar{r}_i}{a_{i.}}, \text{ where } a_{i.} = \sum_{j=1}^{n_p} a_{ij}.$$

For other choices, see [51]. When the γ_j values are all zero, then the E-ML-EM-3 algorithm is identical to the classical E-ML-EM algorithm. The $[\cdot]_+$ operator in (18.4.4) enforces the nonnegativity constraint. The form of (18.4.5) shows that these algorithms are each **diagonally preconditioned gradient projection** methods (*cf.* (12.2.1)), having the matrix-vector form

$$\boldsymbol{x}^{(n+1)} = [\boldsymbol{x}^{(n)} + \boldsymbol{D}(\boldsymbol{x}^{(n)}) \nabla \mathsf{L}(\boldsymbol{x}^{(n)})]_{+},$$

where the **diagonal preconditioner** here is: $D(x^{(n)}) \triangleq \mathsf{Diag}\left\{\frac{x_j^{(n)} + \gamma_j}{a_j}\right\}$.

Both of the above algorithms are parallelizable in the sense that they update all pixels simultaneously. They both enforce the nonnegativity constraint naturally, and both increase the likelihood monotonically each iteration. An interesting difference is that, when initialized with $x^{(0)} \succ 0$, ordinary E-ML-EM iterates remain positive for all iterations, (for $j \in \mathcal{J}_+$ defined below), whereas E-ML-EM-3 iterates can move on and off of the boundary of the nonnegative orthant. This flexibility can help accelerate convergence somewhat [50].

s, eml, em, def 18.4.1 E-ML-EM-3 iterates are well defined (s, eml, em, def)

The iteration (18.4.3)-(18.4.4) contains ratios, so before further analysis we must first verify that a "divide by zero" condition cannot arise. This section confirms that (18.4.4) is well defined under conditions on A and y that are quite reasonable in practice.

We begin by defining some more index sets (cf. [52]):

$$\mathcal{J}_{+} \triangleq \left\{ j : \sum_{i=1}^{n_{d}} a_{ij} y_{i} > 0 \right\}$$
$$\mathcal{J}_{i} \triangleq \left\{ j : a_{ij} > 0 \right\}, \qquad i = 1, \dots, n_{d}$$
$$\mathcal{I}_{j} \triangleq \left\{ i : a_{ij} > 0 \right\}, \qquad j = 1, \dots, n_{p}$$
$$\mathcal{F}_{+} \triangleq \left\{ \boldsymbol{x} \succeq \boldsymbol{0} : \forall j \in \mathcal{J}_{+}, \text{ either } x_{j} > 0 \text{ or } \gamma_{j} > 0 \right\}$$

Assumption 18.4.1 For all $i \in \mathcal{I}_+$, the set \mathcal{J}_i is nonempty or $\bar{r}_i > 0$ (or both).

This assumption is simply an expression of part of the Poisson statistical model; one cannot have $y_i > 0$ yet have $\bar{y}_i = 0$ under the Poisson model. We make this assumption explicit rather than relying on the Poisson model because often these algorithms are applied to measurements that are not exactly Poisson distributed. In such cases, the user must verify the validity of this assumption.

Here are two properties of $e_i(\cdot)$ that follow immediately from these definitions and (18.4.3).

• On the set \mathcal{G}_+ defined in (18.2.7), the term e_j in (18.4.3) is well defined because the denominators are nonzero.

• If $j \in \mathcal{J}_+$, then $e_j(\boldsymbol{x}) > 0$ for all $\boldsymbol{x} \in \mathcal{G}_+$.

Thus, if $x^{(n)} \in \mathcal{G}_+$, then the following refinement of (18.4.4) is well-defined:

$$x_{j}^{(n+1)} = \mathcal{M}_{j}(\boldsymbol{x}^{(n)}) \triangleq \begin{cases} \left[(x_{j}^{(n)} + \gamma_{j}) e_{j}(\boldsymbol{x}^{(n)}) / a_{j} - \gamma_{j} \right]_{+}, & j \in \mathcal{J}_{+} \\ 0, & j \notin \mathcal{J}_{+}. \end{cases}$$
(18.4.7)

(Recall $a_j > 0$ by Assumption 18.2.1.) When using this algorithm, we will always choose $x^{(0)} > 0$, so that $x^{(0)} \in \mathcal{G}_+$. To apply the above algorithm recursively, *i.e.*, $x^{(n+1)} = \mathcal{M}(x^{(n)})$, we must establish that $x^{(n)} \in \mathcal{G}_+$ for all $n \in \mathbb{N}$.

Lemma 18.4.2 If $i \in \mathcal{I}_+$, then either $\bar{r}_i > 0$ or $\mathcal{J}_i \cap \mathcal{J}_+$ is nonempty (or both). Proof:

Suppose not, *i.e.*, suppose $\exists i \in \mathcal{I}_+$ such that $\bar{r}_i = 0$ and $\mathcal{J}_i \cap \mathcal{J}_+$ is empty. Then $j \in \mathcal{J}_i \Longrightarrow j \notin \mathcal{J}_+$. By Assumption 18.4.1, \mathcal{J}_i is nonempty when $\bar{r}_i = 0$. Picking any $j \in \mathcal{J}_i$ we have $a_{ij} > 0$ by definition. But $j \notin \mathcal{J}_+$ implies $\sum_{i=1}^{n_d} a_{ij} y_i = 0$, and hence $a_{ij} y_i = 0$. Because $y_i > 0$ for $i \in \mathcal{I}_+$, we have $a_{ij} = 0$, a contradiction.

l,eml,em,2

,eml,em,poisso

Lemma 18.4.3 If $x^{(n)} \in \mathcal{G}_+$, then $\forall j \in \mathcal{J}_+$, if $x_j^{(n)} > 0$ and $\gamma_j = 0$, then $x_j^{(n+1)} > 0$. Proof:

Suppose $\gamma_j = 0$ and $x_j^{(n+1)} = 0$ for some $j \in \mathcal{J}_+$ with $x_j^{(n)} > 0$. Because in this case $x_j^{(n+1)} = x_j^{(n)} e_j(\boldsymbol{x}^{(n)}) / a_j$, having $x_j^{(n+1)} = 0$ would imply $e_j(\boldsymbol{x}^{(n)}) = 0$, which is impossible for $j \in \mathcal{J}_+$ and $\boldsymbol{x}^{(n)} \in \mathcal{G}_+$. \Box

l,eml,em,3

Lemma 18.4.4 If $x^{(n)} \in \mathcal{F}_+ \cap \mathcal{G}_+$ then $x^{(n+1)} \in \mathcal{F}_+ \cap \mathcal{G}_+$.

Proof:

Suppose $\boldsymbol{x}^{(n)} \in \mathcal{F}_+$ and $\boldsymbol{x}^{(n)} \in \mathcal{G}_+$. For $j \in \mathcal{J}_+$, if $\gamma_j = 0$ then $x_j^{(n)} > 0$, so by Lemma 18.4.3 $x_j^{(n+1)} > 0$. Thus $\boldsymbol{x}^{(n+1)} \in \mathcal{F}_+$.

Now we must show that $x^{(n+1)} \in \mathcal{G}_+$, *i.e.*, $\sum_{j=1}^{n_p} a_{ij} x_j^{(n+1)} + \bar{r}_i > 0$ for all $i \in \mathcal{I}_+$.

Suppose not, *i.e.*, $\sum_{j=1}^{n_{\rm p}} a_{ij} x_j^{(n+1)} + \bar{r}_i = 0$ for some $i \in \mathcal{I}_+$. Then $\bar{r}_i = 0$ and $x_j^{(n+1)} = 0$ for all $j \in \mathcal{J}_i$. Furthermore, by (18.4.6), we have $\gamma_j = 0$ for all $j \in \mathcal{J}_i$. By Lemma 18.4.2, $\mathcal{J}_i \cap \mathcal{J}_+$ is nonempty, so there is some $j \in \mathcal{J}_+ \cap \mathcal{J}_i$ for which $\gamma_j = 0$ and $x_j^{(n+1)} = 0$. This contradicts the previous conclusion $\boldsymbol{x}^{(n+1)} \in \mathcal{F}_+$.

Thus, because $x^{(0)} > 0$ implies $x^{(0)} \in \mathcal{F}_+ \cap \mathcal{G}_+$, it follows by induction that $x^{(n)} \in \mathcal{G}_+$ for all $n \in \mathbb{N}$, so the iteration (18.4.7) is well defined. Because \mathcal{F}_+ and \mathcal{G}_+ are open sets, some care is required when examining limit points of $\{x^{(n)}\}$ in convergence proofs.

The braces in (18.4.7) are a nuisance, so hereafter we assume that \mathcal{J}_+ is all n_p columns of A, *i.e.*, the degenerate columns where $\sum_{i=1}^{n_d} a_{ij} y_i$ is zero have been removed prior to iterating. This assumption is implicit when writing (18.4.4). This modification is needed only for unregularized EM algorithms.

s, eml, em, bound 18.4.2 E-ML-EM-3 iterates are bounded (s, eml, em, bound)

When $\bar{r} = 0$ (and hence the γ_i values are all zero), one can show from (18.4.1) that

$$\sum_{j=1}^{n_{\rm p}} a_j x_j^{(n)} = \sum_{i=1}^{n_{\rm d}} y_i, \qquad \forall n > 0,$$
(18.4.8)

which is known as the (sensitivity weighted) "count preserving" property of the E-ML-EM algorithm. In fact (18.4.8) is a property of the ML solution \hat{x} that E-ML-EM happens to inherit for each iterate $x^{(n)}$ when $\bar{r} = 0$. This property has been used to simplify many E-ML-EM convergence proofs because it ensures that $\{x^{(n)}\}$ and \hat{x} lie in the simplex described by (18.4.8). The simplex (18.4.8) is a bounded set (within the nonnegative orthant).

Unfortunately, the count-preserving property (18.4.8) *does not hold* when $\bar{r} \neq 0$. However, the following Lemma provides a reasonable generalization.

Lemma 18.4.5 The iterates produced by the E-ML-EM-3 algorithm (18.4.4) are bounded. Proof:

Because $a_j > 0$, $a_j x_j^{(n+1)} = \left[(x_j^{(n)} + \gamma_j) e_j(\boldsymbol{x}^{(n)}) - a_j \gamma_j \right]_+ \le (x_j^{(n)} + \gamma_j) e_j(\boldsymbol{x}^{(n)})$. Thus

$$\sum_{j=1}^{n_{\rm p}} a_j x_j^{(n+1)} \le \sum_{j=1}^{n_{\rm p}} (x_j^{(n)} + \gamma_j) e_j(\boldsymbol{x}^{(n)}) = \sum_{i=1}^{n_{\rm d}} y_i \frac{\bar{y}_i^{(n)} - \bar{r}_i + \sum_{j'=1}^{n_{\rm p}} a_{ij'} \gamma_{j'}}{\bar{y}_i^{(n)}} \le \sum_{i=1}^{n_{\rm d}} y_i$$

provided (18.4.6) holds. Thus

$$0 \le x_j^{(n)} \le \frac{1}{a_j} \sum_{i=1}^{n_d} y_i.$$
(18.4.9)

Combining these inequalities yields the following bound on sensitivity-weighted total counts:

$$\sum_{j=1}^{n_{\rm p}} a_j x_j^{(n)} \le \sum_{i=1}^{n_{\rm d}} y_i.$$
(18.4.10)

Using Lemma 18.5.1, one can also show that all ML estimates \hat{x} satisfy the above upper bounds.

Finding a nontrivial lower bound to complement (18.4.10) is an open problem for $\bar{r} \neq 0$. Intuitively, we would expect that as n increases, we should have $\bar{y}_i(\boldsymbol{x}^{(n)}) \approx y_i$, so $\sum_{j=1}^{n_p} a_j x_j^{(n+1)} \approx \sum_{i=1}^{n_d} y_i - \bar{r}_i$. However, translating this approximation into a suitable bound is an elusive open problem.

s, em1, em, k1 18.4.3 Relationship to Kullback-Leibler divergence (s, em1, em, k1)

The Poisson emission log-likelihood (18.2.6) has a close relationship with the **Kullback-Leibler divergence** [53] and the **generalized KL divergence** and this relation permeates papers on EM algorithms. This section summarizes some of its properties because they will be used in subsequent derivations and in the convergence proof.

The KL divergence of two scalars is defined as follows [54, p. 67]:

$$\kappa(u,v) \triangleq \begin{cases} v, & u = 0, \ v \ge 0\\ u \log \frac{u}{v} - u + v, & u > 0, \ v > 0. \end{cases}$$
(18.4.11)

The domain of κ is $\mathcal{D}_{\kappa} = \{(0,\infty) \times (0,\infty)\} \cup (0,0)$, so $\kappa : \mathcal{D}_{\kappa} \to [0,\infty)$. Sometimes one writes $\kappa(u,0) = \infty$, for u > 0.

This function has the following simple but important properties.

• $\kappa(u, v) \ge 0$ with equality iff u = v.

This is easily verified using the inequality $\log x \le x - 1$ with equality if and only if x = 1.

If
$$\kappa(u, v) \to \kappa(u, u)$$
 then $v \to u$. (18.4.12)

If u = 0, then $\kappa(u, v) = v$, so the result is clear.

If u > 0, then $\kappa(u, v) - \kappa(u, u) = u \log \frac{u}{v} \to 0 \Longrightarrow u/v \to 1 \Longrightarrow v \to u$.

• For any $u \ge 0$, the minimizer of $\kappa(u, v)$ over v is the unique value v = u.

- $\kappa(u, v)$ is continuous (in both arguments) for $(u, v) \in (0, \infty) \times (0, \infty)$.
- The partial derivatives of κ are

$$\frac{\partial}{\partial u}\kappa(u,v) = \log\frac{u}{v}, \qquad u,v > 0$$

$$\frac{\partial}{\partial v}\kappa(u,v) = \begin{cases} 1 - \frac{u}{v}, & u,v > 0\\ 1, & u = 0. \end{cases}$$
(18.4.13)

(For the case u = 0, we can extend the domain of $\kappa(0, v)$ to all of \mathbb{R} so that $\frac{\partial}{\partial v} \kappa(0, v) = 1$ for any v, including v = 0.) • $\frac{\partial^2}{\partial u^2} \kappa(u, v) = 1/u$ for $u > 0, v > 0; \frac{\partial^2}{\partial v^2} \kappa(u, v) = u/v^2$ for $u \ge 0, v > 0$. So κ is convex in u and in v. • One can also show that κ is jointly convex in u and v. (See Problem 18.1.)

- $\kappa(\alpha u, v) = \alpha \kappa(u, v/\alpha)$ for $\alpha > 0$.
- In general, κ is not symmetric, hence the term **divergence** rather than **distance**.
- For $(u, v) \in \mathcal{D}_{\kappa}$ [55, Lemma 2.2]:

$$(u-v)^2 \le \left(\frac{2}{3}u + \frac{4}{3}v\right)\kappa(u,v). \tag{18.4.14}$$

A special case of (18.4.14) is **Pinsker's inequality** [56] [57, p. 88]:

$$(u-v)^2 \le 2\max(u,v)\,\kappa(u,v)\,. \tag{18.4.15}$$

These properties have been central to most convergence proofs for E-ML-EM, including the one in §18.10. For example, for $x, z \in \mathcal{G}_+$ one can express differences of the Poisson log-likelihood as follows:

$$\mathsf{L}(\boldsymbol{x}) - \mathsf{L}(\boldsymbol{z}) = \left(\sum_{i=1}^{n_{\rm d}} y_i \log \bar{y}_i(\boldsymbol{x}) - \bar{y}_i(\boldsymbol{x}) \right) - \left(\sum_{i=1}^{n_{\rm d}} y_i \log \bar{y}_i(\boldsymbol{z}) - \bar{y}_i(\boldsymbol{z}) \right)$$

= $\sum_{i=1}^{n_{\rm d}} y_i \log \frac{\bar{y}_i(\boldsymbol{x})}{\bar{y}_i(\boldsymbol{z})} - \bar{y}_i(\boldsymbol{x}) + \bar{y}_i(\boldsymbol{z}) = \sum_{i=1}^{n_{\rm d}} \kappa(y_i, \bar{y}_i(\boldsymbol{z})) - \kappa(y_i, \bar{y}_i(\boldsymbol{x})) .$ (18.4.16)

18.5 E-ML-EM-3 algorithm derivations (s,eml,em,derive)

This section presents several different derivations of the E-ML-EM-3 algorithm (18.4.4). Each distinct derivation provides a unique approach that may be useful for future algorithm development.

Fixed-point "derivation" (s,eml,em,fix) 18.5.1

As described in §11.2, a simple approach to "deriving" iterative algorithms is to examine the stationary points of the cost function. A useful expression for the partial derivatives of the Poisson log-likelihood (18.2.6) is

$$\frac{\partial}{\partial x_j} \mathsf{L}(\boldsymbol{x}) = \sum_{i=1}^{n_{\mathrm{d}}} a_{ij} \left(y_i / \bar{y}_i(\boldsymbol{x}) - 1 \right) = e_j(\boldsymbol{x}) - a_j,$$

where e_i was defined in (18.4.3).

Lemma 18.5.1 For any set of nonnegative values $\{\gamma_i\}$, if a nonnegative vector \hat{x} satisfies the KKT conditions (18.2.11), then \hat{x} is a fixed point of the iteration (18.4.4), i.e.,

$$\hat{x}_{j} = \left[\left(\hat{x}_{j} + \gamma_{j} \right) e_{j}(\hat{x}) / a_{j} - \gamma_{j} \right]_{+}.$$
(18.5.1)

Proof:

Suppose $\hat{x} \succeq 0$ satisfies the KKT conditions (18.2.11).

If $\hat{x}_j > 0$ then from (18.2.11) $e_j(\hat{x}) = a_j$, so $\hat{x}_j = (\hat{x}_j + \gamma_j) e_j(\hat{x}) / a_j - \gamma_j > 0$ and thus (18.5.1) holds. If $\hat{x}_j = 0$, then from (18.2.11) $e_j(\hat{x}) \le a_j$, so $(0 + \gamma_j) e_j(\hat{x}) / a_j - \gamma_j \le 0$, so $[(0 + \gamma_j) e_j(\hat{x}) / a_j - \gamma_j]_+ = 0$.

Convergence analysis would be simplified if the converse of this result held, but it need not. If (18.5.1) holds and $\hat{x}_j > 0$ then $\hat{x}_j = (\hat{x}_j + \gamma_j) e_j(\hat{x}) / a_j - \gamma_j$ so

$$e_j(\hat{\boldsymbol{x}}) = a_j. \tag{18.5.2}$$

e,eml,em,e=

e,eml,em,kl,pinsker

However, if $\hat{x}_j = 0$ then $\gamma_j e_j(\hat{x}) / a_j - \gamma_j \leq 0$ and from this equality alone we cannot conclude that $e_j(\hat{x}) \leq a_j$, because we may have $\gamma_j = 0$. Consequently, convergence proofs must treat this issue more carefully.

The necessary condition (18.5.1) is suggestive of the "fixed point" iteration (18.4.4). However, this "derivation" gives absolutely no insight into how to choose the γ_j values to ensure convergence, or even whether convergence is possible. The rigorous derivations that follow overcome this shortcoming of fixed-point "derivations" [16].

s, eml, em, adp 18.5.2 Optimization transfer derivation (s, eml, em, adp)

This section presents a derivation based on the **optimization transfer** principles described in Chapter 14. Although this is not how E-ML-EM was derived originally, it is perhaps the simplest rigorous derivation that clearly illuminates the monotonicity properties of E-ML-EM-3.

Considering the form of the log-likelihood (18.2.9), we can generalize an observation of De Pierro [58] to see that if $x^{(n)} > 0$, then (for $i \in \mathcal{I}_+$):

$$\begin{split} \bar{y}_{i}(\boldsymbol{x}) &= \left[\boldsymbol{A}\boldsymbol{x}\right]_{i} + \bar{r}_{i} = \sum_{j=1}^{n_{\mathrm{p}}} a_{ij}(x_{j} + \gamma_{j}) + \check{r}_{i} \\ &= \sum_{j=1}^{n_{\mathrm{p}}} \left(\frac{a_{ij}(x_{j}^{(n)} + \gamma_{j})}{\bar{y}_{i}^{(n)}}\right) \frac{x_{j} + \gamma_{j}}{x_{j}^{(n)} + \gamma_{j}} \, \bar{y}_{i}^{(n)} + \left(\frac{\check{r}_{i}}{\bar{y}_{i}^{(n)}}\right) \bar{y}_{i}^{(n)}, \end{split}$$
(18.5.3)

where $\bar{y}_i^{(n)} \triangleq \bar{y}_i(\boldsymbol{x}^{(n)})$ and we define the following nonnegative constants (due to (18.4.6)):

$$\check{r}_i \triangleq \bar{r}_i - \sum_{j=1}^{n_{\rm p}} a_{ij} \gamma_j \ge 0.$$
(18.5.4)

Recall from Lemma 18.4.4 that $\mathbf{x}^{(n)} \in \mathcal{F}_+ \cap \mathcal{G}_+$, so $\bar{y}_i^{(n)} > 0$ for $i \in \mathcal{I}_+$ and $x_j + \gamma_j > 0$ for $j \in \mathcal{J}_+$, so the ratios above are all well defined. Because $h_i(l) \triangleq y_i \log l - l$ is concave, by the convexity inequality (see §29.9 and *cf*. §14.6.7) we have

$$\begin{aligned} \mathsf{L}(\boldsymbol{x}) &= \sum_{i=1}^{n_{\rm d}} \mathsf{h}_i(\bar{y}_i(\boldsymbol{x})) = \sum_{i=1}^{n_{\rm d}} \mathsf{h}_i \left(\sum_{j=1}^{n_{\rm p}} \left(\frac{a_{ij}(x_j^{(n)} + \gamma_j)}{\bar{y}_i^{(n)}} \right) \frac{x_j + \gamma_j}{x_j^{(n)} + \gamma_j} \, \bar{y}_i^{(n)} + \left(\frac{\check{r}_i}{\bar{y}_i^{(n)}} \right) \bar{y}_i^{(n)} \right) \\ &\geq Q(\boldsymbol{x}; \boldsymbol{x}^{(n)}) \triangleq \sum_{i=1}^{n_{\rm d}} \sum_{j=1}^{n_{\rm p}} \left(\frac{a_{ij}(x_j^{(n)} + \gamma_j)}{\bar{y}_i^{(n)}} \right) \mathsf{h}_i \left(\frac{x_j + \gamma_j}{x_j^{(n)} + \gamma_j} \, \bar{y}_i^{(n)} \right) + \left(\frac{\check{r}_i}{\bar{y}_i^{(n)}} \right) \mathsf{h}_i (\bar{y}_i^{(n)}) \\ &= \sum_{j=1}^{n_{\rm p}} Q_j(x_j; \boldsymbol{x}^{(n)}), \end{aligned}$$
(18.5.5)

where

$$Q_{j}(x_{j}; \boldsymbol{x}^{(n)}) = \sum_{i=1}^{n_{d}} \left(\frac{a_{ij}(x_{j}^{(n)} + \gamma_{j})}{\bar{y}_{i}^{(n)}} \right) \left[y_{i} \log \left(\frac{x_{j} + \gamma_{j}}{x_{j}^{(n)} + \gamma_{j}} \bar{y}_{i}^{(n)} \right) - \frac{x_{j} + \gamma_{j}}{x_{j}^{(n)} + \gamma_{j}} \bar{y}_{i}^{(n)} \right]$$

$$\stackrel{c}{=} e_{j}(\boldsymbol{x}^{(n)})(x_{j}^{(n)} + \gamma_{j}) \log(x_{j} + \gamma_{j}) - (x_{j} + \gamma_{j})a_{j}.$$
(18.5.6)

The "M-step" (cf. (14.1.1)) for the separable surrogate function Q reduces into n_p 1D maximization problems of the form

$$x_j^{(n+1)} = \operatorname*{arg\,max}_{x_j \ge 0} Q_j(x_j; \boldsymbol{x}^{(n)}).$$

Because

$$\frac{\partial}{\partial x_j}Q_j(x_j;\boldsymbol{x}^{(n)}) = e_j(\boldsymbol{x}^{(n)})\frac{x_j^{(n)} + \gamma_j}{x_j + \gamma_j} - a_j,$$

equating to zero and considering the nonnegativity constraint yields

$$x_{j}^{(n+1)} = \left[\frac{x_{j}^{(n)} + \gamma_{j}}{a_{j}} e_{j}(\boldsymbol{x}^{(n)}) - \gamma_{j}\right]_{+},$$

which is exactly (18.4.4).

s, em1, em, em 18.5.3 Expectation-Maximization (EM) derivation (s, em1, em, em)

This section presents a derivation based on the general form of the EM algorithm described in §14.10.

Recall that to derive an EM algorithm, one must first define a collection of random variables called the **complete data**. This collection need not correspond to anything physical, and indeed it does not in the approach taken here, following [50]. Define the complete data to be the following collection of independently distributed Poisson random variables:

$$\boldsymbol{Z} = \left\{ \{M_{ij}\}_{j=1}^{n_{\rm p}}, B_i \right\}_{i=1}^{n_{\rm d}}$$

where

$$M_{ij} \sim \mathsf{Poisson}\{a_{ij}(x_j + \gamma_j)\}$$
(18.5.7)

$$B_i \sim \mathsf{Poisson}\left\{\bar{r}_i - \sum_{j=1}^{n_{\mathrm{p}}} a_{ij}\gamma_j\right\},\tag{18.5.8}$$

where the condition (18.4.6) ensures that the mean of each B_i is nonnegative. With these definitions, clearly

$$Y_i = \sum_{j=1}^{n_{\rm p}} M_{ij} + B_i$$

has the appropriate distribution (18.2.2). The complete-data log-likelihood is given by

$$\log \mathbf{p}(\mathbf{Z}; \mathbf{x}) \stackrel{c}{=} \sum_{i=1}^{n_{d}} \sum_{j=1}^{n_{p}} M_{ij} \log(a_{ij}(x_{j} + \gamma_{j})) - a_{ij}(x_{j} + \gamma_{j}),$$

ignoring constants independent of x as always, including $p(B_i)$.

By the multinomial property (31.3.4) for conditional distributions of Poisson sums, we have

$$\mathsf{E}[M_{ij} \mid \boldsymbol{Y} = \boldsymbol{y}; \boldsymbol{x}^{(n)}] = \frac{\mathsf{E}[M_{ij}; \boldsymbol{x}^{(n)}]}{\mathsf{E}[y_i; \boldsymbol{x}^{(n)}]} y_i = \begin{cases} y_i \frac{a_{ij}(x_j^{(n)} + \gamma_j)}{\bar{y}_i^{(n)}}, & \bar{y}_i^{(n)} > 0\\ 0, & \bar{y}_i^{(n)} = 0, \ y_i = 0. \end{cases}$$
(18.5.9)

To avoid writing braces, we treat $y_i/y_i^{(n)}$ as zero when both are zero hereafter. So the surrogate "Q" function of the E-step (14.10.5) of the EM algorithm is

$$Q(\boldsymbol{x}; \boldsymbol{x}^{(n)}) = \mathsf{E}[\log \mathsf{p}(\boldsymbol{Z}) | \boldsymbol{Y} = \boldsymbol{y}; \boldsymbol{x}^{(n)}]$$

$$= \sum_{i=1}^{n_{d}} \sum_{j=1}^{n_{p}} \mathsf{E}[M_{ij} | \boldsymbol{Y} = \boldsymbol{y}; \boldsymbol{x}^{(n)}] \log(a_{ij}(x_{j} + \gamma_{j})) - a_{ij}(x_{j} + \gamma_{j})$$

$$= \sum_{i=1}^{n_{d}} \sum_{j=1}^{n_{p}} y_{i} \frac{a_{ij}(x_{j}^{(n)} + \gamma_{j})}{\bar{y}_{i}^{(n)}} \log(a_{ij}(x_{j} + \gamma_{j})) - a_{ij}(x_{j} + \gamma_{j})$$

$$\stackrel{c}{=} \sum_{j=1}^{n_{p}} e_{j}(\boldsymbol{x}^{(n)})(x_{j}^{(n)} + \gamma_{j}) \log(x_{j} + \gamma_{j}) - a_{j}(x_{j} + \gamma_{j}).$$
(18.5.10)

This EM-based surrogate function is identical (to within constants independent of x) to the surrogate (18.5.6) derived by the concavity property using De Pierro's optimization transfer approach. Thus, the M-step (14.10.6) of the EM algorithm is identical to (18.4.4) once again.

In the context of image reconstruction, every EM algorithm I have seen can be derived by an alternate algebraic approach using optimization-transfer principles. In contrast, however, there are many optimization-transfer methods for which no equivalent EM algorithm is apparent. Thus, the optimization-transfer approach seems more fruitful for deriving algorithms than the EM approach, although familiarity with EM-based surrogates can be helpful in designing surrogates by algebraic methods.

e,eml,em,Mij

Ball Market Posinomial derivation (s,eml,em,pos)

Matt Jacobson [59] and Arkadi Nemirovski⁴ independently re-derived the E-ML-EM algorithm by using the (nonobvious) convexity of the **posinomial** functions described below. The derivation in this section is generalized to include the γ_j values and \bar{r}_i values. Despite being a remarkably different approach, it leads to the same iteration once again!

Define the reparameterization

$$y_j = \log(x_j + \gamma_j),$$

for some $\gamma_j > 0$, where clearly $z_j \ge \log \gamma_j$ for $x_j \ge 0$. Then

$$F(\boldsymbol{z}) \triangleq -\mathsf{L}(\mathrm{e}^{\boldsymbol{z}} - \boldsymbol{\gamma}) \stackrel{\mathrm{c}}{=} \sum_{j=1}^{n_{\mathrm{p}}} a_j (\mathrm{e}^{z_j} - \gamma_j) - \sum_{i=1}^{n_{\mathrm{d}}} y_i \log\left(\sum_{j=1}^{n_{\mathrm{p}}} a_{ij} (\mathrm{e}^{z_j} - \gamma_j) + \bar{r}_i\right).$$

By examining the Hessian, one can verify that the following posinomial function

$$f_i(\boldsymbol{z}) \triangleq \log\left(\sum_{j=1}^{n_{\rm p}} a_{ij}(\mathrm{e}^{z_j} - \gamma_j) + \bar{r}_i\right) = \log\left(\sum_{j=1}^{n_{\rm p}} a_{ij}\,\mathrm{e}^{z_j} + \check{r}_i\right)$$

is **convex** in z if the γ_i values satisfy (18.4.6). Thus, by the support property of convex functions (cf. §29.9),

$$f_i(\boldsymbol{z} + \boldsymbol{\delta}) \geq f_i(\boldsymbol{z}) + \nabla f_i(\boldsymbol{z})\boldsymbol{\delta},$$

i.e., the tangent plane can provide a majorizer of $-f_i$. So a surrogate function for F is

$$F(\boldsymbol{z} + \boldsymbol{\delta}) \leq G(\boldsymbol{z} + \boldsymbol{\delta}) \triangleq \sum_{j=1}^{n_{\mathrm{p}}} a_j (\mathrm{e}^{z_j + \delta_j} - \gamma_j) - \sum_{i=1}^{n_{\mathrm{d}}} y_i \left(f_i(\boldsymbol{z}) + \nabla f_i(\boldsymbol{z}) \boldsymbol{\delta} \right)$$

$$\stackrel{\mathrm{c}}{=} \sum_{j=1}^{n_{\mathrm{p}}} a_j \, \mathrm{e}^{z_j} \, \mathrm{e}^{\gamma_j} - \sum_{i=1}^{n_{\mathrm{d}}} y_i \sum_{j=1}^{n_{\mathrm{p}}} \delta_j a_{ij} \, \mathrm{e}^{z_j} / \bar{y}_i(\boldsymbol{z}) = \sum_{j=1}^{n_{\mathrm{p}}} a_j \, \mathrm{e}^{z_j} \, \mathrm{e}^{\gamma_j} - \sum_{j=1}^{n_{\mathrm{p}}} \delta_j \, \mathrm{e}^{z_j} \, e_j(\boldsymbol{z})$$

where $e_j(\cdot)$ was defined in (18.4.3). This approach is also related to the **difference of convex functions** (**DC**) approach to optimization (see §14.9) [60–62]. Now

$$\frac{\partial}{\partial \delta_j} G(\boldsymbol{z} + \boldsymbol{\delta}) = a_j e^{z_j} e^{\delta_j} - e^{z_j} e_j(\boldsymbol{z}),$$

so the unconstrained minimizer over δ is

$$e^{\delta_j^{(n)}} = e_j^{(n)} / a_j.$$

Thus, following (14.1.1), the update is

$$\begin{aligned} x_{j}^{(n+1)} &= \left[e^{z_{j}^{(n+1)}} - \gamma_{j} \right]_{+} = \left[e^{z_{j}^{(n)} + \delta_{j}^{(n)}} - \gamma_{j} \right]_{+} = \left[\left(x_{j}^{(n)} + \gamma_{j} \right) e^{\delta_{j}^{(n)}} - \gamma_{j} \right]_{+} \\ &= \left[\left(x_{j}^{(n)} + \gamma_{j} \right) e_{j}(\boldsymbol{x}^{(n)}) / a_{j} - \gamma_{j} \right]_{+}, \end{aligned}$$

which is exactly (18.4.4). (One easily makes the presentation rigorous by considering carefully the domain of each function involved.)

In this derivation, the sufficient condition (18.4.6) on the γ_j values arises as a sufficient condition for ensuring convexity of each f_i . To verify this convexity, consider:

$$\frac{\partial}{\partial z_k} f_i(\boldsymbol{z}) = \frac{1}{\bar{y}_i} a_{ik} e^{z_k}, \qquad \frac{\partial^2}{\partial z_k \partial z_l} f_i(\boldsymbol{z}) = \frac{1}{\bar{y}_i^2} \begin{cases} -a_{ik} a_{il} e^{z_k} e^{z_l}, & l \neq k \\ \bar{y}_i a_{ik} e^{z_k} - [a_{ik} e^{z_l}]^2, & l = k \end{cases} = \begin{cases} -\alpha_k \alpha_l, & k \neq l \\ \alpha_k - \alpha_k^2, & k = l \end{cases}$$

where $\alpha_k \triangleq \frac{a_{ij} e^{z_k}}{\bar{y}_i^2}$. Thus

$$\sum_{k} \sum_{l} \beta_{k} \beta_{l} \frac{\partial^{2}}{\partial z_{k} \partial z_{l}} f_{i} = \sum_{k} \alpha_{k} \beta_{k}^{2} - \left[\sum_{k} \alpha_{k} \beta_{k}\right]^{2} \ge 0$$

because $\sum_k \alpha_k \leq 1$ using the convexity inequality (29.9.7) since $g(\beta) = \beta^2$ is convex.

⁴In July 2001 email forwarded to me by Matt Jacobson.

s, eml, em, amp 18.5.5 Alternating minimization derivation (s, eml, em, amp)

Several authors have used the alternating minimization approach [63] to derive the classical version of the E-ML-EM algorithm and some regularized variants, *e.g.*, [46, 64]. Most such derivations have ignored the \bar{r}_i values and often make unnecessary assumptions about the a_{ij} values. This section presents an extension of those previous treatments to derive the algorithm (18.4.4) yet again. This derivation is more complicated than using optimization transfer, but is important because it seems to facilitate convergence proofs.

Following §14.8, define the following convex sets

$$\mathcal{P} = \left\{ \boldsymbol{p} = \{ p_{ij} \ge 0 \} : \sum_{j=0}^{n_{p}} p_{ij} = y_{i} \right\}$$
$$\mathcal{Q} = \left\{ \boldsymbol{q} = \{ q_{ij} \} : \exists \boldsymbol{x} \succeq \boldsymbol{0} \text{ s.t. } q_{ij} = q_{ij}(\boldsymbol{x}) \right\}$$

where, using (18.5.4):

$$q_{ij}(\boldsymbol{x}) \triangleq \begin{cases} a_{ij}(x_j + \gamma_j), & j = 1, \dots, n_{\mathrm{p}} \\ \check{r}_i, & j = 0. \end{cases}$$

Also define the divergence

$$\mathsf{D}(\boldsymbol{p} \| \boldsymbol{q}) = \sum_{i=1}^{n_{\rm d}} \sum_{j=0}^{n_{\rm p}} \kappa(p_{ij}, q_{ij}),$$
(18.5.11)

where κ was defined in (18.4.11). Starting with a previous guess $x^{(n)} \in \mathcal{G}_+$, the first step of the alternating minimization algorithm (14.8.2) is

$$\boldsymbol{p}^{(n+1)} = \underset{\boldsymbol{p} \in \mathcal{P}}{\arg\min} \mathsf{D}(\boldsymbol{p} \,\|\, \boldsymbol{q}(\boldsymbol{x}^{(n)})) \,. \tag{18.5.12}$$

Because of the constraints imposed by the set \mathcal{P} , we perform the constrained minimization using Lagrange multipliers via the following augmented functional:

$$F(\boldsymbol{p}) \triangleq \mathsf{D}(\boldsymbol{p} \| \boldsymbol{q}(\boldsymbol{x}^{(n)})) - \sum_{i=1}^{n_{\mathrm{d}}} \lambda_i \left[\sum_{j=0}^{n_{\mathrm{p}}} p_{ij} - y_i \right].$$

Using (18.4.13), for $i \in \mathcal{I}_+$:

$$\frac{\partial}{\partial p_{ij}}F(\boldsymbol{p}) = \log(p_{ij}/q_{ij}(\boldsymbol{x}^{(n)})) - \lambda_i.$$

Equating to zero and solving yields $p_{ij} = q_{ij}(\boldsymbol{x}^{(n)})e^{\lambda_i}$, and applying the constraints we have

$$y_i = \sum_{j=0}^{n_{\rm p}} q_{ij}(\boldsymbol{x}^{(n)}) e^{\lambda_i} = e^{\lambda_i} \, \bar{y}_i^{(n)},$$

where $\bar{y}_i^{(n)}$ was defined in (18.2.3). Thus $e^{\lambda_i} = y_i / \bar{y}_i^{(n)}$, so for $i \in \mathcal{I}_+$:

$$p_{ij}^{(n+1)} = q_{ij}(\boldsymbol{x}^{(n)})y_i/\bar{y}_i^{(n)} = p_{ij}(\boldsymbol{x}^{(n)}),$$

where we define (for $x \in \mathcal{G}_+$):

$$p_{ij}(\boldsymbol{x}) \triangleq \begin{cases} a_{ij}(x_j + \gamma_j)y_i/\bar{y}_i(\boldsymbol{x}), & i \in \mathcal{I}_+, \quad j = 1, \dots, n_{\rm p} \\ \check{r}_i y_i/\bar{y}_i(\boldsymbol{x}), & i \in \mathcal{I}_+, \quad j = 0 \\ 0, & \text{otherwise.} \end{cases}$$
(18.5.13)

The second step of the alternating minimization algorithm (14.8.2) is

$$\boldsymbol{x}^{(n+1)} = \underset{\boldsymbol{x}:\boldsymbol{q}(\boldsymbol{x})\in\mathcal{Q}}{\arg\min} \mathsf{D}(\boldsymbol{p}(\boldsymbol{x}^{(n)}) \| \boldsymbol{q}(\boldsymbol{x})).$$
(18.5.14)

For $oldsymbol{x}\in\mathcal{F}_+\cap\mathcal{G}_+$ and $oldsymbol{z}\in\mathcal{G}_+,$ we have

$$\mathsf{D}(\boldsymbol{p}(\boldsymbol{z}) \| \boldsymbol{q}(\boldsymbol{x})) = \sum_{i=1}^{n_{\mathrm{d}}} \left(\left[\sum_{j=1}^{n_{\mathrm{p}}} \kappa \left(a_{ij}(z_j + \gamma_j) \frac{y_i}{\bar{y}_i(\boldsymbol{z})}, a_{ij}(x_j + \gamma_j) \right) \right] + \kappa \left(\check{r}_i \frac{y_i}{\bar{y}_i(\boldsymbol{z})}, \check{r}_i \right) \right)$$
(18.5.15)

e,eml,amp,vpnn

_

$$= \left(\sum_{j=1}^{n_{\rm p}} \sum_{i=1}^{n_{\rm d}} a_{ij}(z_j + \gamma_j) \frac{y_i}{\bar{y}_i(z)} \log \frac{(z_j + \gamma_j)y_i}{(x_j + \gamma_j)\bar{y}_i(z)} - a_{ij}(z_j + \gamma_j) \frac{y_i}{\bar{y}_i(z)} + a_{ij}(x_j + \gamma_j)\right)$$
(18.5.16)

$$+\left(\sum_{i=1}^{n_{\rm d}}\check{r}_i\frac{y_i}{\bar{y}_i(z)}\log\frac{y_i}{\bar{y}_i(z)}-\check{r}_i\frac{y_i}{\bar{y}_i(z)}+\check{r}_i\right)$$
(18.5.17)

$$=\sum_{j=1}^{n_{\rm p}}\sum_{i=1}^{n_{\rm d}} \left[a_{ij}(z_j+\gamma_j) \frac{y_i}{\bar{y}_i(z)} \log \frac{y_i}{\bar{y}_i(z)} + a_{ij}(z_j+\gamma_j) \frac{y_i}{\bar{y}_i(z)} \log \frac{z_j+\gamma_j}{x_j+\gamma_j} \right]$$
(18.5.18)

$$-\sum_{i=1}^{n_{\rm d}} y_i + \sum_{i=1}^{n_{\rm d}} \check{r}_i \frac{y_i}{\bar{y}_i(\boldsymbol{z})} \log \frac{y_i}{\bar{y}_i(\boldsymbol{z})} + \sum_{i=1}^{n_{\rm d}} \bar{y}_i(\boldsymbol{x})$$
(18.5.19)

$$=\sum_{i=1}^{n_{\rm d}} y_i \log \frac{y_i}{\bar{y}_i(\boldsymbol{z})} + \sum_{j=1}^{n_{\rm p}} e_j(\boldsymbol{z})(z_j + \gamma_j) \log \frac{z_j + \gamma_j}{x_j + \gamma_j} - \sum_{i=1}^{n_{\rm d}} y_i + \sum_{i=1}^{n_{\rm d}} \bar{y}_i(\boldsymbol{x})$$
(18.5.20)

$$=\sum_{i=1}^{n_{\rm d}}\kappa(y_i,\bar{y}_i(\boldsymbol{z})) + \sum_{j=1}^{n_{\rm p}}e_j(\boldsymbol{z})(z_j+\gamma_j)\log\frac{z_j+\gamma_j}{x_j+\gamma_j} - a_j(z_j+\gamma_j) + a_j(x_j+\gamma_j).$$
(18.5.21)

Ignoring irrelevant constants independent of x, this function is equivalent to the optimization transfer surrogate (18.5.6) and the EM-based surrogate (18.5.10). Substituting $z = x^{(n)}$, for $x \in \mathcal{F}_+ \cap \mathcal{G}_+$:

$$\frac{\partial}{\partial x_j} \mathsf{D}(\boldsymbol{p}(\boldsymbol{x}^{(n)}) \,\|\, \boldsymbol{q}(\boldsymbol{x})) = a_j - \frac{x_j^{(n)} + \gamma_j}{x_j + \gamma_j} e_j^{(n)},$$

where $e_j^{(n)}$ was defined in (18.4.3). Equating to zero while enforcing the nonnegativity constraint in Q again leads directly to the iteration (18.4.4).

e,eml,D,pz,qx

18.6 E-ML-EM acceleration

s, eml, em, slow 18.6.1 Slow convergence (s, eml, em, slow)

The E-ML-EM algorithm is notorious for its slow convergence, and the generalization E-ML-EM-3 given in (18.4.4) provides only modest acceleration [50]. Using classical convergence-rate analysis, one can show that the original E-ML-EM algorithm can converge at **sublinear** rates [65]. This slow convergence can be understood through frequency-domain analysis, as described by Tanaka et al. [66]. First rewrite (18.4.2) as follows:

$$\boldsymbol{x}^{(n+1)} = \mathsf{Diag}\left\{1 + \frac{1}{a_j} \sum_{i=1}^{n_d} a_{ij} \frac{y_i - \bar{y}_i(\boldsymbol{x}^{(n)})}{\bar{y}_i(\boldsymbol{x}^{(n)})}\right\} \boldsymbol{x}^{(n)}.$$
(18.6.1)

Now suppose $x^{(n)} = \hat{x} + \delta$ where δ is small and where

$$\mathbf{0} \approx \nabla \mathsf{L}(\hat{\boldsymbol{x}}) = \sum_{i=1}^{n_{\rm d}} a_{ij} \frac{y_i - \bar{y}_i(\hat{\boldsymbol{x}})}{\bar{y}_i(\hat{\boldsymbol{x}})}$$

(This approximation ignores the nonnegativity constraint.) Then the multiplicative correction factor in (18.6.1) is

$$\sum_{i=1}^{n_{d}} a_{ij} \frac{y_{i} - \bar{y}_{i}(\boldsymbol{x}^{(n)})}{\bar{y}_{i}(\boldsymbol{x}^{(n)})} = \sum_{i=1}^{n_{d}} a_{ij} \frac{y_{i} - \bar{y}_{i}(\hat{\boldsymbol{x}}) - [\boldsymbol{A}\boldsymbol{\delta}]_{i}}{\bar{y}_{i}(\boldsymbol{x}^{(n)})} \approx \sum_{i=1}^{n_{d}} a_{ij} \frac{y_{i} - \bar{y}_{i}(\hat{\boldsymbol{x}}) - [\boldsymbol{A}\boldsymbol{\delta}]_{i}}{\bar{y}_{i}(\hat{\boldsymbol{x}})}$$
$$\approx \sum_{i=1}^{n_{d}} a_{ij} \frac{1}{\bar{y}_{i}(\hat{\boldsymbol{x}})} [\boldsymbol{A}\boldsymbol{\delta}]_{i} = \boldsymbol{A}' \operatorname{Diag}\left\{\frac{1}{\bar{y}_{i}(\hat{\boldsymbol{x}})}\right\} \boldsymbol{A}\boldsymbol{\delta}.$$
(18.6.2)

So the correction factor is related to δ after passing δ through the operator $A' \text{Diag}\{\bar{y}_i(\hat{x})\} A$. This operator significantly attenuates high spatial frequencies; for ordinary tomography, A'A has a frequency response of $1/\rho$. Thus the E-ML-EM algorithm adjusts high spatial frequency components more slowly than the low frequencies.

m, accel **18.6.2** Acceleration methods (s,eml,em,accel)

Numerous authors have attempted to accelerate the convergence rate of E-ML-EM by various numerical methods. None of these methods gained wide popularity, either due to potential algorithm instabilities, inconvenience of implementation, and/or the paltry gains that resulted. This section reviews some of those attempts.

18.6.2.1 Algorithm variations

Many acceleration methods modify the iterative algorithm.

18.6.2.1.1 Grid refinement Instead of using the same parameterization (18.2.1) throughout the iterations, one can use a multiresolution strategy with larger pixels in early iterations [67] [68], or a closely related multigrid strategy [69–71].

18.6.2.1.2 Raised powers One can raise the multiplicative correction factor in (18.4.1) to a power to attempt acceleration [66, 72–77].

18.6.2.1.3 Over-relaxation Many forms of **over-relaxation** have been attempted [78–82] including using bigger step sizes [83] and enhancing high-spatial-frequency components [66, 84]. Another variation uses **vector extrapolation**, based on multiple previous iterates, [85]. All such methods have potential problems with stability.

18.6.2.1.4 Line search Most acceleration methods lose the monotonicity property of E-ML-EM. One way to preserve it is to use a modified line search [68, 84, 86]. Enforcing nonnegativity can be more challenging.

18.6.2.2 Implementation tricks

Another family of acceleration methods focus on the implementation.

e,eml,em,slow

18.6.2.2.1 System matrix Because the projection and back-projection operations are the computational bottleneck in E-ML-EM, many methods have been proposed for accelerating these operations such as precomputing A and storing it in sparse format [74, 87, 88], exploiting symmetries of the system matrix, and factoring the system matrix as a product of several simpler terms [12, 72]. Some of these techniques are used routinely.

18.6.2.2.2 Factoring attenuation In PET, by factoring attenuation as a diagonal matrix: $A = \text{Diag}\{c_i\}G$, it partly disappears from the iteration in the ratio part (without randoms) [89]. Letting $a_{ij} = c_i g_{ij}$:

$$e_j(\boldsymbol{x}) = \sum_{i=1}^{n_{\rm d}} a_{ij} \frac{y_i}{\bar{y}_i(\boldsymbol{x})} = \sum_{i=1}^{n_{\rm d}} c_i g_{ij} \frac{y_i}{\sum_{j'=1}^{n_{\rm p}} c_i g_{ij'} x_{j'} + \bar{r}_i} = \sum_{i: y_i^+ > 0} g_{ij} \frac{y_i^+}{\sum_{j'=1}^{n_{\rm p}} g_{ij'} x_{j'} + \bar{r}_i^+}$$

where

$$y_i^+ \triangleq \begin{cases} y_i, & c_i > 0\\ 0, & \text{otherwise,} \end{cases} \quad \overline{r}_i^+ \triangleq \begin{cases} \overline{r}_i/c_i, & c_i > 0\\ 0, & \text{otherwise} \end{cases}$$

This eliminates one array of length $n_{\rm d}$ from the iteration.

18.6.2.2.3 Hardware acceleration Many special-purpose computing hardware and parallel processing methods have been explored (a partial list) [90–99] including using finite-field arithmetic [100].

18.6.2.2.4 Removing low-intensity pixels Instead of updating all pixels every iteration, one might identify pixels that fall below threshold, set them to zero, and exclude from further updates [101].

18.6.2.2.5 Mismatched back-projectors For some system models, implementing the back-projector A' as the exact adjoint (transpose) of the forward projector can be expensive. Many researchers have investigated mismatched back-projectors that are simplified compared to the forward projector [102–104]. Such modifications void the convergence assurances of E-ML-EM, but may be useful in unregularized cases where the iterations are aborted before convergence anyway.

MIRT | The back-projectors in MIRT are all exact transposes, verified using test_adjoint.m

18.7 Block iterative variations

, os, intro 18.7.1 Ordered-subsets methods (s,eml,os,intro)

Of all the iterative image reconstruction methods described in this *book*, the E-ML-OSEM algorithm published⁵ in 1994 by Hudson and Larkin [106] has had perhaps the greatest impact in the field of nuclear medicine. This algorithm was introduced for commercial use in PET and SPECT scanners only a few years after its publication, and remains the *de facto* standard reconstruction method in commercial PET scanners.

The essential idea of the OSEM method is to replace summations over all projection views with summations over only a subset of those views. Hudson and Larkin's paper caught the attention of the imaging community, but with the benefit of hindsight one can usually find precedents. As noted by Lewitt and Muehllehner [78], Tanaka et al. proposed using subsets of projection angles for acceleration [74], as did D. Politte in his 1983 dissertation [67]. Hebert et al. also suggested using ordered subsets for EM [107]. Schmidlin [108] also proposed a related method. The "subsets" ideas in these earlier papers were secondary considerations that did not capture attention the way that Hudson and Larkin's paper did. Independently, Brown et al. [109] and Desmedt et al. [110] used a subset of the acquired measurements to begin iterations prior to the completion of a SPECT or PET scan.

Richard Larkin (personal communication) has described the development of OSEM as something of a fortuitous programming "accident." In the course of developing software to implement the E-ML-EM algorithm, he first implemented a version that updated the image immediately after the reprojection of each view. Later he implemented "correctly" the classical E-ML-EM algorithm but was surprised to find that it gave worse images (in the early iterations, due of course to its slow convergence). The "immediate update" version turns out to be OSEM with 1 view per subset. Of course, as in the famous saying by Louis Pasteur "le hasard ne favorise que les esprits préparés" (chance favors only the prepared mind). Brian Hutton (personal communication) has pointed out that Larkin's work was supported by a grant, held by Malcolm Hudson and Brian from 1990-92, specifically to investigate and develop methods for acceleration. In this context, the "accidental" discovery of OSEM's acceleration was readily noticed!

s, eml, os 18.7.1.1 E-ML-OSEM (s,eml,os)

A simple expression for E-ML-OSEM is obtained by replacing the summations in the E-ML-EM algorithm (18.4.1) with partial summations as follows:

$$x_{j}^{(n+m/M)} = x_{j}^{(n+(m-1)/M)} \frac{1}{\sum_{i \in \mathcal{S}_{m}} a_{ij}} \sum_{i \in \mathcal{S}_{m}} a_{ij} \frac{y_{i}}{\left[\boldsymbol{A}\boldsymbol{x}^{(n+(m-1)/M)}\right]_{i} + \bar{r}_{i}}$$

for m = 1, ..., M, where $\mathbf{x}^{(n+1)} = \mathbf{x}^{(n+M/M)}$. However, the above expression can fail if implemented directly because the partial sensitivity factors

$$s_{mj} \triangleq \sum_{i \in \mathcal{S}_m} a_{ij}$$

can be zero. To rectify this, first define the "preconditioning factors"

$$p_{mj} \triangleq \begin{cases} 1/s_{mj}, & s_{mj} \neq 0\\ 0, & \text{otherwise.} \end{cases}$$

Next define the partial E-step factors

$$e_{mj}(\boldsymbol{x}) \triangleq \sum_{i \in \mathcal{S}_m} a_{ij} \frac{y_i}{[\boldsymbol{A}\boldsymbol{x}]_i + \bar{r}_i},$$

where 0/0 is treated as zero. Then E-ML-OSEM is defined by

$$\boldsymbol{x}^{(n+m/M)} = \mathcal{M}^{(m)} \left(\boldsymbol{x}^{(n+(m-1)/M)} \right)$$

where

$$\mathcal{M}_{j}^{(m)}(\boldsymbol{x}) \triangleq \begin{cases} x_{j} p_{mj} e_{mj}(\boldsymbol{x}), & p_{mj} \neq 0\\ x_{j}, & \text{otherwise} \end{cases} = x_{j} + x_{j} p_{mj} \left[e_{mj}(\boldsymbol{x}) - s_{mj} \right].$$

Because the algorithm is a multiplicative update, if the initial estimate is positive, then all subsequent iterates also remain positive.

⁵It was presented in 1992 by Hudson, Hutton and Larkin [105] which means that the idea was developed by late 1991.

If we define the partial log-likelihood function

$$\mathsf{L}_m(oldsymbol{x}) = \sum_{i\in\mathcal{S}_m}\mathsf{h}_i([oldsymbol{A}oldsymbol{x}]_i),$$

then one can verify that

$$e_{mj}(\boldsymbol{x}) - s_{mj} = \frac{\partial}{\partial x_j} \mathsf{L}_m(\boldsymbol{x})$$

Thus E-ML-OSEM is an incremental gradient algorithm with diagonal scaling matrices $D_m(x) = \text{Diag}\{x_j p_{mj}\}$. As shown in [48], subiteration-dependent scaling matrices preclude convergence even when relaxation is included.

Although E-ML-OSEM fails to converge in general, it often gives reasonable looking images in a small number of iterations when initialized with a uniform image.

18.7.1.2 RBI

Several variants of E-ML-OSEM have been proposed, including the rescaled block-iterative (RBI) method [111–116].

18.7.1.3 RAMLA

See also the related variants: the row-action maximum likelihood algorithm (RAMLA) [117]

$$x_{j}^{(n+1,k)} = x_{j}^{(n,k)}(1 - \lambda_{k}a_{ij}) + x_{j}^{(n,k)}\lambda_{k}\sum_{i \in \mathcal{S}_{n}} a_{ij}\frac{y_{i}}{\bar{y}_{i}(\boldsymbol{x}^{(n,k)})} = \lambda_{k}x_{j}^{(n,k)}\sum_{i \in \mathcal{S}_{n}} a_{ij}\left(\frac{y_{i}}{\bar{y}_{i}(\boldsymbol{x}^{(n,k)})} - 1\right)$$

where $0 < \lambda_k a_{ij} \le 1$ whenever $a_{ij} \ne 0$. This iteration converges to an ML estimate if $\lambda_k \rightarrow 0$ yet $\sum_k \lambda_k = \infty$.

For generalizations of these block iterative EM-type algorithms, see **dynamic RAMLA** (**DRAMA**) [118] in which the relaxation parameters change not only with iteration but also between each subset update. See also [119].

18.7.1.4 BSREM

A generalization of RAMLA is the **block sequential regularized expectation maximization (BSREM)** algorithm [48, 117, 120].

s, eml, em, inc 18.7.2 Incremental EM methods (s, eml, em, inc)

As described in §14.10.8, an interesting generalization of the EM approach is the **incremental EM algorithm** proposed by Neal and Hinton [121], which has been shown to converge [122], albeit non-monotonically in $L(\cdot)$ [123]. It has been applied to PET [124, 125]. Here we describe a slight generalization for ML emission reconstruction.

First let S_1, \ldots, S_M denote a partition of the data indices $\{1, \ldots, n_d\}$. Extending the notation in §18.5.3, define

$$Z_m = \left\{ \{M_{ij}\}_{j=1}^{n_{\rm p}}, B_i : i \in \mathcal{S}_m \right\}, \qquad m = 1, \dots, M_n$$

where M_{ij} was defined in (18.5.7), and

$$\boldsymbol{Y}_m = \{Y_i : i \in \mathcal{S}_m\}, \qquad m = 1, \dots, M.$$

It follows from the construction of Z and Poisson statistical properties that

$$\mathsf{P}\{\boldsymbol{Z} = \boldsymbol{z} \mid \boldsymbol{Y} = \boldsymbol{y}; \boldsymbol{x}\} = \prod_{m=1}^{M} \mathsf{P}\{\boldsymbol{Z}_{m} = \boldsymbol{z}_{m} \mid \boldsymbol{Y}_{m} = \boldsymbol{y}_{m}; \boldsymbol{x}\}$$
$$\mathsf{P}\{\boldsymbol{Z}_{m} = \boldsymbol{z}_{m} \mid \boldsymbol{Y}_{m} = \boldsymbol{y}_{m}; \boldsymbol{x}\} = \prod_{i \in \mathcal{S}_{m}} \mathsf{P}\Big\{\{M_{ij} = m_{ij}\}_{j=1}^{n_{\mathsf{P}}}, B_{i} = b_{i} \mid Y_{i} = y_{i}; \boldsymbol{x}\Big\},$$

where

$$\mathsf{P}\Big\{\{M_{ij} = m_{ij}\}_{j=1}^{n_{\rm p}}, B_i = b_i \,|\, Y_i = y_i; \boldsymbol{x}\Big\}$$

© J. Fessler. [license] August 13, 2020

$$= \begin{pmatrix} y_{i} \\ m_{i1} \dots m_{in_{p}} b_{i} \end{pmatrix} \left(\frac{\bar{r}_{i} - \sum_{j=1}^{n_{p}} a_{ij} \gamma_{j}}{\bar{y}_{i}(\boldsymbol{x})} \right)^{b_{i}} \prod_{j=1}^{n_{p}} \left(\frac{a_{ij}(x_{j} + \gamma_{j})}{\bar{y}_{i}(\boldsymbol{x})} \right)^{m_{ij}} \mathbb{I}_{\{y_{i} = b_{i} + \sum_{j=1}^{n_{p}} m_{ij}\}}$$

$$= \begin{pmatrix} y_{i} \\ m_{i1} \dots m_{in_{p}} b_{i} \end{pmatrix} \left(\frac{1}{\bar{y}_{i}(\boldsymbol{x})} \right)^{y_{i}} \left(\bar{r}_{i} - \sum_{j=1}^{n_{p}} a_{ij} \gamma_{j} \right)^{b_{i}} \prod_{j=1}^{n_{p}} (a_{ij}(x_{j} + \gamma_{j}))^{m_{ij}} \mathbb{I}_{\{y_{i} = b_{i} + \sum_{j=1}^{n_{p}} m_{ij}\}}.$$

$$(18.7.1)$$

Thus

$$\frac{\mathsf{P}\{\boldsymbol{Z}_m = \boldsymbol{z}_m \,|\, \boldsymbol{Y}_m = \boldsymbol{y}_m; \bar{\boldsymbol{x}}\}}{\mathsf{P}\{\boldsymbol{Z}_m = \boldsymbol{z}_m \,|\, \boldsymbol{Y}_m = \boldsymbol{y}_m; \boldsymbol{x}\}} = \prod_{i \in \mathcal{S}_m} \left(\frac{\bar{y}_i(\boldsymbol{x})}{\bar{y}_i(\bar{\boldsymbol{x}})}\right)^{y_i} \prod_{j=1}^{n_{\mathrm{p}}} \left(\frac{\bar{x}_j + \gamma_j}{x_j + \gamma_j}\right)^{m_{ij}}$$

and hence

$$\log \frac{\mathsf{P}\{\boldsymbol{Z}_m = \boldsymbol{z}_m \,|\, \boldsymbol{Y}_m = \boldsymbol{y}_m; \bar{\boldsymbol{x}}\}}{\mathsf{P}\{\boldsymbol{Z}_m = \boldsymbol{z}_m \,|\, \boldsymbol{Y}_m = \boldsymbol{y}_m; \boldsymbol{x}\}} = \sum_{i \in \mathcal{S}_m} \left[y_i \log \frac{\bar{y}_i(\boldsymbol{x})}{\bar{y}_i(\bar{\boldsymbol{x}})} - \sum_{j=1}^{n_{\mathrm{p}}} m_{ij} \log \left(\frac{x_j + \gamma_j}{\bar{x}_j + \gamma_j}\right) \right],$$

so

$$\begin{split} \mathsf{D}_{m}(\bar{\boldsymbol{x}} \parallel \boldsymbol{x}) &\triangleq \mathsf{D}(\mathsf{P}\{\boldsymbol{Z}_{m} = \boldsymbol{z}_{m} \mid \boldsymbol{Y}_{m} = \boldsymbol{y}_{m}; \bar{\boldsymbol{x}}\} \parallel \mathsf{P}\{\boldsymbol{Z}_{m} = \boldsymbol{z}_{m} \mid \boldsymbol{Y}_{m} = \boldsymbol{y}_{m}; \boldsymbol{x}\}) \\ & \stackrel{\mathrm{c}}{=} \sum_{i \in \mathcal{S}_{m}} y_{i} \log \frac{\bar{y}_{i}(\boldsymbol{x})}{\bar{y}_{i}(\bar{\boldsymbol{x}})} - \sum_{j=1}^{n_{\mathrm{p}}} \left[\sum_{i \in \mathcal{S}_{m}} \mathsf{E}[M_{ij} \mid Y_{i} = y_{i}; \bar{\boldsymbol{x}}] \right] \log \left(\frac{x_{j} + \gamma_{j}}{\bar{x}_{j} + \gamma_{j}} \right) \\ & = \sum_{i \in \mathcal{S}_{m}} y_{i} \log \frac{\bar{y}_{i}(\boldsymbol{x})}{\bar{y}_{i}(\bar{\boldsymbol{x}})} - \sum_{j=1}^{n_{\mathrm{p}}} f_{mj}(\bar{\boldsymbol{x}}) \log \left(\frac{x_{j} + \gamma_{j}}{\bar{x}_{j} + \gamma_{j}} \right), \end{split}$$

where we define

$$f_{mj}(\boldsymbol{x}) \triangleq \sum_{i \in \mathcal{S}_m} \mathsf{E}[M_{ij} \mid Y_i = y_i; \boldsymbol{x}] = (x_j + \gamma_j) e_{mj}(\boldsymbol{x}), \quad e_{mj}(\boldsymbol{x}) \triangleq \sum_{i \in \mathcal{S}_m} a_{ij} \frac{y_i}{\bar{y}_i(\boldsymbol{x})}.$$
(18.7.2)

Defining the partial negative log-likelihood

$$\mathbf{L}_m(\boldsymbol{x}) \triangleq \left[\sum_{i \in \mathcal{S}_m} \bar{y}_i(\boldsymbol{x}) - y_i \log \bar{y}_i(\boldsymbol{x})\right],$$

we see that

$$f_{mj}(\boldsymbol{x}) = (x_j + \gamma_j) \left[\frac{\partial}{\partial x_j} \mathbf{L}_m(\boldsymbol{x}) - \sum_{i \in \mathcal{S}_m} a_{ij} \right].$$
(18.7.3)

Now consider the augmented cost function

$$F(\boldsymbol{x}, \bar{\boldsymbol{x}}_1, \dots, \bar{\boldsymbol{x}}_M) = \boldsymbol{\mathsf{L}}(\boldsymbol{x}) + \sum_{m=1}^M \mathsf{D}_m(\bar{\boldsymbol{x}}_m \| \boldsymbol{x}) = \sum_{m=1}^M \left[\boldsymbol{\mathsf{L}}_m(\boldsymbol{x}) + \mathsf{D}_m(\bar{\boldsymbol{x}}_m \| \boldsymbol{x}) \right],$$

where

$$\begin{aligned} \mathbf{L}_{m}(\boldsymbol{x}) + \mathsf{D}_{m}(\bar{\boldsymbol{x}}_{m} \| \boldsymbol{x}) &= \sum_{i \in \mathcal{S}_{m}} \bar{y}_{i}(\boldsymbol{x}) - y_{i} \log \bar{y}_{i}(\boldsymbol{x}) + \sum_{i \in \mathcal{S}_{m}} y_{i} \log \frac{\bar{y}_{i}(\boldsymbol{x})}{\bar{y}_{i}(\bar{\boldsymbol{x}}_{m})} - \sum_{j=1}^{n_{\mathrm{p}}} f_{mj}(\bar{\boldsymbol{x}}_{m}) \log \left(\frac{x_{j} + \gamma_{j}}{\bar{x}_{mj} + \gamma_{j}}\right) \\ &\stackrel{c}{=} \sum_{i \in \mathcal{S}_{m}} \bar{y}_{i}(\boldsymbol{x}) - \sum_{j=1}^{n_{\mathrm{p}}} f_{mj}(\bar{\boldsymbol{x}}_{m}) \log(x_{j} + \gamma_{j}) \,. \end{aligned}$$

Because $\mathsf{D}(\bar{x}_m \| x)$ is minimized over \bar{x}_m when $\bar{x}_m = x$, the ML estimate can be expressed as follows

$$\hat{\boldsymbol{x}} = \operatorname*{arg\,min}_{\boldsymbol{x} \succeq \boldsymbol{0}} \min_{\bar{\boldsymbol{x}}_1, \dots, \bar{\boldsymbol{x}}_M} F(\boldsymbol{x}, \bar{\boldsymbol{x}}_1, \dots, \bar{\boldsymbol{x}}_M).$$

This suggests the following iterative algorithm. Start with an initial collection of estimates $\{x^{(0+(m-1)/M)}\}\$ for m = 1, ..., M, and apply a block-coordinate descent approach to $F(\cdot)$ as follows.

for m = 1, ..., M:

$$\boldsymbol{x}^{(n+1+(m-1)/M)} = \operatorname*{arg\,min}_{\boldsymbol{x} \succeq \boldsymbol{0}} F\left(\boldsymbol{x}, \boldsymbol{x}^{(n+1+0/M)}, \dots, \boldsymbol{x}^{(n+1+(m-2)/M)}, \boldsymbol{x}^{(n+(m-1)/M)}, \dots, \boldsymbol{x}^{(n+(M-1)/M)}\right)$$

For minimization over x, observe that

$$\frac{\partial}{\partial x_j}F(\boldsymbol{x},\bar{\boldsymbol{x}}_1,\ldots,\bar{\boldsymbol{x}}_M) = \sum_{m=1}^M \sum_{i\in\mathcal{S}_m} a_{ij} - \frac{1}{x_j + \gamma_j} \sum_{m=1}^M f_{mj}(\bar{\boldsymbol{x}}_m) = a_j - \frac{1}{x_j + \gamma_j} \sum_{m=1}^M f_{mj}(\bar{\boldsymbol{x}}_m),$$

where a_j was defined in (18.2.4). Equating to zero (while considering the **KKT conditions**), one can implement this algorithm efficiently as follows. As usual, $\mathbf{x}^{(n+1)} = \mathbf{x}^{(n+M/M)}$.

Incremental emission EM algorithm (E-ML-INC-EM-3) Initialize $\{x^{(0+(m-1)/M)} : m = 1, ..., M\}$ using, for example, subiterates of E-ML-OS-EM. Initialize $\{f_{mj}\}$ using (18.7.2) as follows: $f_{mj} := f_{mj}(x^{(0+(m-1)/M)}), \quad m = 1, ..., M, \quad j = 1, ..., n_p$ Initialize the following sum: $f_j := \sum_{m=1}^M f_{mj}, \quad j = 1, ..., n_p.$ for n = 1, 2, ...for m = 1, ..., M $x_j^{(n+1+(m-1)/M)} = \left[\frac{f_j}{a_j} - \gamma_j\right]_+, \quad j = 1, ..., n_p$ $f_j = f_{mj}$ $f_{mj} := f_{mj}(x^{(n+1+(m-1)/M)})$ $f_j + f_{mj}.$ (18.7.5) end end

MIRT See eml_inc_em.

This algorithm monotonically decreases $F(\cdot)$. For convergence analysis, see [124, 125]. Empirically, it has been found to converge faster than E-ML-EM, albeit non-monotonically in L(x). However, in the early iterations the OS-EM algorithm tends to "converge" faster, so the incremental EM method is perhaps most useful after first applying several OS-EM iterations, if one wishes to converge to the ML estimate.

Like OS-EM, in every subiteration (18.7.4), the entire parameter vector x is updated. In contrast, the SAGE algorithm only updates a subset of the parameters for each subiteration. Whereas OS-EM only uses a subset of the data for each subiteration, the incremental EM method always uses the most recent estimates of all of the complete data. This incremental approach requires one to store M image-sized arrays f_{m_i} .

Using (18.7.3), we can also write the update (18.7.4) in terms of the gradients of L_m , but the expression seems not particularly insightful.

18.8 Alternative algorithms (s,eml,alg)

A variety of other algorithms largely unrelated to EM methods have been applied to emission image reconstruction.

alg, opt 18.8.1 General-purpose optimization methods

Several of the general-purpose optimization methods discussed in Chapter 11 have been investigated, including: steepest descent with step size chosen approximately [126], CG and PCG [42].

More advance optimization algorithms have also been studied, such as primal-dual methods [127].

18.8.2 Special-purpose methods

Lantéri et al. present a general method to devise maximum-likelihood multiplicative algorithms with non-negativity constraints [77, 133]. First write the cost function as follows: $\nabla \mathbf{L}(\mathbf{x}) = \mathbf{u}(\mathbf{x}) - \mathbf{v}(\mathbf{x})$, where \mathbf{u} and \mathbf{v} are both positive. Then try the iteration $\mathbf{x}^{(n+1)} = \mathbf{x}^{(n)} \mathbf{u}^k(\mathbf{x}^{(n)}) / \mathbf{v}^k(\mathbf{x}^{(n)})$, where k is a power (applied element-wise) that affects convergence. Choosing $\mathbf{u}(\mathbf{x}) = \mathbf{A}'(\mathbf{y}/\mathbf{A}\mathbf{x})$, $\mathbf{v}(\mathbf{x}) = \mathbf{A}'\mathbf{1}_{n_d}$ yields the conventional ML-EM algorithm.

Lane [134] proposes to let $x_j = z_j^2$ and to use gradient descent or a CG algorithm to optimize with respect to z_j . The usual gradient descent update would be

$$\boldsymbol{z}^{(n+1)} = \boldsymbol{z}^{(n)} - \alpha_n \, \nabla \Psi(\boldsymbol{z}^{(n)}) = \boldsymbol{z}^{(n)} - \alpha_n 2 \, \mathsf{Diag}\{\boldsymbol{z}^{(n)}\} \, \boldsymbol{A}' \, (\boldsymbol{y} \oslash \, \bar{\boldsymbol{y}}(\boldsymbol{z}^{(n)}) - 1)$$

This is remarkable similar to E-ML-EM, but it loses the guarantee of monotonicity and of ensuring nonnegativity of the iterates. But the final step will be to let $\hat{x}_j = \hat{z}_j^2$ which enforces nonnegativity at the end. Being a gradient descent method, one must use a line search to find α_n to ensure continuity.

s, eml, nested **18.8.3** Nested or cascade EM algorithms (s, eml, nested)

In some applications, the image x may itself be modeled as a function of other parameters, *e.g.*, $x = B\theta$ where B is a blurring matrix, or a kinetic model $x(\theta)$ for radiotracers [135–137]. In these applications one can derive a **nested EM algorithm** or **cascade EM algorithm** [138]. For the E-step, one first (re)derives the Q function in (18.5.10), but now written as follows:

$$Q(\boldsymbol{\theta};\boldsymbol{\theta}^{(n)}) = \sum_{j=1}^{n_{\rm p}} e_j(\boldsymbol{x}^{(n)}) \left(x_j^{(n)} + \gamma_j \right) \log(x_j(\boldsymbol{\theta}) + \gamma_j) - a_j \left(x_j(\boldsymbol{\theta}) + \gamma_j \right),$$

where $x^{(n)} \triangleq x(\theta^{(n)})$. Then for the M-step one maximizes Q with respect to the parameters θ :

$$\boldsymbol{\theta}^{(n+1)} = \operatorname*{arg\,max}_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(n)}).$$

In general this maximization also requires an iterative algorithm.

One can also use an inner EM loop to compute a search direction followed by a line-search for an outer (P)CG loop [135, 139]. In this form, the EM inner loop serves as a type of **implicit preconditioner**.

18.9 E-ML-EM modifications

18.9.1 Stopping rules (s,eml,em,stop)

For typical emission tomography scans, unregularized algorithms like E-ML-EM lead to very noisy images as the iterations proceed. Several papers have described **stopping rules** that describe criteria for terminating the iteration before the noise becomes too excessive. Examples include [141–151].

The results of such methods are completely algorithm dependent, because the trajectory $\{x^{(n)}\}$ taken from the initial guess $x^{(0)}$ to the limit \hat{x} depends on the type of iteration. In particular, because of effects like the "Diag $\{1/\bar{y}_i(\hat{x})\}$ " term in (18.6.2), the spatial-resolution properties of iterations like E-ML-EM converge nonuniformly, so terminating prematurely leads to nonuniform spatial-resolution properties. For such reasons, the trend of late is to run many iterations (particularly with OSEM), thereby hopefully largely eliminating the algorithm-dependent component, and then post-filter to reduce noise—thereby maintaining good control of the spatial-resolution properties.

s, eml, ems 18.9.2 Smoothing and EMS (s, eml, ems)

s,eml,em,stop

Many variations of the ML-EM algorithm have been proposed that involve smoothing in an attempt to reduce noise, including sieves [152], inter-update filtering [153, 154], post-backprojection filtering [155], and the expectationmaximization smooth (EMS) approach [156–163]. Of all such methods, the one used most frequently in practice is to "over iterate" until the image becomes a bit "too noisy" and then post-filter the final iterate to reduce noise.

18.10 E-ML-EM convergence proof (s,eml,em,proof)

An elegant proof of convergence of E-ML-EM-3 is given in [43].

18.11 Problems (s,eml,prob)

Problem 18.1 Show that κ is jointly convex in u and v over $(0, \infty) \times (0, \infty)$.

Problem 18.2 For simplicity, consider $\bar{r}_i = 0$ and consider trying to generalize (18.5.3) as follows:

$$[\mathbf{A}\mathbf{x}]_{i} = \sum_{j=1}^{n_{\mathrm{p}}} a_{ij} x_{j} = \sum_{j=1}^{n_{\mathrm{p}}} \left(\frac{b_{ij} x_{j}^{(n)}}{[\mathbf{B}\mathbf{x}^{(n)}]_{i}} \right) \left[\frac{a_{ij} x_{j}}{b_{ij} x_{j}^{(n)}} [\mathbf{B}\mathbf{x}^{(n)}]_{i} \right]$$

where $b_{ij} \ge 0$ and b_{ij} is zero only if a_{ij} is zero. (But in general **B** could be different than **A**.) The terms in parentheses sum to unity, so we have the inequality

$$\psi_i([\boldsymbol{A}\boldsymbol{x}]_i) \leq \sum_{j=1}^{n_{\mathrm{p}}} \left(\frac{b_{ij} x_j^{(n)}}{[\boldsymbol{B}\boldsymbol{x}^{(n)}]_i} \right) \psi_i \left(\frac{a_{ij} x_j}{b_{ij} x_j^{(n)}} [\boldsymbol{B}\boldsymbol{x}^{(n)}]_i \right).$$

Can we use this to design a surrogate function and hence a new, broader family of optimization transfer methods? *Explain why or why not.*

Problem 18.3 Prove (18.7.1).

Problem 18.4 Veklerov [189] proposes an EM algorithm for estimating the difference (assumed to be nonnegative) of two emission distributions, from the model $Y_i^1 \sim \text{Poisson}\{\sum_{j=1}^{n_p} a_{ij}x_j\}, Y_i^2 \sim \text{Poisson}\{\sum_{j=1}^{n_p} a_{ij}(x_j + \delta_j)\}$, where the goal is to estimate $\delta_j \ge 0$. He used the complete data spaces $\{X_{ij}^1, X_{ij}^2, X_{ij}^3\}$ where $X_{ij}^1 \sim \text{Poisson}\{a_{ij}x_j\}, X_{ij}^2 \sim \text{Poisso}\{a_{ij}x_j\}, X_{ij}^2 \sim \text{Poiss}\{a_{ij}x_j\}, X_{ij}^2 \sim \text{Poiss}\{a_{ij}x_j\}, X_{ij}^2 \sim \text{Poiss}\{a_{ij}x_j\}, X_{ij}^2 \sim \text{Poiss}$

Problem 18.5 From (18.4.16) it is clear that minimizing $\sum_{i=1}^{n_d} \kappa(y_i, \bar{y}_i(\boldsymbol{x}))$ yields a ML estimate. Suppose instead we minimize $\sum_{i=1}^{n_d} \kappa(\bar{y}_i(\boldsymbol{x}), y_i)$, for some ad hoc reason. Derive a monotone iterative algorithm for this cost function.

Problem 18.6 Study the α -EM algorithm of [191] and derive an iterative algorithm for the Poisson log-likelihood (18.2.2) based on α -EM. The possibility of such a derivation (for list-mode data) is mentioned in the closing paragraph of [192]. Hint: see (31.3.5).

Problem 18.7 Apply the optimization transfer derivation of §18.5.2 to the list-mode likelihood function (8.5.10) to derive a list-mode E-ML-EM algorithm.

(Need typed.)

(Need typed.)

18.12 Bibliography

- T. J. Holmes. "Blind deconvolution of quantum-limited incoherent imagery: maximum-likelihood approach." In: J. Opt. Soc. Am. A 9.7 (July 1992), 1052–61. DOI: 10.1364/JOSAA.9.001052 (cit. on p. 18.2).
- [2] W. A. Wheaton et al. "Multiparameter linear least-squares fitting to Poisson data one count at a time." In: *The Astrophysical Journal* 438.1 (Jan. 1995), 322–40. DOI: 10.1086/175077 (cit. on pp. 18.2, 18.5).
- [3] M. I. Miller et al. "Maximum-likelihood estimation applied to electron-microscope autoradiography." In: J. of Electron Microscopy Technique 2.6 (1985), 611–36. DOI: 10.1002/jemt.1060020612 (cit. on p. 18.2).
- [4] D. Xu and Z. He. "Gamma-ray energy-imaging integrated spectral deconvolution." In: *Nucl. Instr. Meth. Phys. Res. A.* 574.1 (Apr. 2007), 98–109. DOI: 10.1016/j.nima.2007.01.171 (cit. on p. 18.2).
- [5] L. B. White. "On the convergence of the estimate posterior density weighted algorithm." In: *IEEE Trans. Sig. Proc.* 39.1 (Jan. 1991), 254–5. DOI: 10.1109/78.80800 (cit. on p. 18.2).

18.22

s,eml,prob p,kl,convex

p,eml,surr

p,eml,inc,mby

.eml.kl.back.min

p,eml,diff

holmes:92:bdo

wheaton:95:mll

miller:85:mle

xu:07:gre

white:91:otc

wells:00:cfb

[7] J. Nuyts et al. "Reducing loss of image quality due to the attenuation artifact in uncorrected PET whole body images." In: *J. Nuc. Med.* 43.8 (Aug. 2002), 1054–62. URL:

http://jnm.snmjournals.org/cgi/content/abstract/43/8/1054 (cit.on p. 18.2).

- [8] R. G. Wells et al. "Comparing filtered backprojection and ordered-subsets expectation maximization for small lesion detection and localization in Ga-67 SPECT." In: J. Nuc. Med. 41.8 (Aug. 2000), 1391–9. URL: http://jnm.snmjournals.org/cgi/content/abstract/41/8/1391 (cit. on p. 18.2).
- [9] M. V. Narayanan et al. "Human-observer receiver-operating-characteristic evaluation of attenuation, scatter, and resolution compensation strategies for 99mTc myocardial perfusion imaging." In: J. Nuc. Med. 44.11 (Nov. 2003), 1725–34. URL:
 - http://jnm.snmjournals.org/cgi/content/abstract/44/11/1725 (cit. on p. 18.2).
- [10] E. Asma et al. "Accurate and consistent lesion quantitation with clinically acceptable penalized likelihood images." In: *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.* 2012, 4062–66. DOI: 10.1109/NSSMIC.2012.6551928 (cit. on p. 18.2).
- [11] S. Ahn et al. "Quantitative comparison of OSEM and penalized likelihood image reconstruction using relative difference penalties for clinical PET." In: *Phys. Med. Biol.* 60.15 (Aug. 2015), 5733–52. DOI: 10.1088/0031-9155/60/15/5733 (cit. on p. 18.2).
 - [12] J. Qi et al. "High resolution 3D Bayesian image reconstruction using the microPET small-animal scanner." In: *Phys. Med. Biol.* 43.4 (Apr. 1998), 1001–14. DOI: 10.1088/0031-9155/43/4/027 (cit. on pp. 18.2, 18.3, 18.16).
- [13] A. J. Rockmore and A. Macovski. "A maximum likelihood approach to emission image reconstruction from projections." In: *IEEE Trans. Nuc. Sci.* 23.4 (Aug. 1976), 1428–32. DOI: 10.1109/TNS.1976.4328496 (cit. on p. 18.2).
 - [14] F. L. Vermeulen. "Comments on a maximum likelihood approach to emission image reconstruction from projections." In: *IEEE Trans. Nuc. Sci.* 29.5 (Oct. 1982), p. 1384. DOI: 10.1109/TNS.1982.4336358 (cit. on p. 18.2).
 - [15] A. J. Rockmore. "Reply to comments on a maximum likelihood approach to emission image reconstruction from projections." In: *IEEE Trans. Nuc. Sci.* 29.5 (Oct. 1982), p. 1385. DOI: 10.1109/TNS.1982.4336359 (cit. on p. 18.2).
 - [16] L. A. Shepp and Y. Vardi. "Maximum likelihood reconstruction for emission tomography." In: *IEEE Trans. Med. Imag.* 1.2 (Oct. 1982), 113–22. DOI: 10.1109/TMI.1982.4307558 (cit. on pp. 18.2, 18.10).
 - [17] D. L. Snyder and D. G. Politte. "Image reconstruction from list-mode data in emission tomography system having time-of-flight measurements." In: *IEEE Trans. Nuc. Sci.* 20.3 (June 1983), 1843–9. DOI: 10.1109/TNS.1983.4332660 (cit. on p. 18.2).
 - [18] D. L. Snyder. "Parameter estimation for dynamic studies in emission-tomography systems having list-mode data." In: *IEEE Trans. Nuc. Sci.* 31.2 (Apr. 1984), 925–31. DOI: 10.1109/TNS.1984.4333400 (cit. on p. 18.2).
 - [19] D. L. Snyder. "Utilizing side information in emission tomography." In: *IEEE Trans. Nuc. Sci.* 31.1 (Feb. 1984), 533–7. DOI: 10.1109/TNS.1984.4333313 (cit. on p. 18.2).
 - [20] L. A. Shepp et al. "Maximum likelihood PET with real data." In: *IEEE Trans. Nuc. Sci.* 31.2 (Apr. 1984), 910–3. DOI: 10.1109/TNS.1984.4333398 (cit. on p. 18.2).
 - [21] D. G. Politte and D. L. Snyder. "Results of a comparative study of a reconstruction procedure for producing improved estimates of radioactivity distributions in time-of-flight emission tomography." In: *IEEE Trans. Nuc. Sci.* 31.1 (Feb. 1984), 614–9. DOI: 10.1109/TNS.1984.4333330 (cit. on p. 18.2).
 - M. I. Miller, D. L. Snyder, and T. R. Miller. "Maximum-likelihood reconstruction for single-photon emission computed-tomography." In: *IEEE Trans. Nuc. Sci.* 32.1 (Feb. 1985), 769–78. DOI: 10.1109/TNS.1985.4336939 (cit. on p. 18.2).
 - [23] C. T. Chen and C. E. Metz. "A simplified EM reconstruction for TOFPET." In: *IEEE Trans. Nuc. Sci.* 32.1 (Feb. 1985), 885–8. DOI: 10.1109/TNS.1985.4336960 (cit. on p. 18.2).

richardson:72:bbi	[24]	W. H. Richardson. "Bayesian-based iterative method of image restoration." In: <i>J. Opt. Soc. Am.</i> 62.1 (Jan. 1972), 55–9. DOI: 10.1364/JOSA.62.000055 (cit. on p. 18.2).
lucy:74:ait	[25]	L. Lucy. "An iterative technique for the rectification of observed distributions." In: <i>The Astronomical Journal</i> 79.6 (June 1974), 745–54. URL: http://adsabs.harvard.edu/cgi-bin/nph-bib_query?bibcode=1974AJ79745L (cit. on p. 18.2).
metz:/l:nan	[26]	C. E. Metz and S. M. Pizer. Nonstationary and nonlinear scintigram processing. 1971 (cit. on p. 18.2).
pizer:73:ron	[27]	S. M. Pizer et al. "Results of nonlinear and nonstationary processing." In: <i>Proc. Third Intl. Conf. on Data Handling and Image Processing in Scintigraphy</i> . Cambridge, MA, June 6-9, 1973. 1973, 93–101 (cit. on p. 18.2).
pizer:03:tmi	[28]	S. M. Pizer. "The medical image display and analysis group at the University of North Carolina: Reminiscences and philosophy." In: <i>IEEE Trans. Med. Imag.</i> 22.1 (Jan. 2003), 2–10. DOI: 10.1109/TMI.2003.809707 (cit. on p. 18.2).
darroch:72:gis	[29]	J. N. Darroch and D. Ratcliff. "Generalized iterative scaling for log-linear models." In: <i>Ann. Math. Stat.</i> 43.5 (Oct. 1972), 1470–80. DOI: 10.1214/aoms/1177692379 (cit. on p. 18.3).
kosarev:90:ssl	[30]	E. L. Kosarev. "Shannon's superresolution limit for signal recovery." In: <i>Inverse Prob.</i> 6.1 (Feb. 1990). Correction in 6(3):479, June 1990., 55–76, DOI: 10, 1088/0266–5611/6/1/007 (cit. on p. 18.3).
digesu:84:bdd	[31]	V. Di Gesu and M. C. Maccarone. "Bayesian direct deconvolution method: Properties and applications." In: Signal Processing 6.3 (June 1984), 201–11 (cit. on p. 18.3).
titterington:85:ala	[32]	D. M. Titterington and C. Rossi. "Another look at a Bayesian direct deconvolution method." In: <i>Signal</i> <i>Processing</i> 9.2 (Sept. 1985), 101–6, DOI: 10.1016/0165–1684 (85) 90032–5 (cit. on p. 18.3).
hutton:97:acp	[33]	B. F. Hutton, H. M. Hudson, and F. J. Beekman. "A clinical perspective of accelerated statistical reconstruction." In: <i>Eur. J. Nuc. Med.</i> 24.7 (July 1997), 797–808. DOI: 10.1007/BF00879671 (cit. on p. 18.3).
lewitt:03:00m	[34]	R. M. Lewitt and S. Matej. "Overview of methods for image reconstruction from projections in emission computed tomography." In: <i>Proc. IEEE</i> 91.10 (Oct. 2003), 1588–611. DOI: 10.1109/JPROC.2003.817882 (cit. on p. 18.3).
qi:06:irt	[35]	J. Qi and R. M. Leahy. "Iterative reconstruction techniques in emission computed tomography." In: <i>Phys. Med. Biol.</i> 51.15 (Aug. 2006), R541–78. DOI: 10.1088/0031-9155/51/15/R01 (cit. on p. 18.3).
defrise:06:ir	[36]	M. Defrise and G. T. Gullberg. "Image reconstruction." In: <i>Phys. Med. Biol.</i> 51.13 (July 2006), R139–54. DOI: 10.1088/0031-9155/51/13/R09 (cit. on p. 18.3).
leahy:00:rdi	[37]	R. Leahy and C. Byrne. "Recent development in iterative image reconstruction for PET and SPECT." In: <i>IEEE Trans. Med. Imag.</i> 19.4 (Apr. 2000), 257–60. DOI: 10.1109/TMI.2000.848177 (cit. on p. 18.3).
leahy:00:sai	[38]	R. M. Leahy and J. Qi. "Statistical approaches in quantitative positron emission tomography." In: <i>Statistics and Computing</i> 10.2 (Apr. 2000), 147–65. DOI: 10.1023/A:1008946426658 (cit. on p. 18.3).
nuyts:01:mle	[39]	J. Nuyts, C. Michel, and P. Dupont. "Maximum-likelihood expectation-maximization reconstruction of sinograms with arbitrary noise distribution using NEC-transformations." In: <i>IEEE Trans. Med. Imag.</i> 20.5 (May 2001), 365–75. DOI: 10.1109/42.925290 (cit. on p. 18.3).
politte:91:cfa	[40]	D. G. Politte and D. L. Snyder. "Corrections for accidental coincidences and attenuation in maximum-likelihood image reconstruction for positron-emission tomography." In: <i>IEEE Trans. Med. Imag.</i> 10.1 (Mar. 1991), 82–9. DOI: 10.1109/42.75614 (cit. on p. 18.3).
aubewitherspoon:92:sci	[41]	M. E. Daube-Witherspoon et al. "Scatter correction in maximum likelihood reconstruction of PET data." In: <i>Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.</i> Vol. 2. 1992, 945–7. DOI: 10.1109/NSSMIC.1992.301098 (cit. on p. 18.3).
qi:98:f3b	[42]	J. Qi et al. "Fully 3D Bayesian image reconstruction for the ECAT EXACT HR+." In: <i>IEEE Trans. Nuc. Sci.</i> 45.3-II (June 1998), 1096–103. DOI: 10.1088/0031–9155/43/4/027 (cit. on pp. 18.3, 18.20).
salvo:19:acp	[43]	K. Salvo and M. Defrise. "A convergence proof of MLEM with fixed background." In: <i>IEEE Trans. Med.</i> <i>Imag.</i> 38.3 (Mar. 2019), 721–9. DOI: 10.1109/TMI.2018.2870968 (cit. on pp. 18.4, 18.22).
vardi:93:fid	[44]	Y. Vardi and D. Lee. "From image deblurring to optimal investments: maximum likelihood solutions for positive linear inverse problems." In: <i>J. Royal Stat. Soc. Ser. B</i> 55.3 (1993), 569–612. URL: http://www.jstor.org/stable/2345873 (cit. on p. 18.4).

tao:98:adc

an:10:dtb

- [45] N. H. Clinthorne et al. "Preconditioning methods for improved convergence rates in iterative reconstructions." In: *IEEE Trans. Med. Imag.* 12.1 (Mar. 1993), 78–83. DOI: 10.1109/42.222670 (cit. on p. 18.5).
 - [46] C. L. Byrne. "Iterative image reconstruction algorithms based on cross-entropy minimization." In: *IEEE Trans. Im. Proc.* 2.1 (Jan. 1993). Erratum and addendum: 4(2):226-7, Feb. 1995., 96–103. DOI: 10.1109/83.210869 (cit. on pp. 18.5, 18.13).
 - [47] J. Nuyts. "Unconstrained image reconstruction with resolution modelling does not have a unique solution." In: *EJNMMI Physics* 1.98 (2014). DOI: 10.1186/s40658-014-0098-4 (cit. on p. 18.5).
 - [48] S. Ahn and J. A. Fessler. "Globally convergent image reconstruction for emission tomography using relaxed ordered subsets algorithms." In: *IEEE Trans. Med. Imag.* 22.5 (May 2003), 613–26. DOI: 10.1109/TMI.2003.812251 (cit. on pp. 18.5, 18.18).
 - [49] S. Ahn. "Convergent algorithms for statistical image reconstruction in emission tomography." PhD thesis. Ann Arbor, MI: Univ. of Michigan, Ann Arbor, MI, 48109-2122, 2004. URL: https://hdl.handle.net/2027.42/124609 (cit. on p. 18.6).
 - [50] J. A. Fessler and A. O. Hero. "Penalized maximum-likelihood image reconstruction using space-alternating generalized EM algorithms." In: *IEEE Trans. Im. Proc.* 4.10 (Oct. 1995), 1417–29. DOI: 10.1109/83.465106 (cit. on pp. 18.6, 18.7, 18.11, 18.15).
 - [51] J. A. Fessler and A. O. Hero. Space-alternating generalized EM algorithms for penalized maximum-likelihood image reconstruction. Tech. rep. 286. Univ. of Michigan, Ann Arbor, MI, 48109-2122: Comm. and Sign. Proc. Lab., Dept. of EECS, Feb. 1994. URL: http://web.eecs.umich.edu/~fessler/papers/files/tr/94, 286, sage.pdf (cit. on p. 18.7).
 - [52] H. N. Multhei and B. Schorr. "On properties of the iterative maximum likelihood reconstruction method." In: *Mathematical Methods in the Applied Sciences* 11.3 (May 1989), 331–42. DOI: 10.1002/mma.1670110303 (cit. on p. 18.7).
 - [53] S. Kullback and R. A. Leibler. "On information and sufficiency." In: Ann. Math. Stat. 22.1 (Mar. 1951), 79–86. URL: http://www.jstor.org/stable/2236703 (cit. on p. 18.9).
 - [54] C. Byrne. Applied iterative methods. Wellesley, MA: A K Peters, 2008 (cit. on p. 18.9).
 - [55] J. M. Borwein and A. S. Lewis. "Convergence of best entropy estimate." In: *SIAM J. Optim.* 1.2 (May 1991), 191–205. DOI: 10.1137/0801014 (cit. on p. 18.9).
- [56] M. S. Pinsker. Information and information stability of random variables and processes. San Francisco: Holden-Day, 1964 (cit. on p. 18.9).
 - [57] A. Tsybakov. *Introduction to nonparametric estimation*. Springer, 2009. DOI: 10.1007/b13794 (cit. on p. 18.9).
 - [58] A. R. De Pierro. "On the relation between the ISRA and the EM algorithm for positron emission tomography." In: *IEEE Trans. Med. Imag.* 12.2 (June 1993), 328–33. DOI: 10.1109/42.232263 (cit. on p. 18.10).
 - [59] M. Jacobson and J. A. Fessler. "Simultaneous estimation of attenuation and activity images using optimization transfer." In: *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.* Vol. 4. 2001, 2085–9. DOI: 10.1109/NSSMIC.2001.1009235 (cit. on p. 18.12).
 - [60] P. D. Tao and L. T. H. An. "A D.C. optimization algorithm for solving the trust-region subproblem." In: *SIAM J. Optim.* 8.2 (May 1998), 476–505. DOI: 10.1137/S1052623494274313 (cit. on p. 18.13).
 - [61] T. Schule et al. "Discrete tomography by convex-concave regularization and D.C. programming." In: *Discrete Applied Mathematics* 151.1-3 (Oct. 2005), 229–43. DOI: 10.1016/j.dam.2005.02.028 (cit. on p. 18.13).
 - [62] L. T. H. An, N. T. Phuc, and P. D. Tao. "Discrete tomography based on DC programming and DCA." In: *Computing and Communication Technologies, Research, Innovation, and Vision for the Future (RIVF), 2010 IEEE RIVF International Conference on.* 2010, 1–6. DOI: 10.1109/RIVF.2010.5633367 (cit. on p. 18.13).
 - [63] I. Csiszar and G. Tusnady. "Information geometry and alternating minimization procedures." In: *Statistics and Decisions, Supplement Issue* 1 (1984), 205–37 (cit. on p. 18.13).

vermeulen:83:otp

- [64] J. A. O'Sullivan. "Alternating minimization algorithms: From Blahut-Arimoto to expectation-maximization." In: *Codes, curves, and signals.* Ed. by A Vardy. Hingham, MA: Kluwer, 1998, pp. 173–92. DOI: 10.1007/978-1-4615-5121-8_13 (cit. on p. 18.13).
- [65] J. A. Fessler, N. H. Clinthorne, and W. L. Rogers. "On complete data spaces for PET reconstruction algorithms." In: *IEEE Trans. Nuc. Sci.* 40.4 (Aug. 1993), 1055–61. DOI: 10.1109/23.256712 (cit. on p. 18.15).
- [66] E. Tanaka. "A fast reconstruction algorithm for stationary positron emission tomography based on a modified EM algorithm." In: *IEEE Trans. Med. Imag.* 6.2 (June 1987), 98–105. DOI: 10.1109/TMI.1987.4307809 (cit. on p. 18.15).
 - [67] D. G. Politte. Reconstruction algorithms for time-of-flight assisted positron-emission tomographs. Master's thesis, Department of Electrical Engineering, Sever Institute of Technology, Washington University, St. Louis. 1983 (cit. on pp. 18.15, 18.17).
 - [68] L. Kaufman. "Implementing and accelerating the EM algorithm for positron emission tomography." In: *IEEE Trans. Med. Imag.* 6.1 (Mar. 1987), 37–51. DOI: 10.1109/TMI.1987.4307796 (cit. on p. 18.15).
 - [69] M. V. Ranganath, A. P. Dhawan, and N. Mullani. "A multigrid expectation maximization reconstruction algorithm for positron emission tomography." In: *IEEE Trans. Med. Imag.* 7.4 (Dec. 1988), 273–8. DOI: 10.1109/42.14509 (cit. on p. 18.15).
 - [70] T. S. Pan and A. E. Yagle. "Numerical study of multigrid implementations of some iterative image reconstruction algorithms." In: *IEEE Trans. Med. Imag.* 10.4 (Dec. 1991), 572–88. DOI: 10.1109/42.108592 (cit. on p. 18.15).
 - [71] A. Raheja, T. F. Doniere, and A. P. Dhawan. "Multiresolution expectation maximization reconstruction algorithm for positron emission tomography using wavelet processing." In: *IEEE Trans. Nuc. Sci.* 46.3 (June 1999), 594–602. DOI: 10.1109/23.775585 (cit. on p. 18.15).
- [72] F. L. Vermeulen. "On the performance of maximum-likelihood reconstruction in positron emission tomography." In: *Proc. SPIE 397 Appl. of Digital Image Proc.* 1983, 274–9. DOI: 10.1117/12.935311 (cit. on pp. 18.15, 18.16).
 - [73] S. Vishampayan et al. "Maximum likelihood image reconstruction for SPECT." In: J. Nuc. Med. (Abs. Book) 26.5 (1985), 20:72. URL: http://www.osti.gov/scitech/biblio/6843152 (cit. on p. 18.15).
 - [74] E. Tanaka. "Utilization of non-negativity constraints in reconstruction of emission tomograms." In: *Information Processing in Medical Im.* Ed. by S L Bacharach. Boston: Martinus-Nijhoff, 1985, pp. 379–93. DOI: 10.1007/978-94-009-4261-5_27 (noUMaccess; gotfromILL) (cit. on pp. 18.15, 18.16, 18.17).
 - [75] A. N. Iusem. "Convergence analysis for a multiplicatively relaxed EM algorithm." In: *Mathematical Methods in the Applied Sciences* 14.8 (Oct. 1991), 573–93. DOI: 10.1002/mma.1670140805 (cit. on p. 18.15).
 - [76] T. S. Zaccheo and R. A. Gonsalves. "Iterative maximum-likelihood estimators for positively constrained objects." In: J. Opt. Soc. Am. A 13.2 (Feb. 1996), 236–42. DOI: 10.1364/JOSAA.13.000236 (cit. on p. 18.15).
 - [77] H. Lantéri et al. "A general method to devise maximum-likelihood signal restoration multiplicative algorithms with non-negativity constraints." In: *Signal Processing* 81.5 (May 2001), 945–74. DOI: 10.1016/S0165-1684 (00) 00275-9 (cit. on pp. 18.15, 18.21).
 - [78] R. M. Lewitt and G. Muehllehner. "Accelerated iterative reconstruction for positron emission tomography based on the EM algorithm for maximum likelihood estimation." In: *IEEE Trans. Med. Imag.* 5.1 (Mar. 1986), 16–22. DOI: 10.1109/TMI.1986.4307734 (cit. on pp. 18.15, 18.17).
 - [79] H. Hart and Z. Liang. "Bayesian image processing in two dimensions." In: *IEEE Trans. Med. Imag.* 6.3 (Sept. 1987), 201–8. DOI: 10.1109/TMI.1987.4307828 (cit. on p. 18.15).
 - [80] C. E. Metz and C. T. Chen. "On the acceleration of maximum likelihood algorithms." In: *Proc. SPIE 914 Med. Im. II: Im. Formation, Detection, Proc., and Interpretation.* 1988, 344–9. DOI: 10.1117/12.968650 (cit. on p. 18.15).
 - [81] P. Schmidlin, M. E. Bellemann, and G. Brix. "Iterative reconstruction of PET images using a high-overrelaxation single-projection algorithm." In: *Phys. Med. Biol.* 42.3 (Mar. 1997), 569–82. DOI: 10.1088/0031-9155/42/3/009 (cit. on p. 18.15).

chen:94:opt

- [82] P. Schmidlin et al. "Computation of high overrelaxation parameters in iterative image reconstruction." In: *IEEE Trans. Nuc. Sci.* 45.3 (June 1998), 1737–42. DOI: 10.1109/23.685297 (cit. on p. 18.15).
- [83] DS. Hwang and G. L. Zeng. "Convergence study of an accelerated ML-EM algorithm using bigger step size." In: *Phys. Med. Biol.* 51.2 (Jan. 2006), 237–52. DOI: 10.1088/0031-9155/51/2/004 (cit. on p. 18.15).
- [84] J. Nuyts, P. Suetens, and L. Mortelmans. "Acceleration of maximum likelihood reconstruction, using frequency amplification and attenuation compensation." In: *IEEE Trans. Med. Imag.* 12.4 (Dec. 1993), 643–52. DOI: 10.1109/42.251114 (cit. on p. 18.15).
- [85] N. Rajeevan, K. Rajgopal, and G. Krishna. "Vector extrapolated fast maximum likelihood estimation algorithms for emission tomography." In: *IEEE Trans. Med. Imag.* 11.1 (Mar. 1992), 9–20. DOI: 10.1109/42.126905 (cit. on p. 18.15).
 - [86] T. J. Holmes and Y. H. Liu. "Acceleration of maximum-likelihood image restoration for fluorescence microscopy and other non-coherent imagery." In: J. Opt. Soc. Am. A 8.6 (June 1991), 893–907. DOI: 10.1364/JOSAA.8.000893 (cit. on p. 18.15).
- [87] T. R. Miller and J. W. Wallis. "Fast maximum-likelihood reconstruction." In: J. Nuc. Med. 33.9 (Sept. 1992), 1710–11. URL: http://jnm.snmjournals.org/cgi/content/abstract/33/9/1710 (cit. on p. 18.16).
 - [88] J. A. Fessler. ASPIRE 3.0 user's guide: A sparse iterative reconstruction library. Tech. rep. 293. Available from http://web.eecs.umich.edu/~fessler. Univ. of Michigan, Ann Arbor, MI, 48109-2122: Comm. and Sign. Proc. Lab., Dept. of EECS, July 1995. URL: http://web.eecs.umich.edu/~fessler/papers/files/tr/95, 293, aspire3.pdf (cit. on p. 18.16).
- [89] T. J. Hebert and R. Leahy. "Fast methods for including attenuation in the EM algorithm." In: *IEEE Trans. Nuc. Sci.* 37.2 (Apr. 1990), 754–8. DOI: 10.1109/23.106710 (cit. on p. 18.16).
- [90] M. S. Atkins, D. Murray, and R. L. Harrop. "Use of transputers in a 3D positron emission tomograph." In: *IEEE Trans. Med. Imag.* 10.3 (Sept. 1991), 276–83. DOI: 10.1109/42.97576 (cit. on p. 18.16).
- [91] C. M. Chen, S. Y. Lee, and Z. H. Cho. "Parallelization of the EM algorithm for 3-D PET image reconstruction." In: *IEEE Trans. Med. Imag.* 10.4 (Dec. 1991), 513–22. DOI: 10.1109/42.108585 (cit. on p. 18.16).
- [92] K. Bastiaens, I. Lemahieu, and P. Desmedt. "On the use of a multi-threaded operating system for an efficient parallel implementation of the ML-EM algorithm for PET image reconstruction." In: *IFIP Transactions A* (*Computer Science and Technology*) A44 (1994). Applications in parallel and distribute computing, 31–9 (cit. on p. 18.16).
- [93] C. M. Chen and S. Y. Lee. "On parallelizing the EM algorithm for PET image reconstruction." In: *IEEE Trans. Parallel. Dist. Sys.* 5.8 (Aug. 1994), 860–73. DOI: 10.1109/71.298213 (cit. on p. 18.16).
 - [94] C. A. Johnson et al. "A system for the 3D reconstruction of retracted-septa PET data using the EM algorithm." In: *IEEE Trans. Nuc. Sci.* 42.4 (Aug. 1995), 1223–7. DOI: 10.1109/23.467875 (cit. on p. 18.16).
 - [95] C. M. Chen and S. Y. Lee. "Optimal data replication: a new approach to optimizing parallel EM algorithms for 3D PET image reconstruction." In: *IEEE Trans. Nuc. Sci.* 42.4 (Aug. 1995), 1235–45. DOI: 10.1109/23.467873 (cit. on p. 18.16).
- [96] J. L. Cruz-Rivera et al. "Parallelized formulation of the maximum likelihood-expectation maximization algorithm for fine-grain message-passing architectures." In: *IEEE Trans. Med. Imag.* 14.4 (Dec. 1995), 758–62. DOI: 10.1109/42.476118 (cit. on p. 18.16).
- [97] Z. W. Ju, E. C. Frey, and B. M. W. Tsui. "Distributed 3-D iterative reconstruction for quantitative SPECT." In: *IEEE Trans. Nuc. Sci.* 42.4 (Aug. 1995), 1301–9. DOI: 10.1109/23.467865 (cit. on p. 18.16).
 - [98] K. Rajan, L. M. Patnaik, and J. Ramakrishna. "Linear array implementation of the EM algorithm for PET image reconstruction." In: *IEEE Trans. Nuc. Sci.* 42.4-2 (Aug. 1995), 1339–44. DOI: 10.1109/23.467723 (cit. on p. 18.16).
 - [99] J. Cui et al. "Distributed MLEM: an iterative tomographic image reconstruction algorithm for distributed memory architectures." In: *IEEE Trans. Med. Imag.* 32.5 (May 2013), 957–67. DOI: 10.1109/TMI.2013.2252913 (cit. on p. 18.16).

klotz:88:aha	[100]	H. A. Klotz and D. L. Snyder. "A hardware architecture using finite-field arithmetic for computing maximum-likelihood estimates in emission tomography." In: <i>IEEE Trans. Med. Imag.</i> 7.4 (Dec. 1988), 279–90. DOI: 10.1109/42.14510 (cit. on p. 18.16).
chuang:05:aml	[101]	K-S. Chuang et al. "A maximum likelihood expectation maximization algorithm with thresholding." In: <i>Computerized Medical Imaging and Graphics</i> 29.7 (Oct. 2005), 571–8. DOI: 10.1016/j.compmedimag.2005.04.003 (cit. on p. 18.16).
zeng:92:fdi	[102]	G. L. Zeng and G. T. Gullberg. "Frequency domain implementation of the three-dimensional geometric point response correction in SPECT imaging." In: <i>IEEE Trans. Nuc. Sci.</i> 39.5-1 (Oct. 1992), 1444–53. DOI: 10.1109/23.173222 (cit. on p. 18.16).
kamphuls:98:amo	[103]	C. Kamphuis et al. "Dual matrix ordered subsets reconstruction for accelerated 3D scatter compensation in single-photon emission tomography." In: <i>Eur. J. Nuc. Med.</i> 25.1 (Jan. 1998), 8–18. DOI: 10.1007/s002590050188 (cit. on p. 18.16).
zeng:00:upb	[104]	G. L. Zeng and G. T. Gullberg. "Unmatched projector/backprojector pairs in an iterative reconstruction algorithm." In: <i>IEEE Trans. Med. Imag.</i> 19.5 (May 2000), 548–55. DOI: 10.1109/42.870265 (cit. on p. 18.16).
hudson:92:aer	[105]	H. M. Hudson, B. F. Hutton, and R. Larkin. "Accelerated EM reconstruction using ordered subsets." In: <i>J. Nuc. Med. (Abs. Book)</i> 33 (1992), p. 960 (cit. on p. 18.17).
hudson:94:air	[106]	H. M. Hudson and R. S. Larkin. "Accelerated image reconstruction using ordered subsets of projection data." In: <i>IEEE Trans. Med. Imag.</i> 13.4 (Dec. 1994), 601–9. DOI: 10.1109/42.363108 (cit. on p. 18.17).
hebert:90:tdm	[107]	T. Hebert, R. Leahy, and M. Singh. "Three-dimensional maximum-likelihood reconstruction for an electronically collimated single-photon-emission imaging system." In: <i>J. Opt. Soc. Am. A</i> 7.7 (July 1990), 1305–13. DOI: 10.1364/JOSAA.7.001305 (cit. on p. 18.17).
schmidlin:94:iii	[108]	P. Schmidlin. "Improved iterative image reconstruction using variable projection binning and abbreviated convolution." In: <i>Eur. J. Nuc. Med.</i> 21.9 (Sept. 1994), 930–6. DOI: 10.1007/BF00238116 (cit. on p. 18.17).
brown:94;icr	[109]	J. K. Brown, B. H. Hasegawa, and T. F. Lang. "Iterative concurrent reconstruction algorithms for emission computed tomography." In: <i>Phys. Med. Biol.</i> 39.7 (July 1994), 1113–32. DOI: 10.1088/0031-9155/39/7/005 (cit. on p. 18.17).
desmedt:94:otu	[110]	P. Desmedt and I. Lemahieu. <i>On the use of successive data in the ML-EM algorithm in PET</i> . 1994 (cit. on p. 18.17).
byrne:96:bim	[111]	C. L. Byrne. "Block-iterative methods for image reconstruction from projections." In: <i>IEEE Trans. Im. Proc.</i> 5.5 (May 1996), 792–3. DOI: 10.1109/83.499919 (cit. on p. 18.18).
byrne:97:cbi	[112]	C. L. Byrne. "Convergent block-iterative algorithms for image reconstruction from inconsistent data." In: <i>IEEE Trans. Im. Proc.</i> 6.9 (Sept. 1997), 1296–304. DOI: 10.1109/83.623192 (cit. on p. 18.18).
byrne:98:ate	[113]	C. L. Byrne. "Accelerating the EMML algorithm and related iterative algorithms by rescaled block-iterative methods." In: <i>IEEE Trans. Im. Proc.</i> 7.1 (Jan. 1998), 100–9. DOI: 10.1109/83.650854 (cit. on p. 18.18).
lalush:98:bit	[114]	D. S. Lalush and B. M. W. Tsui. "Block-iterative techniques for fast 4D reconstruction using a priori motion models in gated cardiac SPECT." In: <i>Phys. Med. Biol.</i> 43.4 (Apr. 1998), 875–86. DOI: 10.1088/0031-9155/43/4/015 (cit. on p. 18.18).
soares:00:nco	[115]	E. J. Soares, C. L. Byrne, and S. J. Glick. "Noise characterization of block-iterative reconstruction algorithms. I. Theory." In: <i>IEEE Trans. Med. Imag.</i> 19.4 (Apr. 2000), 261–70. DOI: 10.1109/42.848178 (cit. on p. 18.18).
byrne:05:cpi	[116]	C. Byrne. "Choosing parameters in block-iterative or ordered subset reconstruction algorithms." In: <i>IEEE Trans. Im. Proc.</i> 14.3 (Mar. 2005), 321–7. DOI: 10.1109/TIP.2004.841193 (cit. on p. 18.18).
browne:96:ara	[117]	J. A. Browne and A. R. De Pierro. "A row-action alternative to the EM algorithm for maximizing likelihoods in emission tomography." In: <i>IEEE Trans. Med. Imag.</i> 15.5 (Oct. 1996), 687–99. DOI: 10.1109/42.538946 (cit. on p. 18.18).
tanaka:03:sdr	[118]	E. Tanaka and H. Kudo. "Subset-dependent relaxation in block-iterative algorithms for image reconstruction in emission tomography." In: <i>Phys. Med. Biol.</i> 48.10 (May 2003), 1405–22. DOI: 10.1088/0031-9155/48/10/312 (cit. on p. 18.18).

depierro:01:fel

lanteri:02:pml

- [119] E. Salomão Helou-Neto and Álvaro Rodolfo De Pierro. "Convergence results for scaled gradient algorithms in positron emission tomography." In: *Inverse Prob.* 21.6 (Dec. 2005), 1905–14. DOI: 10.1088/0266-5611/21/6/007 (cit. on p. 18.18).
- [120] A. R. De Pierro and M. E. B. Yamagishi. "Fast EM-like methods for maximum 'a posteriori' estimates in emission tomography." In: *IEEE Trans. Med. Imag.* 20.4 (Apr. 2001), 280–8. DOI: 10.1109/42.921477 (cit. on p. 18.18).
 - [121] R. Neal and G. E. Hinton. "A view of the EM algorithm that justifies incremental, sparse and other variants." In: *Learning in Graphical Models*. Ed. by M. I. Jordan. Dordrencht: Kluwer, 1998, pp. 355–68. URL: http://www.cs.toronto.edu/~radford/em.abstract.html (cit. on p. 18.18).
 - [122] A. Gunawardana and W. Byrne. *Convergence of EM variants*. Tech. rep. Johns Hopkins University: ECE Dept., Feb. 1999. URL: http://www.clsp.jhu.edu/cgi-bin/zilla/showpage.prl?top (cit. on p. 18.18).
 - [123] W. Byrne and A. Gunawardana. "Comments on "Efficient training algorithms for HMMs using incremental estimation"." In: *IEEE Trans. Speech & Audio Proc.* 8.6 (Nov. 2000), 751–4. DOI: 10.1109/89.876315 (cit. on p. 18.18).
 - [124] A. Gunawardana. "The information geometry of EM variants for speech and image processing." PhD thesis. Baltimore: Johns Hopkins University, 2001. URL: https://dl.acm.org/citation.cfm?id=933549 (cit. on pp. 18.18, 18.20).
 - [125] I. T. Hsiao, A. Rangarajan, and G. Gindi. "A provably convergent OS-EM like reconstruction algorithm for emission tomography." In: *Proc. SPIE 4684 Medical Imaging: Image Proc.* 2002, 10–19. DOI: 10.1117/12.467144 (cit. on pp. 18.18, 18.20).
 - [126] M. Unser and M. Eden. "Maximum likelihood estimation of linear signal parameters for Poisson processes."
 In: *IEEE Trans. Acoust. Sp. Sig. Proc.* 36.6 (June 1988), 942–5. DOI: 10.1109/29.1613 (cit. on p. 18.20).
 - [127] C. A. Johnson and A. Sofer. "A primal-dual method for large-scale image reconstruction in emission tomography." In: *SIAM J. Optim.* 11.3 (2000), 691–715. DOI: 10.1137/S1052623497330148 (cit. on p. 18.20).
 - [128] W. W. Hager, B. A. Mair, and H. Zhang. An affine-scaling interior-point CBB method for box-constrained optimization. 2007.
 - [129] S. Sawada and H. Kudo. "Globally convergent newton-SOR method for statistical image reconstruction." In: *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.* Vol. 2. 2000, 15/263–6. DOI: 10.1109/NSSMIC.2000.950117.
 - [130] J. Nocedal and S. J. Wright. *Numerical optimization*. New York: Springer, 1999.
 - [131] J. Pauli, E-M. Pauli, and G. Anton. "ITEM-QM solutions for EM problems in image reconstruction exemplary for the Compton camera." In: *Nucl. Instr. Meth. Phys. Res. A.* 488.1-2 (Aug. 2002), 323–31. DOI: 10.1016/S0168-9002 (02) 00440-0.
 - [132] J. W. Wallis and T. R. Miller. "Rapidly converging iterative reconstruction algorithms in single-photon emission computed tomography." In: J. Nuc. Med. 34.10 (Oct. 1993), 1793–800. URL: http://jnm.snmjournals.org/cgi/content/abstract/34/10/1793.

[133] H. Lantéri, M. Roche, and C. Aime. "Penalized maximum likelihood image restoration with positivity constraints: multiplicative algorithms." In: *Inverse Prob.* 18.5 (Oct. 2002), 1397–1419. DOI: 10.1088/0266-5611/18/5/313 (cit. on p. 18.21).

- [134] R. G. Lane. "Methods for maximum-likelihood deconvolution." In: J. Opt. Soc. Am. A 13.10 (Oct. 1996), 1992–8. DOI: 10.1364/JOSAA.13.001992 (cit. on p. 18.21).
- [135] G. Wang and J. Qi. "Acceleration of the direct reconstruction of linear parametric images using nested algorithms." In: *Phys. Med. Biol.* 55.5 (Mar. 2010), 1505–17. DOI: 10.1088/0031-9155/55/5/016 (cit. on p. 18.21).
 - [136] J. C. Matthews et al. "Direct reconstruction of parametric images using any spatiotemporal 4D image based model and maximum likelihood expectation maximization." In: *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.* 2010, 2435–41. DOI: 10.1109/NSSMIC.2010.5874225 (cit. on p. 18.21).

- [137] G. I. Angelis et al. "Acceleration of image-based resolution modelling reconstruction using an expectation maximization nested algorithm." In: *Proc. Intl. Mtg. on Fully 3D Image Recon. in Rad. and Nuc. Med.* 2011, 250–3 (cit. on p. 18.21).
- [138] M. Segal and E. Weinstein. "The cascade EM algorithm." In: *Proc. IEEE* 76.10 (Oct. 1988), 1388–90. DOI: 10.1109/5.16341 (cit. on p. 18.21).
 - [139] Z. Yu et al. "Nested loop algorithm for parallel model based iterative reconstruction." In: *Proc. Intl. Mtg. on Fully 3D Image Recon. in Rad. and Nuc. Med.* 2013, 197–200 (cit. on p. 18.21).
- [140] D. Kim et al. "Accelerating ordered subsets image reconstruction for X-ray CT using spatially non-uniform optimization transfer." In: *IEEE Trans. Med. Imag.* 32.11 (Nov. 2013), 1965–78. DOI: 10.1109/TMI.2013.2266898.
- [141] E. Veklerov and J. Llacer. "Stopping rule for the MLE algorithm based on statistical hypothesis testing." In: *IEEE Trans. Med. Imag.* 6.4 (Dec. 1987), 313–9. DOI: 10.1109/TMI.1987.4307849 (cit. on p. 18.21).
 - [142] T. Hebert, R. Leahy, and M. Singh. "Fast MLE for SPECT using an intermediate polar representation and a stopping criterion." In: *IEEE Trans. Nuc. Sci.* 35.1 (Feb. 1988), 615–9. DOI: 10.1109/23.12798 (cit. on p. 18.21).
 - [143] E. Veklerov, J. Llacer, and E. J. Hoffman. "MLE reconstruction of a brain phantom using a Monte Carlo transition matrix and a statistical stopping rule." In: *IEEE Trans. Nuc. Sci.* 35.1 (Feb. 1988), 603–7. DOI: 10.1109/23.12795 (cit. on p. 18.21).
- [144] J. Llacer and E. Veklerov. "Feasible images and practical stopping rules for iterative algorithms in emission tomography." In: *IEEE Trans. Med. Imag.* 8.2 (June 1989). Corrections, 9(1), Mar 1990, 186–93. DOI: 10.1109/42.24867 (cit. on p. 18.21).
 - T. J. Hebert. "Statistical stopping criteria for iterative maximum likelihood reconstruction of emission images." In: *Phys. Med. Biol.* 35.9 (Sept. 1990), 1221–32. DOI: 10.1088/0031-9155/35/9/003 (cit. on p. 18.21).
 - [146] J. Llacer. "On the validity of hypothesis testing for feasibility of image reconstructions." In: *IEEE Trans. Med. Imag.* 9.2 (June 1990), 226–30. DOI: 10.1109/42.56335 (cit. on p. 18.21).
 - [147] E. Veklerov and J. Llacer. "The feasibility of images reconstructed with the method of sieves." In: *IEEE Trans. Nuc. Sci.* 37.2 (Apr. 1990), 835–41. DOI: 10.1109/23.106724 (cit. on p. 18.21).
- [148] K. J. Coakley. "Resampling scheme for improving maximum likelihood reconstructions of positron emission tomography images." In: *Proc. SPIE 2167 Med. Im. 1994: Image Proc.* 1993, 271–81. DOI: 10.1117/12.175061 (cit. on p. 18.21).
 - [149] V. E. Johnson. "A note on stopping rules in EM-ML reconstructions of ECT images." In: *IEEE Trans. Med. Imag.* 13.3 (Sept. 1994), 569–71. DOI: 10.1109/42.310891 (cit. on p. 18.21).
- [150] S. J. Reeves. "Generalized cross-validation as a stopping rule for the Richardson-Lucy algorithm." In: *Intl. J. Imaging Sys. and Tech.* 6.4 (1995), 387–91. DOI: 10.1002/ima.1850060412 (cit. on p. 18.21).
 - [151] F. Benvenuto and M. Piana. "Regularization of multiplicative iterative algorithms with nonnegative constraint." In: *Inverse Prob.* 30.3 (Mar. 2014), p. 035012. DOI: 10.1088/0266-5611/30/3/035012 (cit. on p. 18.21).
 - [152] D. L. Snyder et al. "Noise and edge artifacts in maximum-likelihood reconstructions for emission tomography." In: *IEEE Trans. Med. Imag.* 6.3 (Sept. 1987), 228–38. DOI: 10.1109/TMI.1987.4307831 (cit. on p. 18.21).
 - [153] M. Jacobson et al. "Enhanced 3D PET OSEM reconstruction using inter-update Metz filtering." In: *Phys. Med. Biol.* 45.8 (Aug. 2000), 2417–39. DOI: 10.1088/0031-9155/45/8/325 (cit. on p. 18.21).
 - [154] S. Mustafovic and K. Thielemans. "Object dependency of resolution in reconstruction algorithms with inter-iteration filtering applied to PET data." In: *IEEE Trans. Med. Imag.* 23.4 (Apr. 2004), 433–46. DOI: 10.1109/TMI.2004.824225 (cit. on p. 18.21).
 - [155] B. Zhang and G. L. Zeng. "An immediate after-backprojection filtering method with blob-shaped window functions for voxel-based iterative reconstruction." In: *Phys. Med. Biol.* 51.22 (Nov. 2006), 5825–42. DOI: 10.1088/0031-9155/51/22/007 (cit. on p. 18.21).

silverman:90:ase	[156]	B. W. Silverman et al. "A smoothed EM approach to indirect estimation problems, with particular reference to stereology and emission tomography." In: <i>J. Royal Stat. Soc. Ser. B</i> 52.2 (1990), 271–324. URL: http://www.jstor.org/stable/2345438 (cit. on p. 18.21).
eggermont:95:msl	[157]	P. P. B. Eggermont and V. N. LaRiccia. "Maximum smoothed likelihood density estimation for inverse problems." In: <i>Ann. Stat.</i> 23.1 (1995), 199–220. DOI: 10.1214/aos/1176324463 (cit. on p. 18.21).
eggermont:96:mpl	[158]	P. P. B. Eggermont and V. N. LaRiccia. "Maximum penalized likelihood estimation and smoothed EM algorithm for positive integral equations of the first kind." In: <i>Numer. Funct. Anal. Optimiz.</i> 17.7-8 (1996), 737–54. DOI: 10.1080/01630569608816722 (cit. on p. 18.21).
eggermont:97:nse	[159]	P. P. B. Eggermont and V. N. LaRiccia. "Nonlinearly smoothed EM density estimation with automated smoothing parameter selection for nonparametric deconvolution problems." In: <i>J. Am. Stat. Assoc.</i> 92.440 (Dec. 1997), 1451–8. URL: http://www.jstor.org/stable/2965415 (cit. on p. 18.21).
latham:92:aha	[160]	G. A. Latham and R. S. Anderssen. "A hyperplane approach to the EMS algorithm." In: <i>Appl. Math. Lett.</i> 5.5 (Sept. 1992), 71–4. DOI: 10.1016/0893-9659(92)90068-K (cit. on p. 18.21).
latham:94:aqf	[161]	G. A. Latham and R. S. Anderssen. "Assessing quantification for the EMS algorithm." In: <i>Linear Algebra and its Applications</i> 210 (Oct. 1994), 89–122. DOI: 10.1016/0024-3795 (94) 90467-7 (cit. on p. 18.21).
latham:94:ots	[162]	G. A. Latham and R. S. Anderssen. "On the stabilization inherent in the EMS algorithm." In: <i>Inverse Prob.</i> 10.1 (Feb. 1994), 161–84, DOI: 10.1088/0266-5611/10/1/01.3 (cit. on p. 18.21).
latham:95:eoe	[163]	G. A. Latham. "Existence of EMS solutions and a priori estimates." In: <i>SIAM J. Matrix. Anal. Appl.</i> 16.3 (1995), 943–53, DOI: 10.1137/S0895479893253847 (cit. on p. 18.21).
polack:ll:dpi	[164]	J. K. Polack et al. "Dual-particle imager for standoff detection of special nuclear material." In: <i>Proc. IEEE</i> Nuc Sci Symp Med Im Conf 2011 1494–500 DOI: 10.1109/NSSMIC.2011.6154355
hsiao:10:aao	[165]	I-T. Hsiao and H-M. Huang. "An accelerated ordered subsets reconstruction algorithm using an accelerating power factor for emission tomography." In: <i>Phys. Med. Biol.</i> 55.3 (Feb. 2010), 599–614. DOI: 10.1088/0031-9155/55/3/003.
clinthorne:91:jml	[166]	N. H. Clinthorne et al. "Joint maximum likelihood estimation of emission and attenuation densities in PET." In: <i>Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.</i> Vol. 3. 1991, 1927–32. DOI: 10.1109/NSSMIC.1991.259253.
berker:16:aci	[167]	Y. Berker and Y. Li. "Attenuation correction in emission tomography using the emission data - A review." In: <i>Med. Phys.</i> 43.2 (Feb. 2016), 807–32. DOI: 10.1118/1.4938264.
sitek:08:rop	[168]	A. Sitek. "Representation of photon limited data in emission tomography using origin ensembles." In: <i>Phys. Med. Biol.</i> 53.12 (June 2008), 3201–16. DOI: 10.1088/0031-9155/53/12/009.
aharoni:01:ota	[169]	R. Aharoni and D. Lee. "On the achievability of the Cramer-Rao bound for Poisson distribution." In: <i>IEEE Trans. Info. Theory</i> 47.5 (July 2001), 2096–100, DOI: 10.1109/18.930951.
mair:02:amc	[170]	B. A. Mair, M. Rao, and J. M. M. Anderson. "A minimum chi-squared method for indirect parameter estimation from Poisson data." In: <i>Statistica Neerlandica</i> 56.2 (May 2002), 165–78. DOI: 10.1111/1467-9574.00191
bazaraa:93	[171]	M. S. Bazaraa, H. D. Sherali, and C. M. Shetty. <i>Nonlinear programming: Theory and algorithms</i> . New York: Wiley 1993
yao:00:pco	[172]	R. Yao et al. "Performance characteristics of the 3-D OSEM algorithm in the reconstruction of small animal PET images" In: <i>IEEE Trans. Med. Imag.</i> 19.8 (Aug. 2000) 798–804. DOI: 10.1109/42.876305
meyer:76:scf	[173]	 R. R. Meyer. "Sufficient conditions for the convergence of monotonic mathematical programming algorithms." In: <i>J. Comput. System. Sci.</i> 12.1 (1976), 108–21. DOI: '10.1016/S0022-0000 (76) 80021-9'.
kondor:83:moc	[174]	A. Kondor. "Method of convergent weights—An iterative for solving Fredholm's integral equations of the first kind." In: <i>Nucl. Instr. Meth. Phys. Res.</i> 216.1-2 (Oct. 1983), 177–81. DOI: 10.1016/0167-5087 (83) 90348-4.
miller:87:tro	[175]	M. I. Miller and D. L. Snyder. "The role of likelihood and entropy in incomplete-data problems: Applications to estimating point-process intensities and Toeplitz constrained covariances." In: <i>Proc. IEEE</i> 75.7 (July 1987), 892–907. DOI: 10.1109/PROC.1987.13825.

(cit. on p. 18.22).

mair:96:pet	[176]	B. A. Mair, M. Rao, and J. M. M. Anderson. "Positron emission tomography, Borel measures and weak convergence." In: <i>Inverse Prob.</i> 12.6 (Dec. 1996), 965–76. DOI: 10.1088/0266-5611/12/6/011.
resmerita:0/:tem	[177]	E. Resmerita, H. W. Engl, and A. N. Iusem. "The expectation-maximization algorithm for ill-posed integral equations: a convergence analysis." In: <i>Inverse Prob.</i> 23.6 (Dec. 2007), 2575–2588. DOI: 10.1088/0266-5611/23/6/019.
fessler:94:pwl	[178]	J. A. Fessler. "Penalized weighted least-squares image reconstruction for positron emission tomography." In: <i>IEEE Trans. Med. Imag.</i> 13.2 (June 1994), 290–300. DOI: 10.1109/42.293921.
lange:90:coe	[179]	K. Lange. "Convergence of EM image reconstruction algorithms with Gibbs smoothing." In: <i>IEEE Trans. Med. Imag.</i> 9.4 (Dec. 1990). Corrections, T-MI, 10:2(288), June 1991., 439–46. DOI: 10.1109/42.61759.
dempster.//.mrr	[180]	A. P. Dempster, N. M. Laird, and D. B. Rubin. "Maximum likelihood from incomplete data via the EM algorithm." In: <i>J. Royal Stat. Soc. Ser. B</i> 39.1 (1977), 1–38. URL: http://www.jstor.org/stable/2984875.
boyles:83:otc	[181]	R. Boyles. "On the convergence of the EM algorithm." In: J. Royal Stat. Soc. Ser. B 45.1 (1983), 47–50. URL: http://www.jstor.org/stable/2345622.
wu:83:otc	[182]	C. F. J. Wu. "On the convergence properties of the EM algorithm." In: <i>Ann. Stat.</i> 11.1 (Mar. 1983), 95–103. DOI: 10.1214/aos/1176346060.
hero:95:cin	[183]	A. O. Hero and J. A. Fessler. "Convergence in norm for alternating expectation-maximization (EM) type algorithms." In: <i>Statistica Sinica</i> 5.1 (Jan. 1995), 41–54. URL: http://www3.stat.sinica.edu.tw/statistica/oldpdf/A5n13.pdf.
lange:84:era	[184]	K. Lange and R. Carson. "EM reconstruction algorithms for emission and transmission tomography." In: J. Comp. Assisted Tomo. 8.2 (Apr. 1984), 306–16.
natterer:01	[185]	F. Natterer and F. Wübbeling. <i>Mathematical methods in image reconstruction</i> . Philadelphia: Soc. Indust. Appl. Math., 2001. DOI: 10.1137/1.9780898719284.
lusem:92:asc	[186]	A. N. Iusem. "A short convergence proof of the EM algorithm for a specific Poisson model." In: <i>Braz. Journal of Prob. and Stat.</i> 6 (1992), 57–7.
vard1:85:asm	[187]	Y. Vardi, L. A. Shepp, and L. Kaufman. "A statistical model for positron emission tomography." In: <i>J. Am. Stat. Assoc.</i> 80.389 (Mar. 1985), 8–37. URL: http://www.jstor.org/stable/2288030.
roseniicht:85	[188]	M. Rosenlicht. Introduction to analysis. New York: Dover, 1985.
vekierov.94.ncc	[189]	E. Veklerov. "How to compute the difference between tomographic images." In: <i>IEEE Trans. Med. Imag.</i> 13.3 (Sept. 1994), 566–9. DOI: 10.1109/42.310890 (cit. on p. 18.22).
zhang:01:aat	[190]	S. Zhang and Y. M. Wang. "An approach to positron emission tomography based on penalized cross-entropy minimization." In: <i>Signal Processing</i> 81.5 (May 2001), 1069–74. DOI: 10.1016/S0165-1684 (00) 00245-0.
matsuyama:03:tae	[191]	Y. Matsuyama. "The alpha-EM algorithm: surrogate likelihood maximization using alpha-logarithmic information measures." In: <i>IEEE Trans. Info. Theory</i> 49.3 (Mar. 2003), 692–706. DOI: 10.1109/TIT.2002.808105 (cit. on p. 18.22).
khurd:04:agc	[192]	P. Khurd et al. "A globally convergent regularized ordered-subset EM algorithm for list-mode reconstruction." In: <i>IEEE Trans. Nuc. Sci.</i> 51.3 (June 2004), 719–25. DOI: 10.1109/TNS.2004.829780