The EMML and SMART Algorithms

Charles Byrne (Charles_Byrne@uml.edu)
Department of Mathematical Sciences
University of Massachusetts at Lowell
Lowell, MA 01854, USA

November 6, 2006

1 Background

The expectation maximization maximum likelihood method (EMML) has been the subject of much attention in the medical-imaging literature over the past decade. Statisticians like it because it is based on the well-studied principle of likelihood maximization for parameter estimation. Physicists like it because, unlike its competition, filtered backprojection, it permits the inclusion of sophisticated models of the physical situation. Mathematicians like it because it can be derived from iterative optimization theory. Physicians like it because the images are often better than those produced by other means. No method is perfect, however, and the EMML suffers from sensitivity to noise and slow rate of convergence. Research is ongoing to find faster and less sensitive versions of this algorithm.

Another class of iterative algorithms was introduced into medical imaging by Gordon et al. in [9]. These include the *algebraic reconstruction technique* (ART) and its multiplicative version, MART. These methods were derived by viewing image reconstruction as solving systems of linear equations, possibly subject to constraints, such as positivity. The *simultaneous* MART (SMART) [7, 10] is a variant of MART that uses all the data at each step of the iteration.

Although the EMML and SMART algorithms have quite different histories and are not typically considered together they are closely related [1, 2]. In this paper we examine these two algorithms in tandem, following [3]. Forging a link between the EMML and SMART led to a better understanding of both of these algorithms and to new results. The proof of convergence of the SMART in the inconsistent case [1] was based on the analogous proof for the EMML [6, 11], while discovery of the faster version of the EMML, the *rescaled block-iterative* EMML (RBI-EMML) [4] came from studying the analogous block-iterative version of SMART [5]. The proofs we give here are elementary and rely mainly on easily established properties of the cross-entropy or Kullback-Leibler distance. The alternating minimization method used in these proofs forms the basis for De Pierro's surrogate-function method for regularization.

2 The Kullback-Leibler Distance

For $a > 0$ and $b > 0$, we define

$$KL(a, b) = a \log(\frac{a}{b}) + b - a, \tag{2.1}$$

with $KL(a, 0) = +\infty$, $KL(0, b) = b$, and $KL(0, 0) = 0$. For vectors $\mathbf{x}$ and $\mathbf{z}$ with non-negative entries, we define the Kullback-Leibler distance $KL(\mathbf{x}, \mathbf{z})$ by

$$KL(\mathbf{x}, \mathbf{z}) = \sum_{n=1}^{N} KL(x_n, z_n). \tag{2.2}$$

The function $KL(\mathbf{x}, \mathbf{z})$ is not a distance in the true sense, but is non-negative and equals zero if and only if $\mathbf{x} = \mathbf{z}$. Clearly, the KL distance has the property $KL(c\mathbf{x}, c\mathbf{z}) = cKL(\mathbf{x}, \mathbf{z})$ for all positive scalars c.

Exercise 2.1 *Let* $z_+ = \sum_{j=1}^{J} z_j > 0$. *Then*

$$KL(\mathbf{x}, \mathbf{z}) = KL(x_+, z_+) + KL(\mathbf{x}, (x_+/z_+)\mathbf{z}). \tag{2.3}$$

As we shall see, the KL distance mimics the ordinary square of the Euclidean distance in several ways that make it particularly useful in designing optimization algorithms.

3 The Alternating Minimization Paradigm

Let P be an I by J matrix with entries $P_{ij} \geq 0$, such that, for each $j = 1, ..., J$, we have $s_j = \sum_{i=1}^{I} P_{ij} > 0$. Let $\mathbf{y} = (y_1, ..., y_I)^T$ with $y_i > 0$ for each i. We shall assume throughout this paper that $s_j = 1$ for each j. If this is not the case initially, we replace x_j with $x_j s_j$ and P_{ij} with P_{ij}/s_j; the quantities $(P\mathbf{x})_i$ are unchanged.

For each nonnegative vector $\mathbf{x}$ for which $(P\mathbf{x})_i = \sum_{j=1}^{J} P_{ij} x_j > 0$, let $r(\mathbf{x}) = \{r(\mathbf{x})_{ij}\}$ and $q(\mathbf{x}) = \{q(\mathbf{x})_{ij}\}$ be the I by J arrays with entries

$$r(\mathbf{x})_{ij} = x_j P_{ij} \frac{y_i}{(P\mathbf{x})_i}$$

and

$$q(\mathbf{x})_{ij} = x_j P_{ij}.$$

The KL distances

$$KL(r(\mathbf{x}), q(\mathbf{z})) = \sum_{i=1}^{I} \sum_{j=i}^{J} KL(r(\mathbf{x})_{ij}, q(\mathbf{z})_{ij})$$

and

$$KL(q(\mathbf{x}), r(\mathbf{z})) = \sum_{i=1}^{I} \sum_{j=1}^{J} KL(q(\mathbf{x})_{ij}, r(\mathbf{z})_{ij})$$

will play important roles in the discussion that follows. Note that if there is nonnegative $\mathbf{x}$ with $r(\mathbf{x}) = q(\mathbf{x})$ then $\mathbf{y} = P\mathbf{x}$.

4 Some Pythagorean Identities Involving the KL Distance

The iterative algorithms we discuss in this paper are derived using the principle of *alternating minimization*, according to which the distances $KL(r(\mathbf{x}), q(\mathbf{z}))$ and $KL(q(\mathbf{x}), r(\mathbf{z}))$ are minimized, first with respect to the variable $\mathbf{x}$ and then with respect to the variable $\mathbf{z}$. Although the KL distance is not Euclidean, and, in particular, not even symmetric, there are analogues of Pythagoras' theorem that play important roles in the convergence proofs.

Exercise 4.1 *Establish the following Pythagorean identities:*

$$KL(r(\mathbf{x}), q(\mathbf{z})) = KL(r(\mathbf{z}), q(\mathbf{z})) + KL(r(\mathbf{x}), r(\mathbf{z})); \tag{4.1}$$

$$KL(r(\mathbf{x}), q(\mathbf{z})) = KL(r(\mathbf{x}), q(\mathbf{x}')) + KL(\mathbf{x}', \mathbf{z}), \tag{4.2}$$

for

$$x'_j = x_j \sum_{i=1}^{I} P_{ij} \frac{y_i}{(P\mathbf{x})_i}; \tag{4.3}$$

$$KL(q(\mathbf{x}), r(\mathbf{z})) = KL(q(\mathbf{x}), r(\mathbf{x})) + KL(\mathbf{x}, \mathbf{z}) - KL(P\mathbf{x}, P\mathbf{z}); \tag{4.4}$$

$$KL(q(\mathbf{x}), r(\mathbf{z})) = KL(q(\mathbf{z}''), r(\mathbf{z})) + KL(\mathbf{x}, \mathbf{z}''), \tag{4.5}$$

for

$$z''_j = z_j \exp(\sum_{i=1}^{I} P_{ij} \log \frac{y_i}{(P\mathbf{z})_i}). \tag{4.6}$$

Note that it follows from Equation (2.3) that $KL(\mathbf{x}, \mathbf{z}) - KL(P\mathbf{x}, P\mathbf{z}) \geq 0$.

5 The Two Algorithms

The algorithms we shall consider are the expectation maximization maximum likelihood method (EMML) and the simultaneous multiplicative algebraic reconstruction technique (SMART). When $\mathbf{y} = P\mathbf{x}$ has nonnegative solutions, both algorithms produce such a solution. In general, the EMML gives a nonnegative minimizer of $KL(\mathbf{y}, P\mathbf{x})$, while the SMART minimizes $KL(P\mathbf{x}, \mathbf{y})$ over nonnegative $\mathbf{x}$.

For both algorithms we begin with an arbitrary positive vector $\mathbf{x}^0$. The iterative step for the EMML method is

$$x_j^{k+1} = (\mathbf{x}^k)'_j = x_j^k \sum_{i=1}^{I} P_{ij} \frac{y_i}{(P\mathbf{x}^k)_i}. \tag{5.1}$$

The iterative step for the SMART is

$$x_j^{m+1} = (\mathbf{x}^m)''_j = x_j^m \exp\Big(\sum_{i=1}^{I} P_{ij} \log \frac{y_i}{(P\mathbf{x}^m)_i} \Big). \tag{5.2}$$

Note that, to avoid confusion, we use k for the iteration number of the EMML and m for the SMART.

Exercise 5.1 *Show that, for $\{\mathbf{x}^k\}$ given by Equation (5.1), $\{KL(\mathbf{y}, P\mathbf{x}^k)\}$ is decreasing and $\{KL(\mathbf{x}^{k+1}, \mathbf{x}^k)\} \to 0$. Show that, for $\{\mathbf{x}^m\}$ given by Equation (5.2), $\{KL(P\mathbf{x}^m, \mathbf{y})\}$ is decreasing and $\{KL(\mathbf{x}^m, \mathbf{x}^{m+1})\} \to 0$.*

Hint: Use $KL(r(\mathbf{x}), q(\mathbf{x})) = KL(\mathbf{y}, P\mathbf{x})$, $KL(q(\mathbf{x}), r(\mathbf{x})) = KL(P\mathbf{x}, \mathbf{y})$, and the Pythagorean identities.

Exercise 5.2 *Show that the EMML sequence $\{\mathbf{x}^k\}$ is bounded by showing*

$$\sum_{j=1}^{J} x_j^k = \sum_{i=1}^{I} y_i.$$

Show that the SMART sequence $\{\mathbf{x}^m\}$ is bounded by showing that

$$\sum_{j=1}^{J} x_j^m \le \sum_{i=1}^{I} y_i.$$

Exercise 5.3 *Show that* $(\mathbf{x}^*)' = \mathbf{x}^*$ *for any cluster point* $\mathbf{x}^*$ *of the EMML sequence* $\{\mathbf{x}^k\}$ *and that* $(\mathbf{x}^*)'' = \mathbf{x}^*$ *for any cluster point* $\mathbf{x}^*$ *of the SMART sequence* $\{\mathbf{x}^m\}$.

Hint: Use the facts that $\{KL(\mathbf{x}^{k+1}, \mathbf{x}^k)\} \to 0$ and $\{KL(\mathbf{x}^m, \mathbf{x}^{m+1})\} \to 0$.

Exercise 5.4 *Let* $\hat{\mathbf{x}}$ *and* $\tilde{\mathbf{x}}$ *minimize* $KL(\mathbf{y}, P\mathbf{x})$ *and* $KL(P\mathbf{x}, \mathbf{y})$, *respectively, over all* $\mathbf{x} \geq \mathbf{0}$. *Then,* $(\hat{\mathbf{x}})' = \hat{\mathbf{x}}$ *and* $(\tilde{\mathbf{x}})'' = \tilde{\mathbf{x}}$.

Hint: Apply Pythagorean identities to $KL(r(\hat{\mathbf{x}}), q(\hat{\mathbf{x}}))$ and $KL(q(\tilde{\mathbf{x}}), r(\tilde{\mathbf{x}}))$.

Note that, because of convexity properties of the KL distance, even if the minimizers $\hat{\mathbf{x}}$ and $\tilde{\mathbf{x}}$ are not unique, the vectors $P\hat{\mathbf{x}}$ and $P\tilde{\mathbf{x}}$ are unique.

Exercise 5.5 *For the EMML sequence* $\{\mathbf{x}^k\}$ *with cluster point* $\mathbf{x}^*$ *and* $\hat{\mathbf{x}}$ *as defined previously, we have the double inequality*

$$KL(\hat{\mathbf{x}}, \mathbf{x}^k) \geq KL(r(\hat{\mathbf{x}}), r(\mathbf{x}^k)) \geq KL(\hat{\mathbf{x}}, \mathbf{x}^{k+1}), \tag{5.3}$$

from which we conclude that the sequence $\{KL(\hat{\mathbf{x}}, \mathbf{x}^k)\}$ *is decreasing and* $KL(\hat{\mathbf{x}}, \mathbf{x}^*) < +\infty$.

Hint: For the first inequality calculate $KL(r(\hat{\mathbf{x}}), q(\mathbf{x}^k))$ in two ways. For the second one, use $(\mathbf{x})'_j = \sum_{i=1}^{I} r(\mathbf{x})_{ij}$ and Exercise 2.1.

Exercise 5.6 *Show that, for the SMART sequence* $\{\mathbf{x}^m\}$ *with cluster point* $\mathbf{x}^*$ *and* $\tilde{\mathbf{x}}$ *as defined previously, we have*

$$KL(\tilde{\mathbf{x}}, \mathbf{x}^m) - KL(\tilde{\mathbf{x}}, \mathbf{x}^{m+1}) = KL(P\mathbf{x}^{m+1}, \mathbf{y}) - KL(P\tilde{\mathbf{x}}, \mathbf{y}) +$$

$$KL(P\tilde{\mathbf{x}}, P\mathbf{x}^m) + KL(\mathbf{x}^{m+1}, \mathbf{x}^m) - KL(P\mathbf{x}^{m+1}, P\mathbf{x}^m), \tag{5.4}$$

and so $KL(P\tilde{\mathbf{x}}, P\mathbf{x}^*) = 0$, *the sequence* $\{KL(\tilde{\mathbf{x}}, \mathbf{x}^m)\}$ *is decreasing and* $KL(\tilde{\mathbf{x}}, \mathbf{x}^*) < +\infty$.

Hint: Expand $KL(q(\tilde{\mathbf{x}}), r(\mathbf{x}^m))$ using the Pythagorean identities.

Exercise 5.7 *For* $\mathbf{x}^*$ *a cluster point of the EMML sequence* $\{\mathbf{x}^k\}$ *we have* $KL(\mathbf{y}, P\mathbf{x}^*) = KL(\mathbf{y}, P\hat{\mathbf{x}})$. *Therefore,* $\mathbf{x}^*$ *is a nonnegative minimizer of* $KL(\mathbf{y}, P\mathbf{x})$. *Consequently, the sequence* $\{KL(\mathbf{x}^*, \mathbf{x}^k)\}$ *converges to zero, and so* $\{\mathbf{x}^k\} \to \mathbf{x}^*$.

Hint: Use the double inequality of Equation (5.3) and $KL(r(\hat{\mathbf{x}}), q(\mathbf{x}^*))$.

Exercise 5.8 *For* $\mathbf{x}^*$ *a cluster point of the SMART sequence* $\{\mathbf{x}^m\}$ *we have* $KL(P\mathbf{x}^*, \mathbf{y}) = KL(P\tilde{\mathbf{x}}, \mathbf{y})$. *Therefore,* $\mathbf{x}^*$ *is a nonnegative minimizer of* $KL(P\mathbf{x}, \mathbf{y})$. *Consequently, the sequence* $\{KL(\mathbf{x}^*, \mathbf{x}^m)\}$ *converges to zero, and so* $\{\mathbf{x}^m\} \to \mathbf{x}^*$. *Moreover,*

$$KL(\tilde{\mathbf{x}}, \mathbf{x}^0) \geq KL(\mathbf{x}^*, \mathbf{x}^0)$$

for all $\tilde{\mathbf{x}}$ *as before.*

Hints: Use Exercise 5.6. For the final assertion use the fact that the difference $KL(\tilde{\mathbf{x}}, \mathbf{x}^m) - KL(\tilde{\mathbf{x}}, \mathbf{x}^{m+1})$ is independent of the choice of $\tilde{\mathbf{x}}$, since it depends only on $P\mathbf{x}^* = P\tilde{\mathbf{x}}$. Now sum over the index m.

Both the EMML and the SMART algorithms are slow to converge. For that reason attention has shifted, in recent years, to *block-iterative* versions of these algorithms.

References

[1] Byrne, C. (1993) "Iterative image reconstruction algorithms based on cross-entropy minimization." *IEEE Transactions on Image Processing* **IP-2**, pp. 96–103.

[2] Byrne, C. (1995) "Erratum and addendum to 'Iterative image reconstruction algorithms based on cross-entropy minimization'." *IEEE Transactions on Image Processing* **IP-4**, pp. 225–226.

[3] Byrne, C. (1996) "Iterative reconstruction algorithms based on cross-entropy minimization." in *Image Models (and their Speech Model Cousins)*, S.E. Levinson and L. Shepp, editors, IMA Volumes in Mathematics and its Applications, Volume 80, pp. 1–11. New York: Springer-Verlag.

[4] Byrne, C. (1996) "Block-iterative methods for image reconstruction from projections." *IEEE Transactions on Image Processing* **IP-5**, pp. 792–794.

[5] Censor, Y. and Segman, J. (1987) “On block-iterative maximization.” *J. of Information and Optimization Sciences* **8**, pp. 275–291.

[6] Csiszár, I. and Tusnády, G. (1984) “Information geometry and alternating minimization procedures.” *Statistics and Decisions* **Supp. 1**, pp. 205–237.

[7] Darroch, J. and Ratcliff, D. (1972) “Generalized iterative scaling for log-linear models.” *Annals of Mathematical Statistics* **43**, pp. 1470–1480.

[8] De Pierro, A. (1995) “A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography.” *IEEE Transactions on Medical Imaging* **14**, pp. 132–137.

[9] Gordon, R., Bender, R., and Herman, G.T. (1970) “Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and x-ray photography.” *J. Theoret. Biol.* **29**, pp. 471–481.

[10] Schmidlin, P. (1972) “Iterative separation of sections in tomographic scintigrams.” *Nucl. Med.* **15(1)**.

[11] Vardi, Y., Shepp, L.A. and Kaufman, L. (1985) “A statistical model for positron emission tomography.” *Journal of the American Statistical Association* **80**, pp. 8–20.