Faster Maximum-Likelihood Reconstruction via Explicit Conjugation of Search Directions

Guillem Pratx, Andrew J. Reader, Craig S. Levin

Abstract— **Conjugate gradient (CG) is useful to perform maximum-likelihood (ML) reconstruction in positron emission tomography (PET). Although first derived to solve linear systems of equations, CG has been used for non-quadratic objectives. For the log-likelihood of inhomogeneous Poisson processes, the search directions generated by the generic Polak–Ribière formulation are not quite conjugate. We investigated a new CG formulation specific to the ML criterion in PET that preserves the conjugation. We first established a new relationship between the search direction and the image residual. We then derived a new method to generate a basis of search directions conjugate in the ML sense. Conjugation was enforced explicitly by forming the new search direction from a linear combination of the gradient and all the past search directions.**

The new formulation converges faster to the ML optimal solution. The equivalent of 50 Polak–Ribière iterations is reached in 39 iterations (1.3× faster) and the equivalent of 2000 Polak–Ribière iterations is reached in 451 iterations (4.4× faster). The truncation of the new formulation converges at the same rate as the Polak–Ribière method. The new formulation requires only a negligible amount of extra computation, but very large amounts of memory to store past search directions.

I. Introduction

The conjugate gradient (CG) algorithm was first introduced to solve large systems of linear equations, by optimizing a quadratic objective function [1]. The method was later applied with little modification to non-quadratic objectives (non-linear CG). Both approaches have been used in positron emission tomography (PET) image reconstruction, for the weighted least-square (WLS) [2], [3] and the maximum-likelihood (ML) criterion [4].

CG search directions are formed using the Polak–Ribière formulation. This generic approach assumes a quadratic objective. Thus, for the ML criterion, the search directions are not quite conjugate, and the the convergence rate is affected. We researched a new formulation for ML-CG which derives directly from the ML criterion. We discovered an approximate conjugation relationship for ML that is analogous to the one obtained for the quadratic WLS criterion. This relationship was used to define explicit conditions on β, which control the search direction, and derive an alternate formulation for ML-CG. We show that the new formulation converges faster to the ML estimate than Polak–Ribière.

Work presented at the 2008 *IEEE NSS-MIC* conference in Dresden, Germany. Conference record received November 14, 2008. This work was supported by the National Institutes of Health (NIH) under grants R01CA119056, R33EB003283, R01CA120474 and the Stanford Bio-X graduate fellowship program.

*G. Pratx and C. S. Levin are with Stanford University, Department of Radiology, and Molecular Imaging Program at Stanford. (email:pratx@stanford.edu). A. J. Reader is with McGill University, Montréal, QC, Montréal Neurological Institute.

II. Theory

An emission tomography dataset is a collection of individual events. Inverse methods that account for the counting statistics of the signal are preferable. The number of positrons that are emitted in each of the N image voxels, during a fixed time frame, is well described a Poisson random vector $\mathbf{X}$, the mean of which will be noted $\mathbf{x}$. The number of counts recorded in each of the P projection bins can be modelled by another Poisson random vector $\mathbf{Y}$, with mean $\mathbf{y}$. A PET projection dataset consists of a measured realization $\mathbf{m}$ of $\mathbf{Y}$. The imaging system is linear, thus a system matrix A relates the mean of the two random vectors

$$\mathbf{y} = A\mathbf{x}. \tag{1}$$

The goal of tomographic image reconstruction is to produce an estimate of the parameters $\mathbf{x}$ and $\mathbf{y}$. Statistical methods formulate and optimize an objective function that describes the consistency of the estimate with the measured data. Two such objectives are the WLS and the ML criterion. Section II-A reviews the rationale for using CG for quadratic objective such as WLS [5]. In section II-B, a better formulation for CG, specific to the ML criterion, is derived. This new formulation outperforms applying the conventional Polak–Ribière formulation directly to the non-quadratic ML objective.

A. Weighted Least-Squares

1) Problem Formulation: The WLS estimate $\mathbf{x}^*_{\text{wls}}$ is the solution to the quadratic problem with equality constraints

$$\begin{aligned} \text{maximize} \quad & f_{\mathbf{m}}(\mathbf{y}) = -\tfrac{1}{2}\sum_{i=1}^{P}\left(\frac{y_i - m_i}{m_i}\right)^2 \\ \text{subject to} \quad & A\mathbf{x} = \mathbf{y}. \end{aligned} \tag{2}$$

Inequality constraints ($\mathbf{x} \geq 0$ and $\mathbf{y} \geq 0$) can be added to problem (2). Problem (2) can be formulated as an unconstrained optimization problem by substituting $A\mathbf{x}$ for $\mathbf{y}$.

The gradient of the WLS objective with respect to $\mathbf{x}$ is given by

$$\mathbf{g}_n = \nabla_{\mathbf{x}} f_{\mathbf{m}}(\mathbf{y}_n) \tag{3}$$

$$= A^T M^{-1}(\mathbf{m} - \mathbf{y}_n), \tag{4}$$

where $\mathbf{y}_n = A\mathbf{x}_n$ and $M = \mathbf{diag}(\mathbf{m})$.

2) Conjugate Gradient for WLS: CG is the iterative method of choice to optimize a quadratic objective. This ascent method alternates the computation of a search direction and a step size, producing a sequence of estimates $\mathbf{x}_n$. The CG search direction $\mathbf{d}_n$ combines the gradient of the objective function $f_{\mathbf{m}}$ and the previous search direction

$$\mathbf{d}_n = \mathbf{g}_n + \beta_n \mathbf{d}_{n-1}, \tag{5}$$

where β_n is given by the Polak–Ribière formulation

$$\beta_n = \frac{(\mathbf{g}_n - \mathbf{g}_{n-1})^T \mathbf{g}_n}{\mathbf{g}_{n-1}^T \mathbf{g}_{n-1}}. \tag{6}$$

The step size α_n is computed by performing a line search

$$\alpha_n = \arg\max_{\alpha} f(\mathbf{x}_n + \alpha \mathbf{d}_n), \tag{7}$$

so the new image estimate is

$$\mathbf{x}_{n+1} = \mathbf{x}_n + \alpha_n \mathbf{d}_n. \tag{8}$$

3) Conjugation in WLS-CG: CG has been shown to converge to the optimal solution of a quadratic problem faster than gradient ascent. By maintaining conjugation in the search directions, CG constraints the image residual to an ever decreasing subspace.

Following (7), the gradient at $\mathbf{x}_{n+1}$ is orthogonal to the search direction $\mathbf{d}_n$

$$\mathbf{d}_n^T \mathbf{g}_{n+1} = 0. \tag{9}$$

The image residual is defined as the difference between the current image estimate and the optimal value of the problem

$$\mathbf{e}_{n+1} = \mathbf{x}_{n+1} - \mathbf{x}_{\text{wls}}^*. \tag{10}$$

The optimal solution $\mathbf{x}_{\text{wls}}^*$ of problem (2) satisfies $\nabla_{\mathbf{x}} f_{\mathbf{m}}(A\mathbf{x}_{\text{wls}}^*) = 0$. Using this definition, the gradient can be expressed as a linear function of the residual

$$\mathbf{g}_n = A^T M^{-1} A \mathbf{e}_n. \tag{11}$$

From (9), the search direction and the image residual are conjugate by $C = A^T M^{-1} A$

$$\mathbf{d}_n^T C \mathbf{e}_{n+1} = 0. \tag{12}$$

Following (8) and (10), the residual $\mathbf{e}_{n+1}$ can be reformulated as

$$\mathbf{e}_{n+1} = \underbrace{\mathbf{x}_{i+1} + \sum_{j=i+1}^{n} \alpha_j \mathbf{d}_j}_{\mathbf{x}_{n+1}} - \mathbf{x}_{\text{wls}}^* \tag{13}$$

$$= \mathbf{e}_{i+1} + \sum_{j=i+1}^{n} \alpha_j \mathbf{d}_j \tag{14}$$

for $i < n$. As a result,

$$\mathbf{d}_i^T C \mathbf{e}_{n+1} = \underbrace{\mathbf{d}_i^T C \mathbf{e}_{i+1}}_{0} + \sum_{j=i+1}^{n} \alpha_j \mathbf{d}_i^T C \mathbf{d}_j \tag{15}$$

$$= \sum_{j=i+1}^{n} \alpha_j \mathbf{d}_i^T C \mathbf{d}_j. \tag{16}$$

The key to WLS-CG is to use Gram-Schmidt orthogonalization procedure to construct a basis of conjugate search directions

$$\mathbf{d}_i^T C \mathbf{d}_j = 0, \quad i \neq j. \tag{17}$$

For such a basis of conjugate directions, the residual is conjugate to all the past search directions

$$\mathbf{e}_n^T C \mathbf{d}_i = 0, \quad i < n. \tag{18}$$

In other words, $n-1$ components of $\mathbf{e}_n$ are zero in the C-orthogonal basis defined by $(\mathbf{d}_i)$. As n increases, the image residual $\mathbf{e}_n$ is constrained to a subspace of decreasing dimension. In other words, the n^{th} update does not undo the work achieved during the previous steps. The exact value of the residual $\mathbf{e}_n$ is never known during the optimization, yet after $N+1$ iterations it is exactly zero (at least in theory). Moreover, it has been observed that small residuals can be obtained even after a number of iterations much smaller than N. Convergence is particularly fast when the eigenvalues of C are clustered.

For WLS-CG, conjugate search directions can be interpreted as being orthogonal when projected and normalized by the measured standard deviation of each sinogram bin

$$\left(M^{-1/2} A \mathbf{d}_i\right) \perp \left(M^{-1/2} A \mathbf{d}_j\right). \tag{19}$$

4) Polak–Ribière Formulation: The Polak–Ribière formulation is a method to build recursively a basis of conjugate search directions (17). The new search direction $\mathbf{d}_n$ is chosen in the subspace spanned by the gradient $\mathbf{g}_n$ and all the past search directions:

$$\mathbf{d}_n = \mathbf{g}_n + \sum_{j=1}^{n-1} \beta_{n,j} \mathbf{d}_j. \tag{20}$$

The coefficients β_n must be such that $\mathbf{d}_n$ satisfies (17), which yields

$$\mathbf{g}_n^T C \mathbf{d}_i + \beta_{n,i} \mathbf{d}_i^T C \mathbf{d}_i = 0, \tag{21}$$

for $i < n$. The $n-1$ equations are uncoupled and can be solved independently.

As a result of the Gram-Schmidt procedure, the gradients form an orthogonal basis. This can be shown by combining (11) and (20). Let us first assume that $i < j$, then

$$\mathbf{g}_i^T \mathbf{g}_j = (\mathbf{d}_i - \sum_{k=1}^{i-1} \beta_{i,k} \mathbf{d}_k)^T C \mathbf{e}_j. \tag{22}$$

The residual $\mathbf{e}_j$ is conjugate to all the past search directions (18), therefore

$$\mathbf{g}_i^T \mathbf{g}_j = 0, \quad i \neq j. \tag{23}$$

The quantities used in (21) can be reformulated using the gradients only

$$\begin{aligned} \mathbf{g}_n^T C \mathbf{d}_i &= \frac{1}{\alpha_i} \mathbf{g}_n^T C (\mathbf{x}_{i+1} - \mathbf{x}_i) \\ &= \frac{1}{\alpha_i} \mathbf{g}_n^T C (\mathbf{e}_{i+1} - \mathbf{e}_i) \\ &= \frac{1}{\alpha_i} \mathbf{g}_n^T (\mathbf{g}_{i+1} - \mathbf{g}_i) \\ \mathbf{g}_n^T C \mathbf{d}_i &= \begin{cases} \mathbf{g}_n^T (\mathbf{g}_n - \mathbf{g}_{n-1}) / \alpha_{n-1}, & i = n-1 \\ 0, & \text{otherwise.} \end{cases} \end{aligned} \tag{24}$$

Similarly, using the orthogonality of the gradients,

$$\mathbf{d}_i^T C \mathbf{d}_i = -\mathbf{g}_{i-1}^T \mathbf{g}_{i-1} / \alpha_{i-1}. \tag{25}$$

As a result, $\beta_{n,i} = 0$ for $i < n-1$ (21). Solving for the nonzero coefficient ($\beta_{n,n-1}$) yields the Polak–Ribière formulation (6). Due to the orthogonality of the gradients, the equivalent Fletcher–Reeves formulation can be derived

$$\beta_n^{\mathrm{FR}} = \frac{\mathbf{g}_n^T \mathbf{g}_n}{\mathbf{g}_{n-1}^T \mathbf{g}_{n-1}}. \tag{26}$$

However, for non-quadratic objectives, the gradients are not orthogonal and Polak–Ribière often converges faster than Flectcher–Reeves [5].

B. Maximum-Likelihood

1) Problem Formulation: The ML estimate $\mathbf{x}_{\mathrm{ml}}^*$ solves the convex problem

$$\begin{aligned} \text{max.} \quad & p_{\mathbf{m}}(\mathbf{y}) = \sum_{i=1}^{P} -\mathbf{y}_i + \mathbf{m}_i \log(\mathbf{y}_i) - \log(\mathbf{m}_i!) \\ \text{sub. to} \quad & A\mathbf{x} = \mathbf{y}. \end{aligned} \tag{27}$$

The nonnegativity constraint on the projection $\mathbf{y} \geq 0$ is enforced implicitely by the objective.

Although the log-likelihood objective $p_{\mathbf{m}}(\mathbf{y})$ is non-quadratic, it has been observed that problem (27) could be optimized efficiently using CG with the Polak–Ribière formulation [4]. Fast convergence can be expected when the Hessian matrix varies slowly between iterations. Near the optimal value, the objective is well approximated by a quadratic function, and as a result conventional CG is efficient. However, in the early iterations, CG takes large steps and the Hessian matrix can change substantially between iterations. As a result, the search directions are not quite conjugate and the image residual $\mathbf{e}_n$ does not get constrained to a subspace of the full image volume.

A new CG method, specific to the ML objective, was investigated. Inspiration was drawn from the WLS case to derive a new approximate conjugation relationship specific to the ML criterion, and to design a method to form search directions consistent with that relationship.

2) Conjugation in ML-CG: The image residual is defined similarly to (10)

$$\mathbf{e}_n = \mathbf{x}_n - \mathbf{x}_{\mathrm{ml}}^*, \tag{28}$$

where $\mathbf{x}_{\mathrm{ml}}^*$ satisfies $\nabla_{\mathbf{x}} p_{\mathbf{m}}(A\mathbf{x}_{\mathrm{ml}}^*) = 0$.

The gradient of the ML objective with respect to $\mathbf{x}$ is given by

$$\begin{aligned} \mathbf{h}_n &= \nabla_{\mathbf{x}} p_{\mathbf{m}}(\mathbf{y}_n) && (29) \\ &= A^T Y_n^{-1} (\mathbf{m} - \mathbf{y}_n) && (30) \end{aligned}$$

where $\mathbf{y}_n = A\mathbf{x}_n$ and $Y_n = \mathbf{diag}(\mathbf{y}_n)$. For the ML gradient $\mathbf{h}_n$, the difference between the measured and estimated projection $(\mathbf{m}_i - \mathbf{y}_i)$ is scaled by the inverse of the *estimated* variance $\mathbf{y}_n$, while in the WLS gradient $\mathbf{g}_n$, this difference is scaled by the inverse of the *measured* variance $\mathbf{m}$.

The ML gradient can be expressed as a function of the residual.

$$\mathbf{h}_n = A^T Y_n^{-1} \Lambda A \mathbf{e}_n, \tag{31}$$

where $\Lambda = \mathbf{diag}(\mathbf{m}/A\mathbf{x}_{\mathrm{ml}}^*)$. This expression depends upon the unknown optimal solution to the ML problem, therefore Λ is approximated by the identity matrix. This approximation is equivalent to $A\mathbf{x}_{\mathrm{ml}}^* = \mathbf{m}$. It should be noted that if this relationship is true, then indeed $\mathbf{x}_{\mathrm{ml}}^*$ is the ML optimal solution. An approximate relationship can then be established between the gradient and the image residual, independently of the optimal solution:

$$\mathbf{h}_n \approx A^T Y_n^{-1} A \mathbf{e}_n. \tag{32}$$

The gradient is orthogonal to the search direction when α_n is computed by a line search

$$\mathbf{d}_n^T \mathbf{h}_{n+1} = 0. \tag{33}$$

Following (32), a new conjugation relationship can be formulated for ML

$$\mathbf{d}_n^T B_{n+1} \mathbf{e}_{n+1} \approx 0, \tag{34}$$

where $B_{n+1} = A^T Y_{n+1}^{-1} A$. Unlike in the WLS case, the conjugation matrix B_{n+1} varies with the iterations. Nevertheless, this conjugation relationship can be exploited to constrain the residual.

Similarly to (16), the image residual $\mathbf{e}_n$ satisfies

$$\mathbf{d}_i^T B_{i+1} \mathbf{e}_{n+1} \approx \sum_{j=i+1}^{n} \alpha_j \mathbf{d}_j^T B_{i+1} \mathbf{d}_i \tag{35}$$

for $i = 1 \ldots n-1$. We therefore explored a method that recursively builds a sequence of search directions such that

$$\mathbf{d}_j^T B_{i+1} \mathbf{d}_i = 0, \quad i < j. \tag{36}$$

For such a sequence of search directions, the image residuals $\mathbf{e}_n$ are constrained to a subspace of decreasing dimension as the image iterates $\mathbf{x}_n$ approach optimality

$$\mathbf{d}_i^T B_{i+1} \mathbf{e}_{n+1} \approx 0, \quad i < n. \tag{37}$$

The basis of search directions conjugated in the ML sense (36) can be interpreted as follows. The projected search directions $A\mathbf{d}_i$ and $A\mathbf{d}_j$ ($i < j$), scaled by the estimate of the variance at the i^{th} iteration, are orthogonal:

$$Y_{i+1}^{-1/2} A \mathbf{d}_i \perp Y_{i+1}^{-1/2} A \mathbf{d}_j. \tag{38}$$

3) Explicit Conjugation of Search Directions: An algorithm similar to Gram-Schmidt is employed to produce a basis of search directions that satisfy (36). The search direction $\mathbf{d}_n$ is formed by combining linearly the gradient $\mathbf{h}_n$ and all the past search directions

$$\mathbf{d}_n = \mathbf{h}_n + \sum_{j=1}^{n-1} \beta_{n,j} \mathbf{d}_j. \tag{39}$$

Applying conditions (36) yields a system of $n-1$ linear equations with $n-1$ variables $\beta_{n,j}$

$$\mathbf{h}_n^T B_{i+1} \mathbf{d}_i + \sum_{j=1}^{n-1} \beta_{n,j} \mathbf{d}_j^T B_{i+1} \mathbf{d}_i = 0 \tag{40}$$

for $i = 1 \ldots n-1$. Unlike for the WLS objective, the equations are coupled. They can be equivalently represented in matrix notation

$$H^{(n)}\boldsymbol{\beta}_n = \mathbf{c}_n, \tag{41}$$

where $\boldsymbol{\beta}_n = (\beta_{n,1}, \ldots, \beta_{n,n-1})$, $\mathbf{c}_n$ is the vector defined as

$$c_{n,i} = -\mathbf{h}_n^T B_{i+1}\mathbf{d}_i \tag{42}$$

and $H^{(n)}$ is a lower-triangular matrix

$$H_{ij}^{(n)} = \mathbf{d}_j^T B_{i+1}\mathbf{d}_i. \tag{43}$$

The matrix $H^{(n)}$ can be constructed recursively since $H_{ij}^{(n+1)} = H_{ij}^{(n)}$ for $i, j \leq n-1$.

A truncated formulation can also be implemented by assuming $\beta_{n,j} = 0$ for $j \leq n-2$. In that case, the non-zero component can be formulated as

$$\beta_{n,n-1}^{\text{trunc}} = -\frac{\mathbf{h}_n^T B_n \mathbf{d}_{n-1}}{\mathbf{d}_{n-1}^T B_n \mathbf{d}_{n-1}} \tag{44}$$

III. Evaluation

A 2-D 128×128 Shepp Logan phantom with a positive-valued surrounding background was simulated. No cold region was present in the phantom to circumvent the as yet unresolved issue of constraining voxels to remain nonnegative. A noise-free dataset was obtained by computing 192 parallel-beam projection views. A noisy dataset was also produced by generating a realization of a Poisson random vector parameterized by the noise-free projections, resulting in 35 millions counts.

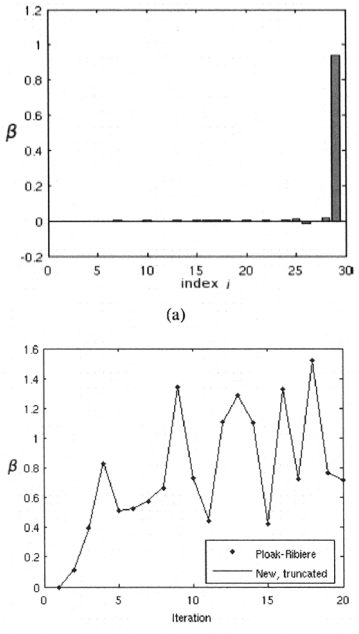

Fig. 1. (a) Typical coefficients for $\beta_{n,i}$ for $n = 30$ and $i = 1 \ldots 29$. The last search direction $\mathbf{d}_{n-1}$ is weighted more heavily than older ones for computing the new search direction $\mathbf{d}_n$, but older search directions also contribute to the final search direction. When the contribution of these older search directions is ignored, the new formulation (44) is within 0.3% of Polak–Ribière (6), as shown (b) by plotting β for both formulations for 20 iterations of ML-CG.

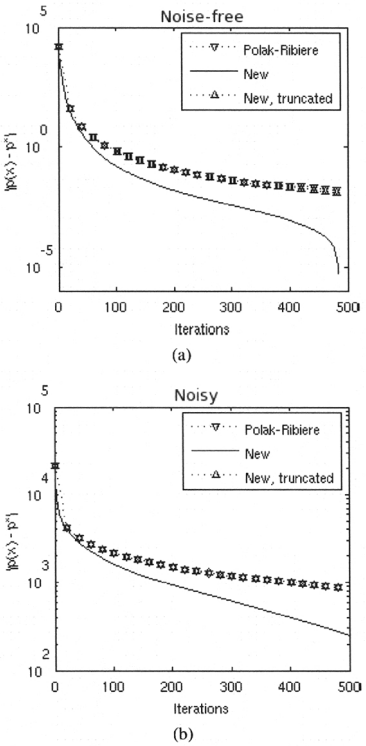

Fig. 2. Progress of reconstruction, measured by $|\log p_{\mathbf{m}} - \log p^*|$, where p^* is the log-likelihood optimal value found by running 10,000 iterations of Polak-Ribière ML-CG. The progress of our new formulation for ML-CG is shown with and without truncation of the search directions. Two datasets were reconstructed: (a) a noise-free dataset and (b) a dataset with Poisson noise based on 35 million counts. For the noise-free case, the log-likelihood of the new formulation exceeded the result of 10,000 iterations with Polak–Ribière after only 485 iterations.

Reconstructions were performed for both datasets (Fig. 3). We compared ML-CG with the Polak–Ribière formulation (6) against the new formulation (40). The truncated formulation (44) was also investigated. The log-likelihood residual, defined as $|\log p_{\mathbf{m}} - \log p^*|$ where $\log p^*$ is the objective at optimality, was the main figure of merit of this study. The value of the objective at optimality was computed by running 10,000 iterations of ML-CG with Polak–Ribière.

Fig. 1 shows typical values for $\beta_{n,j}$. We have observed that, independently of the iteration number, the last coefficient is always far greater than the others. Yet, accounting for all the search directions (and not only the last one) improves the convergence of ML-CG both for noise-free and noisy datasets. The log-likelihood residual converges to zero faster for the new formulation (Fig. 2). For the noisy dataset, it reaches the equivalent of 50 Polak–Ribière iterations in 39 iterations (1.3X

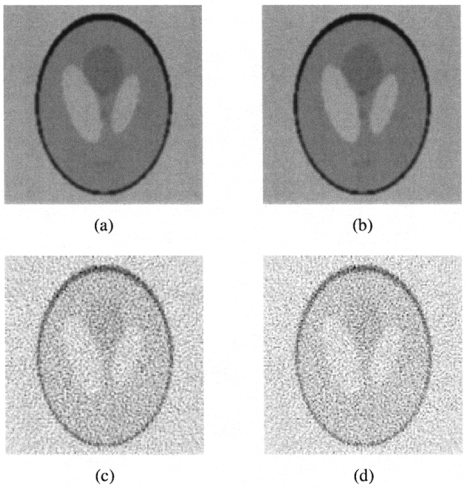

Fig. 3. The noise-free and noisy reconstructed images are shown for 500 iterations of Polak–Ribière ((a) and (c)) and 500 iterations of the new ML-specific formulation ((b) and (d)).

faster), and the equivalent of 2,000 Polak–Ribière iterations in 451 iterations (4.4X faster).

The new formulation can be truncated (44). The resulting formulation converges to the ML solution at the same rate as Polak-Ribière. Furthermore, both formulations yield similar values (average difference smaller than 0.3%) for β for all iterations (Fig. 1(b)).

The assumption (in Section II-B.2) that $\Lambda \approx I_P$ (or, equivalently, $A\mathbf{x}^*_{\mathrm{ml}} \approx \mathbf{m}$) was experimentally studied. Fig. 4 is a histogram of the diagonal coefficients of Λ, computed from the estimate of $\mathbf{x}^*_{\mathrm{ml}}$ obtained by running 10,000 iterations of Polak–Ribière ML-CG for the dataset with noise. The histogram is centered on 1, with a full-width half-maximum of 0.04.

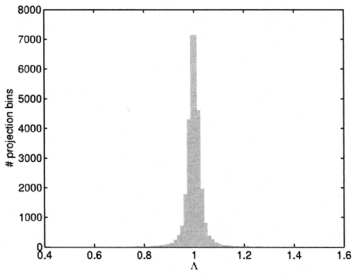

Fig. 4. Histogram of the diagonal coefficients of $\Lambda = \mathbf{diag}(\mathbf{m}/A\mathbf{x}^*_{\mathrm{ml}})$. In the derivation of the ML conjugation relationship, Λ is assumed to be the identity matrix.

IV. Discussion

For quadratic objectives, such as WLS, accounting for the last search search direction alone is sufficient to form a basis of conjugated search directions. For a non-quadratic objective, such as ML, all the past search directions must be combined when forming a new search direction to ensure conjugation. In the new, ML-specific formulation, the last search direction is typically more heavily weighted than older ones. However, accounting for all the search directions improves the convergence rate. If older search directions are truncated, the convergence rate reverts to that of standard Polak–Ribière.

The CG method based on the Polak–Ribière formulation is surprisingly efficient for the non-quadratic ML objective. Furthermore, it can be shown (by doing a full search on β) that no other formulation converges faster. The reason is that Polak–Ribière applied to the ML objective is equivalent to truncating the optimal ML formulation, which uses all the past search directions.

For the new method, only a small amount of additional computation is required to evaluate dot products. The total complexity remaining dominated by the cost of the back- and forward projections. The main drawback of the optimal ML formulation is that it requires that all the past search directions be stored in memory, which is impractical for large image volumes. Reducing the memory footprint while keeping the desirable properties of the new formulation is a goal worthwile pursuing.

The images produced by ML iterative reconstruction becomes increasingly noisy as the sequence of estimates reaches optimality. The noise can obscure the main features of the images. Thus, some researchers terminate the iterations early, before convergence is reached. This indeed limits the attractiveness of the reconstruction methods described in this paper. When an early ML iterate is sought, the method of choice is the ordered-subsets expectation-maximization (OSEM) algorithm. OSEM can produce images suitable for clinical use in as little as two iterations. OSEM, however, does not converge to the ML solution. To reduce noise, other researchers prefer either to apply smoothing filters to the ML estimate, or to incorporate a smoothing penalty in the objective. The CG algorithm and the novel methods described in this paper are suitable for these approaches.

The CG algorithm can be modified to incorporate a preconditioning matrix. Preconditioned CG (PCG), with the right choice of matrix, converges faster than regular CG to the optimal solution. Conventional preconditioners attempt to approximate the inverse of the Hessian matrix. In PET, the EM preconditioner [3] has been shown to improve convergence and provide a non-linear smoothing effect, similar to that obtained by running the EM algorithm. The optimal ML formulation for CG can be readily extended to include preconditioning.

The current problem formulation (27) does not incorporate a nonnegativity constraint for the voxels. The nonnegativity of the estimated projections is however implicitly enforced by the logarithmic term in the ML objective. Largely negative voxel values are therefore prevented, because they would induce negative projection values. In the absence of a nonnegativity constraint, voxels can take slightly negative values. It is very challenging to incorporate nonnegativity constraints in CG while preserving the fast convergence. In the early iterations, during which the cold voxels are identified, CG is not faster than a simple gradient ascent method. A suboptimal ML esti-

mate might be found by truncating the negative voxels at the end of the reconstruction (projection onto convex sets). Such truncation cannot be performed within the iterations because it would destroy the delicate sequence of conjugate directions. A logarithmic barrier can be used to enforce nonnegativity. Positive bias in cold regions might result, but the advantage of this method is that the conjugation of the search directions in CG is preserved. The positive bias can be mitigated by reducing the weight of the log barrier, but this negatively affects the conditioning of the objective.

V. Conclusion

We have investigated a new way to form search directions in ML-CG by enforcing explicit conjugation relationships derived from the expression of the log-likelihood in PET. This new ML conjugation relationship accounts for the non-quadraticity of the objective function. It thus requires that all the past search directions are used when forming a new conjugated search direction.

The new formulation converges faster to the ML objective: it takes 22% fewer iterations to reach the equivalent of 50 Polak–Ribière iterations, and 77% less iterations to reach the equivalent of 2,000 Polak–Ribière iterations. To reduce the memory burden, truncating all but the last search direction was investigated. In that case, the convergence rate reverts to that of the Polak–Ribière formulation. This result provides some insight on the performance of CG with the Polak–Ribière formulation. Polak–Ribière is approximately equivalent to truncating the optimal ML formulation, and for this reason performs relatively well despite the non-quadraticity of the ML criterion.

In future work, we will expand the theory to account for the incorporation of a regularizing prior and a voxel nonnegativity constraint. We will also research a theoretical proof of convergence. Last, we will investigate ways to reduce the memory burden, by for example limiting the number of past search directions that are stored or using efficient lossless compression.

References

[1] R. Fletcher and C. Reeves, "Function minimization by conjugate gradients," *Computer Journal*, vol. 7, pp. 149–154, 1964.

[2] G. Chinn and S.-C. Huang, "A general class of preconditioners for statistical iterative reconstruction of emission computed tomography," *IEEE Trans. Med. Imag.*, vol. 16, no. 1, pp. 1–10, Feb 1997.

[3] L. Kaufman, "Maximum likelihood, least squares, and penalized least squares for PET," *IEEE Trans. Med. Imag.*, vol. 12, no. 2, pp. 200–214, Jun 1993.

[4] E. Mumcuogluyz, R. Leahy, and S. Cherry, "Bayesian reconstruction of PET images: Methodology and performance analysis," *Phys. Med. Bio.*, vol. 41, pp. 1777–1807, 1996.

[5] J. R. Shewchuk, "An introduction to the conjugate gradient method without the agonizing pain," *unpublished paper*, Aug 1994.