An Algorithm for Computerized Tomography

Walter Vasilaky*

July 10, 2001

Abstract

An algorithm is given for the solution of a quadratic optimization problem in Computerized Tomography. It is shown that a dynamic programming approach yields a generalized Gram-Schmidt orthogonalization procedure that solves the quadratic optimization problem.

1. Introduction.

Commercial image reconstruction methods such as Computerized Tomography (CT) and Magnetic Resonance Imaging (MRI) are based on approximate transform methods, such as the Fast Fourier Transform (FFT). The transform based algorithms are simpler, faster, and use less storage than the optimal finite dimensional methods. However optimal finite dimensional algorithms, such as Bayesian [1],[2], minimum-variance [3], minimum-norm [4], and other [5] are more robust e.g., to data collection geometry, noisy data, low photon counts, or small number of views [5], [6]. Most of these finite dimensional reconstruction algorithms utilize quadratic optimization criteria that yield an associated set of linear (normal) equations whose solution is usually obtained by a variant of a Gauss-Seidel relaxation method [5], [7],[8]. We, on the other hand, do not try solve the associated system of normal equations but instead , motivated by the dynamic programming approach [9], [10], expresses the quadratic optimization criterion as a quadratic recurrence relation. The quadratic recurrence relation induces a linear recurrence relation which, quite unexpectedly, turns out to be a generalization of the Gram-Schmidt orthogonalization procedure. The Gram-Schmidt orthogonalization procedure is a special case of our generalized (approximate) Gram-Schmidt procedure and the Gram-Schmidt Orthogonalization solution of the linear least squares problem [11] is then a special case of our solution to the quadratic optimization problem.

1.1 Optimal finite dimensional methods.

In finite dimensional methods a set of basis pictures [12] $(b_1, ..., b_J)$ is chosen then a linear combination of the basis pictures is used to approximate a picture f(x,y). To obtain the simplest example of basis pictures we subdivide the picture region into a grid of n^2 squares (pixels) where $J = n^2$ then for $1 \leq j \leq J$ let

$$b_j(s,t) = \begin{cases} 1 & \text{if (s,t) in } j^{th} \text{ pixel,} \\ 0 & \text{otherwise.} \end{cases}$$

The estimated picture $\hat{f}$(s,t), for (s,t) in the picture region, is defined as

$$\hat{f}(s,t) = \sum_{j=1}^{J} \hat{x}_j b_j(s,t), J = n^2$$

where $\hat{x}_j$ is the average value of f in the j^{th} pixel. The problem is to find the image vector $\hat{X} = (\hat{x}_1, ..., \hat{x}_J)$ such that $\hat{f}$ in some sense is close to f.

*Long Island University, Brooklyn New York 11201, USA. Partial support by NSF, Grant DMS-8815324.

In Computerized Tomography X-ray pencils are passed through a source and for each ray the energy loss is measured. According to Lambert-Beer law, if E^o is the initial energy then $E = E^o e^{(-\int_L f ds)}$ is the final energy, where $\int_L f ds$ is the line integral of f along line L. If we denote the i^{th} ray by L_i and its initial and final energies by E_i^o and E_i respectively then for the i^{th} ray $-\ln(E_i/E_i^o)$ approximates the line integral $\int_{L_i} f ds$. Let $R_i f$ denote the line integral of f along the i^{th} line, then

$$R_i f \simeq R_i \hat{f} = \sum_{j=1}^{J} x_j (R_i b_j), 1 \leq i \leq I$$

where I is the number of rays, J the number pixels, and $R_i b_j$ the computed length of the intersection of the i^{th} ray with the j^{th} pixel. Actually as pointed out in [6] computing the area of the ray-strip with the pixels works better which would change the line integrals to strip integrals. Let r_{ij} denote the calculated values of $R_i b_j$, y_i the physical estimate of $R_i f$ i.e., $-\ln(E_i/E_i^o)$, W^{I1} the error vector and, $X^{J\times 1}$ the picture vector then

$$Y^{I\times 1} = R^{I\times J} X^{J\times 1} + W^{I\times 1}.$$

Minimizing the error vector W by least squares

$$\min_X \|Y - RX\|_2$$

may not give a unique solution since the linear system is overdetermined i.e, the number of rays, I, exceeds the number of pixels, J, and R may be rank deficient i.e., $rank(R) < J$. Instead we consider the following quadratic optimization criteria:

$$\min_X (\|Y - RX\|_2 + c\|X - X_0\|_2), c > 0 \tag{1}$$

where X_0 is the initial estimate f and c is some small positive constant. Intuitively the size of $c > 0$ represents the importance of the initial image X_0 in (1). Let R^T denote the transpose of R, and U the unit matrix, then the associated normal equations

$$(R^T R + cU)X = R^T Y + cX_0$$

have a unique solution, since $R^T R + cU$ is clearly a non-singular square matrix. Most finite dimensional reconstruction algorithms use various iterative methods to solve the associated normal equation. We, on the other hand, will derive an associated recurrence relations and solve the recurrence relations instead. Note that if we assume that X and W are uncorrelated Gaussian distributed with vector means $\vec{\mu_X}$ and $\vec{0_W}$ and covariances $\sigma_X^2 U$ and $\sigma_W^2 U$ respectively, then $c = \frac{\sigma_W^2}{\sigma_X^2}$ and for $X_0 = \vec{\mu_X}$ we get a Bayesian estimate of f [1],[2]. If we set c=1 and $X_0 = \bar{X}$, where $\bar{X}$ is the mean of X which is known in practice, then we get the minimum variance estimate of f [3]. Thus the quadratic minimization criterion in (1) is quite general.

2. Recurrence relations.

To simplify the algebraic manipulations, in what follows we will assume $X_0 = \vec{0}$. Note that the associated normal equations become

$$(R^T R + cU)X = R^T Y.$$

Consider the following quadratic form:

$$f_N(Y) = \min_{x_1,\ldots,x_N} (\sum_{i=1}^{I} (\sum_{j=1}^{N} r_{ij} x_j - y_i)^2 + c \sum_{j=1}^{N} x_j^2), c > 0 \tag{2}$$

defined for N=1,...,J and all Y. For $N \geq 2$ we have the following recurrence relation:

$$f_N(Y) = \min_{x_N}(f_{N-1}(Y - x_N R(:,N)) + cx_N^2), c > 0 \tag{3}$$

where $R(:,N)$ is the N^{th} column of the computed coefficient matrix R. Note that we have replaced vector Y, the observed data, with $Y - x_N R(:,N)$ in f_{N-1} to be minimized over $x_1, ..., x_{N-1}$.

Lemma 1. *$f_N(Y)$ is a quadratic form and is of the form $Y^T V_N Y$ where V_N is an $I \times I$ matrix independent of the data Y.*

Proof. Let R(:,1:J-N) be the $R^{I \times J}$ matrix with the last J-N columns replaced by zeros. The vector $\hat{X}$ that minimizes (2) is the solution to the associated normal equations, namely

$$\hat{X} = (R^T(:,1:J-N)R(:,1:J-N) + cU)^{-1} R^T(:,1:J-N)Y, c > 0$$

where U is a unit matrix. Substituting this expression on the right hand side in (2) yields the above form with

$$V_N = U - R(:,1:J-N)(R^T(:,1:J-N)R(:,1:J-N) + cU)^{-1} R^T(:,1:J-N) \blacksquare$$

Lemma 2. *Set $V_0 = U$, the unit matrix. The following recurrence relation holds for V_N*

$$V_N = V_{N-1} - \frac{(V_{N-1}R(:,N))(V_{N-1}R^T(:,N))}{c + R^T(:,N)V_{N-1}R(:,N)}, N = 1, ..., J.$$

Note: The numerator in the above expression is an outer product of an $I \times 1$ column vector $V_{N-1}R(:,N)$ and a $1 \times I$ row vectors $(V_{N-1}R(:,N))^T$, which is an $I \times I$ matrix, and the denominator is a scalar.

Proof. By Lemma 1 we can replace $f_{N-1}(Y - x_N R(:,N))$ in (3) with

$$(Y - x_N Y(:,N))^T V_{N-1}(Y - x_N Y(:,N))$$

and get

$$f_N(Y) = \min_{x_N}((Y - x_N R(:,N))^T V_{N-1}(Y - x_N R(:,N)) + cx_N^2).$$

Taking the derivative with respect to x_N and setting the result to zero we get

$$cx_N^* - (Y - x_N^* R(:,N))^T V_{N-1} R(:,N) = 0$$

where x_N^* is the minimizing value. Solving for x_N^* we get

$$x_N^* = \frac{Y^T V_{N-1} R(:,N)}{c + R^T(:,N)V_{N-1}R(:,N)}.$$

Now

$$\begin{aligned}
f_N(Y) &= Y^T V_N Y \\
&= (Y - x_N^* R(:,N))^T V_{N-1}(Y - x_N^* R(:,N)) + cx_N^{*2} \\
&= (Y - x_N^* R(:,N))^T V_{N-1} Y - x_N^*(Y - x_N^* R(:,N))^T V_{N-1} R(:,N) + cx_N^{*2} \\
&= (Y - x_N^* R(:,N))^T V_{N-1} Y + x_N^*(cx_N^* - (Y - x_N^* R(:,N))^T V_{N-1}) R(:,N)) \\
&= (Y - x_N^* R(:,N))^T V_{N-1} Y \\
&= (Y^T - x_N^* R^T(:,N)) V_{N-1} Y \\
&= Y^T V_{N-1} Y - x_N^* R^T(:,N) V_{N-1} Y \\
&= Y^T V_{N-1} Y - \frac{Y^T V_{N-1} R(:,N)}{c + R^T(:,N)V_{N-1}R(:,N)} R^T(:,N) V_{N-1} Y \\
&= Y^T V_{N-1} Y - \frac{Y^T (V_{N-1}R(:,N))(V_{N-1}R(:,N))^T Y}{c + R^T(:,N)V_{N-1}R(:,N)} \\
&= Y^T (V_{N-1} - \frac{(V_{N-1}R(:,N))(V_{N-1}R(:,N))^T}{c + R^T(:,N)V_{N-1}R(:,N)}) Y \blacksquare
\end{aligned}$$

The penultimate equality follows from the symmetry of V_N, N=1,2,...,J.

Definition. *Two none-zero independent parameterized vectors* $A(c) = (a_1(c), ...a_n(c))^T$ *and* $B(c) = (b_1(c), ..., b_N(c))^T$ *are said to be pseudo-orthogonal iff*

$$\lim_{c\to 0} A(c)B^T(c) = 0.$$

Theorem 1. *Let* $V_0 = U$*, the* $I \times I$ *unit matrix, and define matrix Q(c) as* $Q(:, N) = V_{N-1}(c)R(:, N)$*, N=1,...,J. Then the columns of matrix Q(c) are pseudo-orthogonal.*

Proof. Expanding the recurrence relation in Lemma 2 we get

$$V_N = U - \sum_{k=1}^{N-1} \frac{Q(:, k)Q^T(:, k)}{c + Q^T(:, k)Q(:, k)}. \tag{4}$$

Note that we replaced in Lemma 2 $R^T(:, N)V_{N-1}R(:, N)$ by $Q^T(:, k)Q(:, k)$ by virtue of the fact that $R^T(:, N)$ can be written as $Q(:, N) + Q^{\perp}(:, N)$. Using (4) the columns of matrix Q(c) can now be computed by the following recursion

$$\begin{aligned}
Q(:, 1) &= UR(:, 1)\\
Q(:, 2) &= (U - \frac{Q(:, 1)Q^T(:, 1)}{c + Q^T(:, 1)Q(:, 1)})R(:, 2)\\
Q(:, N) &= (U - \sum_{k=1}^{N-1} \frac{Q(:, k)Q^T(:, k)}{c + Q^T(:, k)Q(:, k)})R(:, N), N = 3, ..., J.
\end{aligned}$$

Except for the constant $c > 0$ in the denominators the above recursion is the classical Gram-Schmidt orthogonalization procedure. Hence the column vectors of Q(c) are pseudo-orthogonal ■

Note: Accept for the $c > 0$ in the denominator the V_n terms are similar to the Householder Reflections matrices. The roundoff properties associated with Householder matrices are very favorable [14] and with $c > 0$ in the denominator we can expect that the roundoff properties to be even better.

Theorem 2. *The vector* $\hat{X}(c) = (\hat{x}_1(c), ..., \hat{x}_J(c))$ *that minimizes quadratic form (2) is given by*

$$\begin{aligned}
\hat{x}_J(c) &= \frac{Q^T(:, J)}{c + Q^T(:, J)Q(:, J)}Y\\
\hat{x}_{J-1}(c) &= \frac{Q^T(:, J-1)}{c + Q^T(:, J-1)Q(:, J-1)}(Y - \hat{x}_J R(:, J))\\
\hat{x}_{J-N}(c) &= \frac{Q^T(:, J-N)}{c + Q^T(:, J-N)Q(:, J-N)}(Y - \sum_{k=J}^{J-(N-1)} \hat{x}_k(c)R(:, k)), N = 2, ..., J-1.
\end{aligned}$$

.

Proof. If we substitute for f_{J-1} in (3) the expression

$$(Y - x_J R(:, J))^T V_{J-1}(Y - x_J R(:, J))$$

with N=J we get

$$\min_{x_J}((Y - x_J R(:, J))^T V_{J-1}(Y - x_J R(:, J) + cx_J^2)$$

since this is a minimization of a quadratic function in x_J we get

$$\begin{aligned}
\hat{x}_J(c) &= \frac{R^T(:, J))V_{J-1}Y}{c + R^T(:, J))V_{J-1}R(:, J)}, or\\
\hat{x}_J(c) &= \frac{Q^T(:, J)}{c + Q^T(:, J)Q(:, J)}Y.
\end{aligned}$$

To compute $\hat{x}_{J-1}$ we substitute

$$(Y - \hat{x}_J R(:,J) - x_{J-1}R(:,J-1))^T V_{J-2}(Y - \hat{x}_J R(:,J) - x_{J-1}R(:,J-1))$$

in (3) with N= J-1 and minimize with respect to x_{J-1} we get

$$\hat{x}_{J-1}(c) = \frac{Q^T(:,J-1)}{c + Q^T(:,J-1)Q(:,J-1)}(Y - \hat{x}_J R(:,J))$$

We continue this way for N=2,...,J-1 to obtain

$$\hat{x}_{J-N}(c) = \frac{Q^T(:,J-N)}{c + Q^T(:,J-N)Q(:,J-N)}(Y - \sum_{k=J}^{J-(N-1)} \hat{x}_k(c)R(:,k))\blacksquare$$

Note that $\frac{Q^T(:,J-N))}{c+Q^T(:,J-N)Q(:,J-N)}$ are pseudo-orthonormal in the sense that they are pseudo-orthogonal and

$$\lim_{c\to 0} \frac{Q^T(:,J-N)Q(:,J-N)}{c + Q^T(:,J-N)Q(:,J-N)} = 1.$$

Corollary 1. *For $R^{I\times J}$, if I=J, and rank(R) $= J$, then the recursion for Q(:,N) in the proof of Theorem 1 is the Gram-Schmidt orthogonalization procedure.*

Proof. If we set c=0 in the recursion for Q(:,N) we get the Gram-Schmidt orthogonalization procedure $\blacksquare$

Corollary 2. *The $\lim_{c\to 0}\hat{X}(c) = R^+Y$ for $I > J$, and $rank(R) \leq J$, where R^+ is the pseudo-inverse of R.*

Proof. It's shown in [13] that $lim_{c\to 0}(R^TR + cU)^{-1}R^T = lim_{c\to 0}R^T(RR^T + cU)^{-1} = R^+$ exists and for any vector Y R^+Y is the vector of minimum norm among which minimize

$$\min_X \|Y - RX\|_2$$

where R^+ is the pseudo-inverse of R. By Theorem 2

$$\hat{X}(c) = (R^TR + cU)^{-1}R^TY$$

hence

$$\lim_{c\to 0}\hat{X}(c) = R^+Y\blacksquare$$

Corollary 2 is an interesting result even though R^+Y is not very useful for Computerized Tomography. The problem is that R^+Y is not even a continuous function of the data so that small changes in R and Y can induce arbitrarily large changes in R^+Y [14] . In Computerized Tomography the value of $c > 0$ plays the role of an image focusing parameter. For simplicity we took the initial image to be blank i.e, $X_0 = 0$, so that $c > 0$ would be chosen relatively small since a blank image would not be a good initial estimate of the actual image. In actual practice the choice of $c > 0$ would depend on many factors and would be difficult to guess. Error analysis for the Modified Gram-Schmidt solution of the least squares problem in [11], [14] show that the condition number κ_2 determines how sensitive the solution is to perturbations of the R and Y. In our case the condition number is

$$\kappa_2(R^TR + cU) = \frac{\sigma^2_{max} + c}{\sigma^2_{min} + c}$$

,where σ^2_{max} and σ^2_{min} are respectively the maximum and minimum singular values of matrix R. So the choice of $c > 0$ controls the stability of the algorithm. However a large $c > 0$ may distort the picture by giving too much weight to the initial image X_0, thus we have to balance the stability of the algorithm against bias toward X_0. Fortunately in Computerized Tomography R is very sparse, so that out of $I \times J$ entries only about $\sqrt{I \times J}$ entries are none-zero. In fact [14]

$$\sigma_{max} = \|R\|_2 \leq \sqrt{\|R\|_1 \|R\|_\infty}$$

where

$$\|R\|_1 = \max_{1 \leq j \leq J} \sum_{i=1}^{I} \mid r_{ij} \mid$$

and

$$\|R\|_\infty = \max_{1 \leq i \leq J} \sum_{j=1}^{J} \mid r_{ij} \mid$$

so an upper bond for σ_{max} is easy to compute and in practice κ_2 is sufficiency small. However, a systematic error analysis for our algorithm remains to be done.

Conclusion.

Motivated by a problem in Computerized Tomography we have developed an algorithm for the minimum quadratic form criterion. The Gram-Schmidt orthogonalization solution and the minimum norm solutions to the Least Squares problem are special cases of the presented algorithm. The new feature of the presented algorithm is that the solution is a parameterized vector. The best value for the parameter would probably have to be be determined empirically and would be different for different types of images. Having a parameterized solution is an advantage in the sense that any imaging device is better if it can be adjusted to context.

References

[1] G.T. Herman. Application of Maximum Entropy and Bayesian optimization methods to image reconstruction from projections in Maximum Entropy and Bayesian Methods Inverse Problems. D. Reidel, Dordreecht, Holland, 1985, 319–338.

[2] G. T. Herman, M. Hurwitz, A. Lent, H. P. Lung. On the Bayesian Approach to Image Reconstruction, Inf. Conrol, 42, 1979, 60–71.

[3] S. L. Wood, M. Mort, and A. Macovski. Stochastic Methods Applied to Medical Image Reconstuction. Proc. 1977 IEEE Conf. Contr., New Orleans, Louisiana, Dec. 1979.

[4] G. T. Herman, A. Lent, and A. Lutz. Relaxation Methods for Image Reconstruction. Comm. A.C.M. 21, 1978, 152–158.

[5] G. T. Herman. Image Reconstruction From Projections. Academic Press, New York, 1982.

[6] L. A. Shepp and J. B. Kruskal. Computorized Tomogaphy: The New Medical X-Ray Technology. Amer. Math Monthly 1985, 420 - 439.

[7] G. T. Herman, and A. Lent. Quadratic Optimization for Image Reconsruction. Comput. Graphics Image Processing 5, 1976, 319 - 322.

[8] A. V. Lakshinarayanan and A. Lent. Methods of Least Squares and SIRT in Reconstruction. J. Theor., 1976, 267 - 295.

[9] R. Bellman. Dynamic Programming. Printeton University Press, Printceton, New Jersey, 1957.

[10] R. Bellman and J.B. Dreyfus. Applied Dynamic Programming. Printeton University Press, Printceton, New Jersey, 1962.

[11] A Bjorck. Solving Linear Least Squares Problems By Gram-Schmidt Orthogonalization. BIT 7 1967, 1 - 21.

[12] M. H. Buonocore, W. R. Broody, A. Macovski. A Natural Pixel Decomposition Two-Dimensional Image Reconstruction. IEEE Transactions on Biomedical Engineering Vol. BME-28. No. 2, Feb. 1981.

[13] T. Boullion and P. Odell. Generalized Inverse Matrices. Wiley (intercience), New York, 1971.

[14] G. H. Golub and C. F. Van Loan. Matrix Computations, Second Ediion. The John Hopkins University Press, 1989.