PET Regularization by Envelope Guided Conjugate Gradients

Linda Kaufman* and Arnold Neumaier

Abstract—**We propose a new way to iteratively solve large scale ill-posed problems and in particular the image reconstruction problem in positron emission tomography by exploiting the relation between Tikhonov regularization and multiobjective optimization to obtain iteratively approximations to the Tikhonov L-curve and its corner. Monitoring the change of the approximate L-curves allows us to adjust the regularization parameter adaptively during a preconditioned conjugate gradient iteration, so that the desired solution can be reconstructed with a small number of iterations.**

Manuscript received October 27, 1994; revised March 11, 1996. The Associate Editor responsible for coordinating the review of this paper and recommending its publication was M. Vannier. *Asterisk indicates corresponding author.*

*L. Kaufman is with Lucent Technologies, Bell Laboratories, rm.2C-461, 700 Mountain Avenue, Murray Hill, NJ 07974 USA (e-mail: lck@research.bell-labs.com.

A. Neumaier is with the Institute für Mathematik, University of Vienna, Vienna A-1090, Austria.

Publisher Item Identifier S 0278-0062(96)04415-1.

I. Introduction

In positron emission tomography (PET), the patient is given a tagged substance that emits positrons. Each positron annihilates with an electron and emits two photons in opposite directions. The reconstruction problem in PET is to determine the distribution of the annihilations from the data gathered by a ring of detectors surrounding the patient.

Assume one has imposed a grid of boxes on the affected organ and tries to compute the unknown number x_B of annihilations in the box B. Let b_d represent the number of photon pairs detected in tube d and let the matrix entry $A_{d,B}$ represent the probability that an emission in box B is detected in tube d. Then the desired density x satisfies the approximate relation

$$Ax \approx b. \tag{1}$$

The iterative probability matrix approaches to determine x (e.g., [8] and [13]) are plagued by a snowy image problem. This is particularly noticed when there are relatively few photon pairs.

Since the coefficient matrix is very ill-conditioned, the solution of the linear system $Ax = b$ must be regularized. This is often done by truncating some iterative method at a point where the signal is clearly visible but "snowy" contributions from the near null space of the matrix are not yet pronounced. Automating the decision of when to stop the iteration is not always easy as is indicated in [3].

The Tikhonov regularization technique discussed in this paper entails adding a smoothing *penalty term* to the objective function. We will use

$$r(x) = \|Ax - b\|_2^2 \tag{2}$$

as the objective in this paper, but our technique is also applicable to a negative log likelihood measure. In the penalty approach one adds a multiple of a penalty term $q(x)$, measuring the roughness of the image, and thus minimizes

$$f(x) = r(x) + \lambda q(x) \tag{3}$$

for a suitable *regularization parameter* λ.

One of the most satisfying schemes (according to a comparison in [3]) for the automatic selection of a good value of λ was suggested two decades ago by Lawson and Hanson [12] and popularized recently by Hansen [4]. In this approach one optimizes (3) for a number of values of λ and plots the resulting $r(x)$ versus $q(x)$. Usually, the result looks like an L-shaped curve, the Tikhonov L-curve. A good value of λ would be at a *corner* of the L-curve, where the curvature of the L-curve is greatest.

The envelope technique, developed in Section II, is designed to find a point on the $r(x)$, $q(x)$ plot that is near the corner of the L-curve. It does *not* involve minimizing (1) for many values of λ. It may be used for various "inverse problems" including the CAT problem with [11].

In Section III we apply the envelope technique to two-dimensional (2-D) idealized simple PET problems that were generated from a known idealized phantom to facilitate the interpretation of the results. Since the envelope technique is quite independent of the particular application, it is likely to work, too, in sophisticated realistic PET cases, involving three-dimensional (3-D) images and more realistic imaging system models.

To show the generality of our approach we have used a number of candidates for $q(x)$ including the quadratic

$$q(x) = \sum_j \left(\frac{1}{8} \sum_{i \sim j} x_i - x_j \right)^2 \tag{4}$$

where $i \sim j$ denotes the statement that i is adjacent to j (i.e., is one of the eight neighbors of j on the grid), several convex, nonquadratic penalty terms, $q(x) = \sum_j \sum_{i \sim j} H(x_i - x_j, \delta)$ with Huber's [7] function

$$H(d, \delta) = \begin{cases} d^2 & \text{if } |d| < \delta \\ 2\delta|d| - \delta^2 & \text{otherwise} \end{cases} \tag{5}$$

the multiquadric function of Vogel and Oman [15]

$$q(x) = \sum_j \sum_{i \sim j} \sqrt{(x_i - x_j)^2 + \delta} \tag{6}$$

and $q(x) = \sum_j \sum_{i \sim j} \log \cosh (x_i - x_j/\delta)$ of Green [2]; and some redescending functions: Hebert and Leahy's [6]

$$q(x) = \sum_j \sum_{i \sim j} \log \left[1 + \frac{(x_i - x_j)^2}{\delta} \right] \tag{7}$$

and $q(x) = \sum_j \sum_{i \sim j} (x_i - x_j)^2 / [(x_i - x_j)^2 + \delta]$, of Geman and McClure [1], which all have an additional parameter δ. The merit of some of these functions and the need for algorithms to determine good values for δ and λ are described by Lalush and Tsui [10].

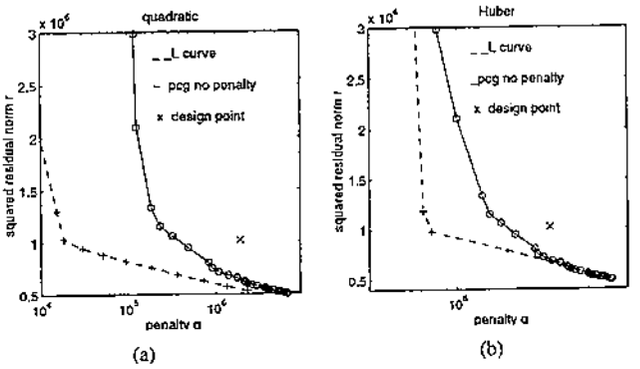


Fig. 1. Tikhonov L-curve for the (a) quadratic and (b) Huber penalty function with $\delta = 1$; 1 million photon pairs.

II. The Envelope Guided Conjugate Gradient Algorithm

The L-curve approach of [5] for determining a good value of λ to be used in (3) is simple to use for small problems and is more robust than the standard generalized cross-validation approach of [16] when the noise is correlated with the signal of the data. Because it entails minimizing (3) for many values of λ, it is too time consuming for the PET problem. However, it can be modified to produce a useful algorithm.

Fig. 1 tries to explain the basic idea behind the L-curve and to indicate its applicability to the PET problem. We generated the matrix for the PET problem defined by Shepp and Vardi [13] using a 128 × 128 grid and 128 detectors. A right-hand side $\hat{b}$ was constructed by simulating 1 million random annihilations from a phantom distribution $\hat{x}$ which looks like the top left-hand corner of Fig. 5. The dashed lines in Fig. 1 is the result of minimizing (3) with various values of λ with $q(x)$ defined in (4) and (5). Notice that the curve is L-shaped. According to [12], the best value of λ would be that at which these approximate L-curves are the most bent.

The cross (×) in the two graphs in the figure give the coordinates of the original data. Note that $\|A\hat{x} - \hat{b}\|_2$ was roughly the same order of magnitude as $\|\hat{b}\|_1$, so that decreasing $r(x)$ much below that means one is doing a good job of fitting the noise in the data as well as its signal. The solid lines show the course of the preconditioned conjugate gradient method (PCG) algorithm in [8] applied to this problem with $\lambda = 0$, i.e., ignoring the smoothing penalty. The solid lines are always above the L-curve. The corner of the envelope of all the points in the (q, r) plane generated for $\lambda = 0$ would have been a

good termination point for that value of λ, but this would still be far from the most bent point on the L-curve. This observation suggests one might try to adjust λ during the course of the algorithm in order to push the corner of the envelope of the points seen so far closer to the unknown point of maximum curvature on the Tikhonov L-curve.

In the algorithm given below we move in the (q, r) plane trying to approximate a corner of the Tikhonov L-curve. Instead of finding points on the curve itself, we use the envelopes of the set of points generated so far to guide us to the desired destination. We try to get better approximations to a point of large curvature of the Tikhonov L-curve by monitoring the corners of the current envelope. The value of λ in the objective function (3) is updated in cycles of three iterations referred to as a *major iteration.*

The theoretical basis of the approach, given in [9], is summarized by the following theorems.

Theorem 1: Suppose $q(x) = \|Ax - b\|_2^2$ and the rows of the matrix (A^T, J^T) are linearly independent. Then for each $\lambda > 0$, the optimization problem (3) has a unique solution x_λ. Let $q_\lambda := q(x_\lambda)$ and $r_\lambda := r(x_\lambda)$. The curve Γ consisting of all pairs (q_λ, r_λ) is strictly monotone decreasing and convex, its envelope contains the envelope of any set of pairs $[q(x), r(x)]$, and Γ is the L-curve, or boundary of the envelope of the set of all pairs $[q(x), r(x)]$.

Theorem 2: Let $(q_0, r_0), (q_1, r_1), \cdots, (q_N, r_N)$ be the vertices of an L-curve of a finite set S, ordered such that $r_0 > r_1 > \cdots > r_N$. Then $q_0 < q_1 < \cdots < q_N$, and the slopes

$$s_k := \frac{r_{k-1} - r_k}{q_k - q_{k-1}} \quad (k = 1, \cdots, N) \tag{8}$$

satisfy the relations $s_1 > s_2 > \cdots > s_N > 0$.

Theorem 2 leads to a specification for determining the envelope of a set of points and imposes an ordering of the vertices which are used in the following algorithm:

A. Generic Envelope Guided Algorithm

1) Set $\lambda = 0$.
2) While the corner is the highest-numbered vertex of the envelope: Take a step with your favorite optimizer for (3) and determine the corner of the envelope of all the points.
3) While the corner of the envelope is near the left endpoint of the envelope: Apply a few (we used three steps of an optimization procedure to the new f in (3). Find the corner of the envelope which includes the new points. Adjust λ.

In Step 3) one would wish to choose λ to decrease both r and q. Locally, at x_k r would decrease if $\nabla r(x_k)^T[-\nabla r(x_k) - \lambda \nabla q(x_k)] < 0$ and q would decrease if $\nabla q(x_k)^T[-\nabla r(x_k) - \lambda \nabla q(x_k)] < 0$ which suggests that $\lambda_{\min} = \max[\epsilon, -g/\nabla q(x_k)^T \nabla q(x_k)] < \lambda < \lambda_{\max} = -\nabla r(x_k)^T \nabla r(x_k)/g$ where ϵ is the machine precision and $g = \nabla q(x_k)^T \nabla r(x_k)$. Note that the gradients of q and r are already required in most optimization procedures.

The most robust case of the general scheme, which we call the *tail strategy,* is similar to the adaptive annealing algorithm in [14]. One can use the envelope as a gatekeeper and begin every major iteration in Step 3) with x from the previous iteration. At the first encounter of Step 3) we set λ to $\hat{\lambda} = (\lambda_{\max}\lambda_{\min})^{1/2}$ and take three steps of the PCG algorithm. At subsequent encounters of Step 3) the current value of λ, λ_k, is reset to λ_{k+1} by the following scheme:

$$\lambda_{k+1} = \begin{cases} \min\left[\sigma\lambda_k, 0.5(\lambda_k + \lambda_{\max})\right] & \text{if } (q_k, r_k) \text{ below corner} \\ \max\left[\dfrac{\lambda_k}{2}, 0.5(\lambda_k + \lambda_{\min})\right] & \text{if } (q_k, r_k) \text{ above corner} \\ \lambda_k & \text{if } (q_k, r_k) \text{ is corner.} \end{cases}$$

For the tail strategy, σ was set to 4.0.

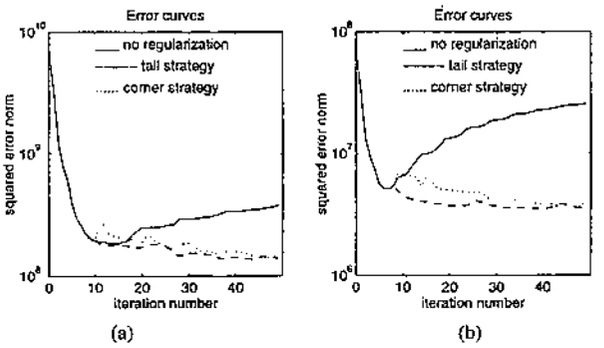


Fig. 2. Sum of squares of the difference between computed images and designed image for (a) 10 million and (b) one million photon pairs, with quadratic penalty term, using no regularization ($\lambda = 0$), the tail strategy and the corner strategy.

The *corner strategy* begins the major iterations instead with the corner of the last envelope (or with its only point if there is no corner). With the corner strategy our experimentation suggests setting σ to 2.0 to increase λ at a slower pace and initially setting λ to $\lambda_{\min}$.

Various measures can be used to assess how much the envelope bends at a vertex. With s_k defined as in (8) both

$$c_k := \frac{s_k}{s_{k-1}} \tag{9}$$

and

$$c_k := q_k \frac{s_k - s_{k+1}}{r_{k-1} - r_{k+1}} \tag{10}$$

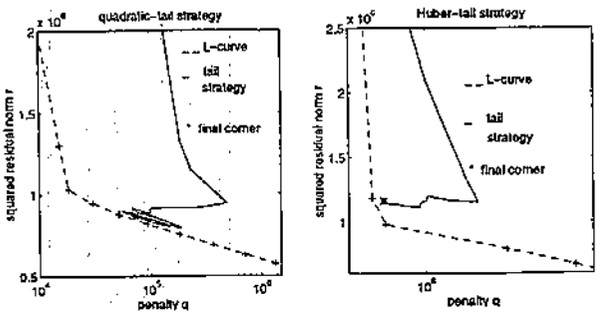
are invariant under rescaling of $q(x)$ and/or $r(x)$, and depend on the vertex and its two neighboring vertices only. In both cases, a large quotient indicates a strong bend, and the envelope is said to be *most bent* at the vertex (q_k, r_k) $(1 \le k \le N - 1)$ where c_k, given by (9) (*version 1*) or (10) (*version 2*), is largest. This point is called the *corner* of the envelope. In many cases, the two versions define the same corner.

III. Numerical Results

For our experiments, the sets of points whose envelope defines the current L-curves were generated by the PCG of Kaufman [8] that successfully handles nonnegativity constraints. Unless indicated otherwise, the starting vector for the first optimization was always uniform, except outside a circle inscribed to the box defining the grid, where zero entries were forced for all images; normalization was such that the total sum of entries equal to the number of photon pairs counted. We defined two problems, one with 1 million annihilations and one with 10 million annihilations.

Fig. 2 compares the size of the squared error norm during the iteration for both tail strategy and corner strategy with the corresponding result for the unregularized PCG method. The penalty term is the quadratic penalty; but for other penalty functions, the situation is qualitatively similar. Both strategies successfully avoid the growth of the error after the initial reduction and reduce the error significantly below that of the best unregularized iterate, with an advantage in speed for the tail strategy.

Fig. 3 gives a more detailed view of the progress by displaying the curves in the (q, r)-plane defined by the iterates and the (approximate) Tikhonov L-curve whose corner is the desired target. The values of λ used in the objective function generally were not close to the values of λ of a nearby point on the L-curve, since in our envelope strategies, the conjugate gradient method was for each λ terminated prematurely to avoid wasting iterations.



(a) (b)

Fig. 3. Tikhonov L-curve and iteration history for the (a) quadratic and (b) Huber penalty function with $\delta = 1$; one million photon pairs.

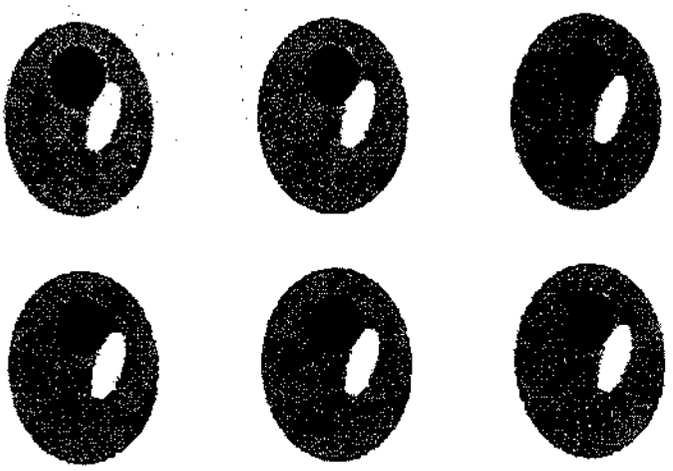


Fig. 4. Images for 10 million annihilations (top row) Quadratic, log for $\delta = 2^{13}$, log cosh for $\delta = 2^5$; (bottom row) Huber for, $\delta = 2^8$, multiquadric for $\delta = 2^8$, rational for $\delta = 2^{17}$.

Figs. 4 and 5 indicate the quality of the computer images using the tail strategy. (The corner strategy usually was slightly inferior). To simulate the effect of a posteriori image processing we transformed the components of the final vector linearly so that they span the range [1, 256], and then replaced all pixel values <100 by 100 and all pixel values >200 by 200, and rescaled the new gray values to fill the interval [1, 256] again.

In general, for optimally tuned δ, all penalty functions gave qualitatively similar results. However, images tended to be more blurred with the quadratic penalty, and redescending penalties like the log penalty in [6] occasionally led to additional artificial spots. The convex, asymptotically linear penalty terms removed most high frequency noise and were less sensitive to the precise choice of δ.

The high photon case of 10 million pairs is easy and produces reconstructed images of good quality. Best results were obtained using 32 iterations of the tail strategy with the quadratic penalty function (4) to compute a good starting point, used with the tail strategy for the nonlinear penalties (another 32 iterations). Fig. 4 shows the results for optimal choices of the scaling parameter δ, found by trying all powers of two in a reasonable range. All figures used version one to compute the corner. The most insensitive formula was the multiquadric penalty (6), which produced excellent results over a wide range of δ's. This insensitivity with respect to δ is very attractive since as yet no automatic method for choosing δ is available.

Fig. 5 shows images for the one million photon example. The noise in the data, the randomness in choosing the direction of the annihilating photons and the inaccuracy resulting from the discretization

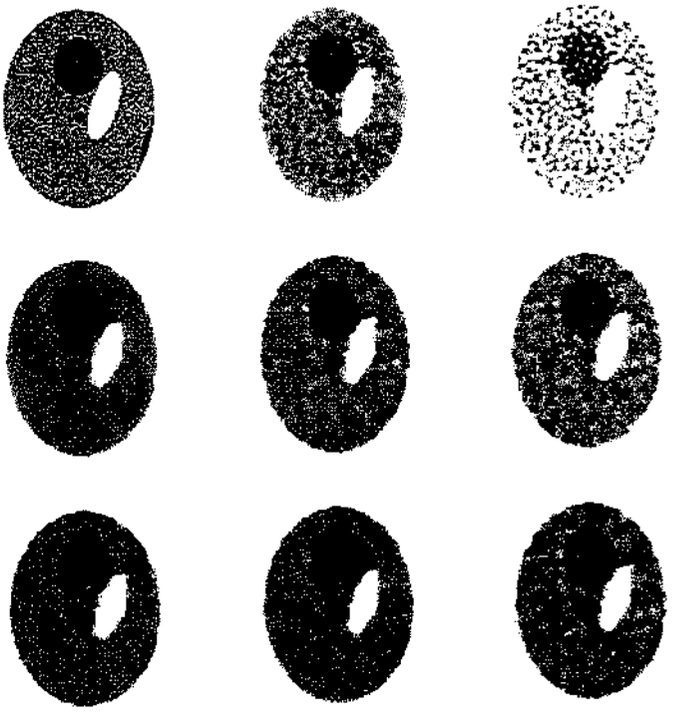


Fig. 5. Images for one million annihilations (top row) Phantom, and no penalty after seven (best) and 15 iterations (middle row) tail strategy, Huber penalty, $\delta = 8, 16, 32$, with (9) (bottom row) tail strategy, Huber penalty, $\delta = 8, 16, 32$, with (10).

used to set up the matrix only barely allows the reconstruction of the small features (representing idealized tumors). The top row displays the phantom used (for comparison), the optimal approximation—with minimal Euclidean distance to the phantom—computed by the PCG method without regularization (after seven iterations), and the deterioration resulting from continuing the unregularized iteration until iteration 15. The second and third rows give the final corner of the tail strategy (terminated after 32 iterations) for the Huber penalty with various scale parameters and the two versions of defining the corner. Here (but not in general), version two produced the most pleasing results.

Our experimentation in [9] suggests that the quality of the images for the nonquadratic penalty functions depends more strongly on δ for the one-million-count case than for the 10-million-count case. For the convex and asymptotically linear penalty functions, the sensitivity of the results with respect to δ is less.

IV. Conclusion

The new technique of envelope guided conjugate gradients was successfully tested for the PET image reconstruction problem with noisy data. At low noise level (high photon count) the resulting images reproduced the test image without the introduction of artifacts, such as spurious oscillations. At high noise level (low photon count) where a truncation method only shows the gross features, envelope guided conjugate gradient methods were able to reproduce all features of the test image, but tiny spurious features could not be fully suppressed.

References

[1] S. Geman and D. McClure, "Bayesian image analysis: An application to single photon emission tomography," in *Proc. Stat., Comput. Sect., Amer. Stat. Assoc.*, Washington, DC, 1985, pp. 12–18.

[2] P. Green, "Bayesian reconstructions from emission tomography using a modified EM algorithm," *IEEE Trans. Med. Imag.*, vol. 9, pp. 84–93, 1990.
[3] M. Hanke and P. C. Hansen, "Regularization methods for large-scale problems," *Surv. Math. Ind.*, vol. 3, pp. 253–315, 1993.
[4] P. C. Hansen, "Analysis of discrete ill-posed problems by means of the L-curve," *SIAM Rev.*, vol. 34, pp. 561–580, 1992.
[5] P. C. Hansen and D. P. O'Leary, "The use of the L-curve in the regularization of discrete ill-posed problems," *SIAM J. Sci. Comput.*, vol. 14, pp. 1487–1506, 1993.
[6] T. Hebert and R. Leahy, "A generalized EM algorithm for the 3-D Bayesian reconstruction from Poisson data using Gibbs priors," *IEEE Trans. Med. Imag.*, vol. 8, pp. 194–202, 1990.
[7] P. J. Huber, *Robust Statistics*. New York: Wiley, 1981.
[8] L. Kaufman, "Maximum likelihood, least squares, and penalized least squares for PET," *IEEE Trans. Med. Imag.*, vol. 12, pp. 200–214, 1993.
[9] L. Kaufman and A. Neumaier, "Image reconstruction through regularization by envelope guided conjugate gradients," AT&T Bell Laboratories, Murray Hill, NJ, Tech. Rep. CS94/4-14,1994. (available through the WWW, http://netlib.att.com/netlib/att/cs/doc/94/4-14.ps, 2MB)
[10] D. S. Lalush and B. M. W. Tsui, "Simulation evaluation of Gibbs prior distributions for use in maximum *a posteriori* SPECT reconstructions," *IEEE Trans. Med. Imag.*, vol. 11, pp. 267–275, 1992.
[11] K. Lange and J. Fessler, "Globally convergent algorithms for maximum a posteriori transmission tomography," *IEEE Trans. Image. Processing*, vol. 4, 1995.
[12] C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems*. Englewood Cliffs, NJ: Prentice-Hall, 1974.
[13] L. Shepp and Y. Vardi, "Maximum likelihood reconstruction in positron emission tomography," *IEEE Trans. Med. Imag.*, vol. MI-1, pp. 113–122, 1982.
[14] M. Symonds, Y.-S. Han, P. Santago, and W. Snyder, "Reconstruction of positron emission tomography images by using maximum a posteriori and mean field annealing techniques," in *Proc. SPIE: Physics of Med. Imag.*, 1994, vol. 2613, pp. 212–222.
[15] C. R. Vogel and M. E. Oman, "Iterative methods for total variation denoising," *SIAM J. Sci. Comput.*, vol. 17, pp. 227–238, 1996.
[16] G. Wahba, "Practical approximate solutions to linear operator equations when the data are noisy," *SIAM J. Numer. Anal.*, vol. 14, pp. 651–657, 1977.