Comparison of SIRT and SQS for Regularized Weighted Least Squares Image Reconstruction

Jens Gregor and
Dept. of Electrical Engr. & Computer Science, Univ. of Tennessee, Knoxville, TN 37996

Jeffrey A. Fessler
Dept. of Electrical Engr. & Computer Science, Univ. of Michigan, Ann Arbor, MI 48109

Jens Gregor: jgregor@utk.edu; Jeffrey A. Fessler: fessler@umich.edu

Abstract

Tomographic image reconstruction is often formulated as a regularized weighted least squares (RWLS) problem optimized by iterative algorithms that are either inherently algebraic or derived from a statistical point of view. This paper compares a modified version of SIRT (Simultaneous Iterative Reconstruction Technique), which is of the former type, with a version of SQS (Separable Quadratic Surrogates), which is of the latter type. We show that the two algorithms minimize the same criterion function using similar forms of preconditioned gradient descent. We present near-optimal relaxation for both based on eigenvalue bounds and include a heuristic extension for use with ordered subsets. We provide empirical evidence that SIRT and SQS converge at the same rate for all intents and purposes. For context, we compare their performance with an implementation of preconditioned conjugate gradient. The illustrative application is X-ray CT of luggage for aviation security.

I. Introduction

Iterative X-ray CT image reconstruction algorithms are often categorized as being either algebraic or statistical, with the former typically based on solving a system of equations centered around a forward model of the imaging process, and the latter based on maximizing a likelihood of the measurements. We are interested in the case where the two are related. For example, Sauer and Bouman [1] showed that a second-order Taylor-series expansion of a Poisson log-likelihood leads to a weighted least-squares (WLS) problem. Separable quadratic surrogate (SQS) methods for optimizing Poisson log-likelihoods also lead to WLS inner minimization problems [2], [3]. In these formulations the WLS weights have statistical meaning related to the modeled variance of the projection data. We show that statistical weighting is easily incorporated into algebraic algorithms such as SIRT (Simultaneous Iterative Reconstruction Technique) and SART (Simultaneous Algebraic Reconstruction Technique) [4].

The paper makes two main contributions. First, we establish similarity of a version of SIRT modified for WLS use with a version of the statistically based SQS algorithm. SIRT and

SQS were shown previously to be similar in their basic unregularized forms [5], [6]. Here we show that they can solve the same Tikhonov regularized WLS problem using the same gradient descent approach, only with different diagonal preconditioners. Second, while SIRT is often relaxed by means of a user-defined step size [7], [8], SQS was not developed with relaxation in mind. We present a practical approach for selecting near-optimal step sizes for both algorithms. We extend the relaxation to apply also for ordered subsets (OS) which have become the *de facto* method for reducing the computational cost of iterative reconstruction in practice [9]. As part of this work, a scaling factor is introduced that accounts for potential imbalance among the subsets.

The SIRT and SQS algorithms considered are those commonly encountered in the current literature. For other versions of SIRT including the original algorithm and generalizations thereof, see [10]–[13]. A broad class of more contemporary SIRT-like algorithms can be found in [14]. Generalized versions of SQS are discussed in [2], [3]. We make no claims that the results presented in this paper apply to these alternative algorithms due to differences in their matrix set-up relative to ours. Comparison with the recently introduced acceleration of SQS based on Nesterov's momentum [15], which likely produces faster convergence than what is shown here, is likewise considered out-of-scope as a comparable version does not exist for SIRT. For context, we provide an empirical comparison with preconditioned conjugate gradient (PCG) which often converges quickly on well-behaved, unconstrained WLS problems.

The illustrative application is X-ray CT of luggage for aviation security for which imaging challenges include beam hardening and metal artifacts. Neither is explicitly addressed by the weighting and the regularization considered here although the effects of both are likely alleviated somewhat. We use the luggage data because the presence of dense objects and the noisy character of the data exacerbates any differences with respect to data-dependent convergence behavior for the algorithms compared.

II. Notation and Problem Definition

Let $A = [a_{ij}]$ denote a non-negative $m \times n$ system matrix. Let $W = \text{diag}\{w_i\}$ denote an $m \times m$ diagonal statistical weighting matrix with positive diagonal entries w_i. That is, we assume that rows for which $w_i = 0$ have been removed from A and W with corresponding columns removed from W as well. Defined in greater detail below, let Q denote an $\eta \times n$ regularization matrix where typically $\eta \quad n$.

The general form of the regularized WLS (RWLS) image reconstruction problem addressed in this paper is

$$x^* = \arg\min_x \Psi(x) \quad \text{(1a)}$$

$$\Psi(x) \triangleq \frac{1}{2}\|Ax - y\|_W^2 + \beta\frac{1}{2}\|Qx\|_2^2, \quad \text{(1b)}$$

where $\|\boldsymbol{u}\|_{\boldsymbol{W}}^{2}$ is shorthand for $\boldsymbol{u}'\boldsymbol{W}\boldsymbol{u}$, and $\boldsymbol{x} = [x_j]$ and $\boldsymbol{y} = [y_i]$ denote $n \times 1$ and $m \times 1$ vectors representing the unknown image and the log-normalized projection data, respectively. User-defined hyperparameter β establishes a trade-off between the data term (left norm) and the regularizer term (right norm).

Matrix $\boldsymbol{Q}$ is usually chosen to stress structural characteristics of $\boldsymbol{x}$ that are undesirable. We focus on two common regularizes, namely, minimum norm for which $\|\boldsymbol{Q}\boldsymbol{x}\|_2^2 = \sum_j x_j^2$, and first-order finite differences for which $\|\boldsymbol{Q}\boldsymbol{x}\|_2^2 = \sum_j \sum_{k \in \mathcal{N}_j} (x_j - x_k)^2$, where $\mathcal{N}_j$ denotes the set of lexicographical predecessor neighbors. The former compensates for the linear system solved being ill-conditioned. The latter penalizes image roughness thereby implicitly encouraging smoothness. Adoption of other quadratic regularizers is trivial. We assume that $\boldsymbol{Q}$ and $\boldsymbol{A}$ have disjoint null spaces so that the cost function in (1) is strictly convex and has a unique minimizer [16].

Preconditioned gradient descent (PGD) solves RWLS problem (1) by means of the iterative update

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - \alpha \boldsymbol{D} \nabla \Psi(\boldsymbol{x}^{(k)}) \quad (2)$$

where relaxation parameter α defines the step size, $\boldsymbol{D}$ is a preconditioning matrix, and k denotes iteration. Algebraic expansion reveals that PGD-RWLS is a Richardson Iteration for solving a linear system of equations; cf. [17] for definition and mathematical properties. That is,

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \alpha \boldsymbol{D}(\boldsymbol{A}'\boldsymbol{W}\boldsymbol{y} - (\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A} + \beta \boldsymbol{Q}'\boldsymbol{Q})\boldsymbol{x}^{(k)}). \quad (3)$$

Convergence is guaranteed for an arbitrary choice of initial estimate $\boldsymbol{x}^{(0)}$ if

$$0 < \alpha < \frac{2}{\lambda_{\max}}, \quad (4)$$

with the fastest convergence rate obtained for

$$\alpha^* = \frac{2}{\lambda_{\max} + \lambda_{\min}}, \quad (5)$$

where $\lambda_{\max}$ and $\lambda_{\min}$ denote the largest and smallest eigenvalues of matrix $\boldsymbol{D}\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A} + \beta \boldsymbol{D}\boldsymbol{Q}'\boldsymbol{Q}$ whose eigenvalues are all assumed to be strictly positive.

The remainder of the paper is devoted to derivation, analysis, and comparison of Richardson Iterations for SIRT and SQS that solve RWLS problem (1) using near-optimal α values. Well-established properties for eigenvalues are used extensively without reference, cf. [18], [19]. The shorthand notation $\boldsymbol{M} = \boldsymbol{M}_1 + \beta \boldsymbol{M}_2$ where $\boldsymbol{M}_1 = \boldsymbol{D}\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}$ and $\boldsymbol{M}_2 = \boldsymbol{D}\boldsymbol{Q}'\boldsymbol{Q}$ will be

used for convenience. The smallest and largest eigenvalues of matrix M corresponding to the particular algorithm studied will be expressed as $\lambda_{\min}(M)$ and $\lambda_{\max}(M)$, with the explicit reference to M dropped when the matrix in question is obvious.

III. Method: SIRT-RWLS

A. Algorithm Derivation

Define the diagonal $m \times m$ and $n \times n$ matrices representing the inverse row and column sums of A, respectively, by $R = \text{diag}\{1/r_i\}$ and $C = \text{diag}\{1/c_j\}$, where $r_i = \Sigma_j \, a_{ij}$, and $c_j = \Sigma_i \, a_{ij}$. We assume that the column sums c_j of A are positive, so by design C is positive definite. In other words, every pixel being reconstructed is intersected by one or more rays[1].

The classical SIRT iteration for "solving" $Ax = y$ is

$$x^{(k+1)} = x^{(k)} + \alpha C A' R(y - A x^{(k)}). \quad (6)$$

This iteration is equivalent to gradient descent with diagonal preconditioner C of the "geometrically weighted" LS cost function $\frac{1}{2}\|Ax - y\|_R^2$. Weighting by R means that minimization of the residual errors is per unit length, allowing rays that intersect larger portions of the field of view to tolerate larger errors than rays that intersect smaller portions. However, statistical weighting is preferable to geometric weighting for noisy data.

Adding the gradient of the regularizer to (6) leads to a version of SIRT that solves the following minimization problem [8]:

$$x^* = \arg \min_x \frac{1}{2}\|Ax - y\|_R^2 + \beta \frac{1}{2}\|Qx\|_2^2. \quad (7)$$

Note use of an R-norm here. Next we modify SIRT using a simple variable transformation so that it solves RWLS problem (1) which is based on a W-norm. Let $u \triangleq Ax - y$ and $\tilde{u} \triangleq Mu$. Choosing $M \triangleq W R^{-1}$ implies that $\|u\|_W^2 = \|\tilde{u}\|_{\tilde{R}}^2$ where $\tilde{R} = M^{-1}R$. More specifically, let $\tilde{A} = [\tilde{a}_{ij}]$ and $\tilde{y} = [\tilde{y}_i]$, where $\tilde{a}_{ij} \triangleq w_i r_i a_{ij}$ and $\tilde{y}_i \triangleq w_i r_i y_i$. Let the diagonal matrices corresponding to the row and column sums of $\tilde{A}$ be $\tilde{R} \triangleq \text{diag}\{1/\tilde{r}_i\}$ and $\tilde{C} \triangleq \text{diag}\{1/\tilde{c}_j\}$ where $\tilde{r}_i \triangleq \sum_j \tilde{a}_{ij} = w_i r_i^2$ and $\tilde{c}_j \triangleq \Sigma_j \tilde{a}_{ij}$. The SIRT-like iteration for "solving" $\tilde{A}x = y$ with quadratic regularization is then obtained by left-multiplying the normal equations associated with the variable transformed version of (7) by $\tilde{C}$, followed by matrix splitting of the form $\tilde{C}\tilde{A}'\tilde{R}\tilde{A} = I - (I - \tilde{C}\tilde{A}'\tilde{R}\tilde{A})$. The resulting modified SIRT iteration is given by

$$x^{(k+1)} = (I - \alpha\beta\tilde{C}Q'Q)x^{(k)} + \alpha\tilde{C}\tilde{A}'\tilde{R}(\tilde{y} - \tilde{A}x^{(k)}). \quad (8)$$

---

[1]This assumption may not hold for certain helical CT geometries with padded end slices. Generalizing SIRT to such geometries could be interesting future work.

The choice of a is discussed in the next two subsections, first for the unregularized case and then when regularization is applied.

Using the identities $\tilde{A}'\mathcal{R}\tilde{A} = A'WA$ and $\tilde{A}'\mathcal{R}\tilde{y} = A'Wy$, SIRT is expressed in terms of the original problem variables as follows:

$$x^{(k+1)} = (I - \alpha\beta\tilde{C}Q'Q)x^{(k)} + \alpha\tilde{C}A'W(y - Ax^{(k)}) = x^{(k)} - \alpha\tilde{C}\nabla\Psi(x^{(k)}). \quad (9)$$

In other words, this modified SIRT algorithm is a Richardson Iteration that solves RWLS problem (1) using relaxed gradient descent with $\tilde{C}$ as a diagonal preconditioner. Note that

$$\tilde{c}_j = \sum_i \tilde{a}_{ij} = \sum_i a_{ij}w_i r_i = [A'WA\mathbf{1}]_j, \quad (10)$$

where $\mathbf{1}$ denotes a vector of n ones corresponding to the number of columns of A. Because $\tilde{C}$ merely serves as a preconditioner, it is amenable to substitution. Indeed, Section IV presents a SIRT-like derivation of the SQS algorithm for a different choice of $\tilde{C}$.

B. Step size for unregularized case

Assuming that A has full column rank and $w_i > 0$, then $\tilde{A}'\mathcal{R}\tilde{A} = A'WA$ is symmetric positive definite. It therefore follows that $C\tilde{A}'\mathcal{R}\tilde{A}$ has strictly positive eigenvalues because it is similar to the symmetric positive definite matrix $SA'WAS$ where $S = \tilde{C}^{1/2}$. Optimal step size $a*$ given by (5) thus applies for unregularized SIRT. Obtaining values for $\lambda_{\max}$ and $\lambda_{\min}$ with respect to $\tilde{C}A'WA$ is addressed next. (See [20]–[22] for convergence analyses.)

The spectral radius of a non-negative matrix is bounded by its smallest and largest row sums. Since $C\tilde{A}'$ and $\mathcal{R}\tilde{A}$ are both non-negative and stochastic, so is their product $C\tilde{A}'\mathcal{R}\tilde{A} = C\tilde{A}'WA$. The row sums of a stochastic matrix all equal 1. Therefore, $\lambda_{\max} = 1$, which implies $1 \quad a* \triangleq 2/(1 + \lambda_{\min}) < 2$. Fig. 1 illustrates this relationship. Empirical comparisons of residual norms for typical 2D and 3D tomography problems have consistently shown that using $a = 1.99$ requires half as many iterations to achieve the same residual error as $a = 1.00$ [7]. This behavior indicates $\lambda_{\min} \ll 1$. While the value of $\lambda_{\min}$ is difficult to establish, we can find a conservative bound to use in its place. The trace of an $n \times n$ matrix equals the sum of its eigenvalues. The smallest eigenvalue can be no more than the average of all eigenvalues. Hence,

$$\lambda_{\min} \leq \frac{\mathrm{tr}(\tilde{C}A'WA)}{n}. \quad (11)$$

C. Step size for regularized case

We now examine selection of optimal step size $a*$ for the regularized case. We use the fact that $C(\tilde{A}'WA + \beta Q'Q)$ is similar to $S(A'WA + \beta Q'Q)S$ where S remains defined as before. To simplify the notation, let $M = M_1 + \beta M_2$ where $M_1 = SA'WAS$ and $M_2 = SQ'QS$. Note that M_1 and M_2 are both Hermitian matrices.

The sum of two positive semi-definite matrices is nonetheless positive definite, if they have different null spaces. This is the case for M_1 and M_2. Optimal step size α^* thus remains given by (5) when regularization is applied. We must therefore determine the values of $\lambda_{\max}(M)$ and $\lambda_{\min}(M)$. As previously shown [8], this is easily done for minimum norm regularization where $Q = I$. However, for finite difference regularization where $Q \neq I$, this is challenging. Instead, we derive bounds that lead to guaranteed convergence while keeping optimality in mind.

Applying Weyl's inequalities for the sum of Hermitian matrices:

$$\lambda_{\max}(M_1) + \beta \lambda_{\min}(M_2) \leq \lambda_{\max}(M) \quad (12a)$$

$$\leq \lambda_{\max}(M_1) + \beta \lambda_{\max}(M_2) \quad (12b)$$

and

$$\lambda_{\min}(M_1) + \beta \lambda_{\min}(M_2) \leq \lambda_{\min}(M) \quad (13a)$$

$$\leq \lambda_{\min}(M_1) + \beta \lambda_{\max}(M_2). \quad (13b)$$

SIRT converges to the RWLS minimizer x^* if (4) holds. Using the upper bound on $\lambda_{\max}(M)$ given by (12) as well as the upper bound on $\lambda_{\min}(M)$ given by (13) would be valid but might be overly conservative. As an alternative, we combine the upper bound on the former with the lower bound on the latter to obtain a step size estimate that satisfies (4) and thus ensures convergence. That is,

$$\tilde{\alpha}^* \triangleq \frac{2}{\lambda_{\max}(M_1) + \lambda_{\min}(M_1) + \beta(\lambda_{\max}(M_2) + \lambda_{\min}(M_2))}. \quad (14)$$

We analyze the similarity of $\tilde{\alpha}^*$ to α^* next by considering two extreme cases.

Case 1: Suppose $\lambda_{\max}(M)$ and $\lambda_{\min}(M)$ both take on their lower bounds. Then $\tilde{\alpha}^*$ underestimates α^* by the factor

$$\frac{\tilde{\alpha}^*}{\alpha^*} = \frac{\lambda_{\max}(M_1) + \lambda_{\min}(M_1) + 2\beta \lambda_{\min}(M_2)}{\lambda_{\max}(M_1) + \lambda_{\min}(M_1) + \beta(\lambda_{\max}(M_2) + \lambda_{\min}(M_2))}.$$

We see that near-optimality is achieved when either of two conditions are met, namely, $\lambda_{\min}(M_2) \approx \lambda_{\max}(M_2)$ or $\beta \lambda_{\max}(M_2) \ll \lambda_{\max}(M_1)$. Sub-optimality results to a greater or lesser degree when neither of these conditions hold.

Case 2: Conversely, suppose $\lambda_{\max}(M)$ and $\lambda_{\min}(M)$ both take on their upper bounds. Then $\tilde{\alpha}^*$ overestimates α^* by the factor

$$\frac{\tilde{\alpha}^*}{\alpha^*} = \frac{\lambda_{\max}(\boldsymbol{M}_1) + \lambda_{\min}(\boldsymbol{M}_1) + 2\beta\lambda_{\max}(\boldsymbol{M}_2)}{\lambda_{\max}(\boldsymbol{M}_1) + \lambda_{\min}(\boldsymbol{M}_1) + \beta(\lambda_{\max}(\boldsymbol{M}_2) + \lambda_{\min}(\boldsymbol{M}_2))}.$$

Note that although $\tilde{\alpha}^*$ may exceed α^*, it is still a valid step size for SIRT. Near-optimality is again achieved when either of two conditions are met, this time $\lambda_{\min}(\boldsymbol{M}_2) \approx \lambda_{\max}(\boldsymbol{M}_2)$ or $2\beta\lambda_{\max}(\boldsymbol{M}_2) \ll \lambda_{\max}(\boldsymbol{M}_1)$. Sub-optimality results to a greater or lesser degree when neither of these conditions hold.

Restating (14) in terms of the original problem variables we get

$$\tilde{\alpha}^* = \frac{2}{\Lambda(\tilde{\boldsymbol{C}}\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}) + \beta\Lambda(\tilde{\boldsymbol{C}}\boldsymbol{Q}'\boldsymbol{Q})} \quad (15)$$

Where

$$\Lambda(\tilde{\boldsymbol{C}}\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}) \triangleq 1 + \lambda_{\min}(\tilde{\boldsymbol{C}}\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A})$$
$$\Lambda(\tilde{\boldsymbol{C}}\boldsymbol{Q}'\boldsymbol{Q}) \triangleq \lambda_{\max}(\tilde{\boldsymbol{C}}\boldsymbol{Q}'\boldsymbol{Q}) + \lambda_{\min}(\tilde{\boldsymbol{C}}\boldsymbol{Q}'\boldsymbol{Q}).$$

Whether applied to 2D or 3D reconstruction, minimum norm regularization yields

$$\lambda_{\max}(\tilde{\boldsymbol{C}}\boldsymbol{Q}'\boldsymbol{Q}) = \lambda_{\max}(\tilde{\boldsymbol{C}}) = 1/\min \tilde{c}_j \quad (16a)$$

$$\lambda_{\min}(\tilde{\boldsymbol{C}}\boldsymbol{Q}'\boldsymbol{Q}) = \lambda_{\min}(\tilde{\boldsymbol{C}}) = 1/\max \tilde{c}_j. \quad (16b)$$

For finite difference regularization, we first note that a multiplicative equivalent to Weyl's inequalities follows from the relation between (maximum) eigenvalues and matrix norms, namely,

$$\lambda_{\max}(\tilde{\boldsymbol{C}}\boldsymbol{Q}'\boldsymbol{Q}) \leq \lambda_{\max}(\tilde{\boldsymbol{C}})\lambda_{\max}(\boldsymbol{Q}'\boldsymbol{Q}) \leq \lambda_{\max}(\tilde{\boldsymbol{C}})\|\boldsymbol{Q}'\boldsymbol{Q}\|_1,$$

where $\|\boldsymbol{Q}'\boldsymbol{Q}\|_1$ denotes the maximum absolute column sum. In 2D, for a first-order neighborhood a typical row of $\boldsymbol{Q}'\boldsymbol{Q}$ has one entry valued 4 and four entries valued −1, leading to $\|\boldsymbol{Q}'\boldsymbol{Q}\|_1 = 8$, In 3D, one entry is valued 6 and six entries are valued −1, leading to $\|\boldsymbol{Q}'\boldsymbol{Q}\|_1 = 12$. Extending to other neighborhoods is trivial. Combined with singularity of $\boldsymbol{Q}'\boldsymbol{Q}$ carrying through to $\boldsymbol{C}\tilde{\boldsymbol{Q}}'\boldsymbol{Q}$, we thus have that

$$\lambda_{\max}(\tilde{\boldsymbol{C}}\boldsymbol{Q}'\boldsymbol{Q}) \leq \frac{4N}{\min_j \tilde{c}_j} \quad (17a)$$

$$\lambda_{\min}(\tilde{\boldsymbol{C}}\boldsymbol{Q}'\boldsymbol{Q})=0 \quad \text{(17b)}$$

where N is 2 or 3 corresponding to the problem dimensionality.

To summarize, RWLS image reconstruction by means of SIRT as given by (9) is carried out by applying weighting to the system matrix and the log-normalized data. Near-optimal step size a^* is obtained by substituting (16) or (17) into (15) and using (11). That is,

$$\tilde{\alpha}^*=\frac{2}{1+\frac{\text{tr}(\tilde{\boldsymbol{C}}\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A})}{n}+\beta\left(\frac{v_1}{\min_j \tilde{c}_j}+\frac{v_2}{\max_j \tilde{c}_j}\right)}, \quad \text{(18)}$$

where $v_1 = v_2 = 1$ for minimum-norm regularization and $v_1 = 4N$ and $v_2 = 0$ for N-dimensional finite difference regularization. Computing the trace term requires the equivalent effort of one combined forward and back-projection.

IV. Method: SQS-RWLS

A. Algorithm Derivation

Although SQS was presented originally for non-quadratic log-likelihoods like the Poisson model [2], it is equally applicable to RWLS problems like (1) [15], [23], where $\boldsymbol{W}$ is the reciprocal of the (modeled) variance of y_i. The SQS approach to solving such problems traditionally has been derived using surrogate functions [2], [23], based on the work of De Pierro [24], [25]. Here we present a different "SIRT-like" derivation for the RWLS case that leads to near-optimal relaxation.

The idea is to first specify an appropriate system of equations of the form $\boldsymbol{Bx} = \boldsymbol{d}$ such that applying the classical SIRT algorithm (6) to that system solves RWLS problem (1) and then add relaxation. As noted earlier, classical SIRT solves a geometrically weighted LS problem rather than the statistical WLS problem, so it requires a judicious choice of $\boldsymbol{B}$ and $\boldsymbol{d}$ to coerce SIRT into working as desired here.

We first use variable transforms akin to those used in (8) to rewrite RWLS cost function (1) in a stacked form, namely,

$$\Psi(\boldsymbol{x})=\frac{1}{2}\|\boldsymbol{Bx} - \boldsymbol{d}\|_{\hat{R}}^2, \quad \text{(19)}$$

Where

$$\boldsymbol{B} \triangleq \left[\begin{array}{c} \tilde{\boldsymbol{A}} \\ \beta\hat{\boldsymbol{Q}} \end{array}\right], \boldsymbol{d} \triangleq \left[\begin{array}{c} \tilde{\boldsymbol{y}} \\ \boldsymbol{0} \end{array}\right], \hat{\boldsymbol{R}} \triangleq \left[\begin{array}{cc} \tilde{\boldsymbol{R}} & \boldsymbol{0} \\ \boldsymbol{0} & (1/\beta)\boldsymbol{P}^2 \end{array}\right]$$

and $Q \stackrel{\triangle}{=} P^{-1}Q$, $P \triangleq \text{diag}\ \{1/p_k\}$, $p_k \triangleq \Sigma_j\ |q_{kj}|$, and $\mathbf{0}$ denotes, as appropriate, a vector of zeros whose length is the number of rows of Q or matrices of zeros whose dimensions correspond to the number of rows and columns of $\tilde{R}$ and P.

The generalization of SIRT to extend beyond non-negative matrices by using such absolute sums has been published previously [20], [26]. That generalization is needed here because Q has negative elements for finite difference regularizers. It is easily shown that $\hat{R} = \text{diag}$ $\{1/r_i\}$ where the absolute row sums of B are $r_i \stackrel{\triangle}{=} \Sigma_j\ |b_{ij}|$. Denoting the absolute column sums via $\hat{C} = \text{diag}\ \{1/\hat{c}_j\}$ where $\hat{c}_j = \Sigma_i|b_{ij}|$, the classic SIRT method for "solving" $Bx = d$ is given by

$$x^{(k+1)} = x^{(k)} + \hat{C}B^{'}\hat{R}(d - Bx^{(k)}).$$

Using the identities $B'R\hat{B} = A'WA + \beta Q'Q$ and $B'R\hat{d} = A'Wy$, we can express this iteration using the original problem variables while also introducing relaxation:

$$x^{(k+1)} = (I - \alpha\beta\hat{C}Q^{'}Q)x^{(k)} + \alpha\hat{C}A^{'}W(y - Ax^{(k)}) = x^{(k)} - \alpha\hat{C}\nabla\Psi(x^{(k)}). \quad (20)$$

Remarkably, iteration (20) with $\alpha = 1$ is exactly the usual SQS iteration (*cf.* [23, eqn. (17)]) for minimizing RWLS problem (1). The convergence proof for SQS is normally based on the fact it decreases $\Psi(x^{(k)})$ monotonically [27]. One can also see that convergence follows from the fact that $\lambda_{\max}(\hat{C}B'R\hat{B}) = \lambda_{\max}(\hat{C}(A'WA + \beta Q'Q)) \quad 1$ by arguments similar to those in [7, eqn.(6)].

We see that the SQS version of SIRT in (20) is also preconditioned gradient descent. The "denominator" elements of preconditioner $\hat{C}$ are given by

$$\hat{c}_j = \sum_i |b_{ij}| = [A^{'}WA\mathbf{1}]_j + \beta\sum_k |q_{kj}|\ p_k. \quad (21)$$

For 2D and 3D minimum norm regularization, use of $Q = I$ implies that $p_k = 1$ and $\Sigma_k|q_{kj}|\ p_k = 1$ leading to $\hat{c}_j = \tilde{c}_j + \beta$. For first-order finite difference regularization, each row of Q has at most one entry valued 1 and one entry valued −1 such that $p_k \quad 2$. In 2D, each column of Q has at most two entries valued 1 and two entries valued −1 while in 3D there are at most three of each. Thus, $\Sigma_k|q_{kj}|\ p_k \quad 4N$ leading to $\hat{c}_j \quad \tilde{c}_j + 4N\beta$ where N again refers to the problem dimensionality. For the experimental results presented in Section VI, we used a simple implementation based on $\hat{c}_j = \tilde{c}_j + 4N\beta$. This choice is exact for periodic boundary conditions and is a value upper bound otherwise.

To summarize, SIRT-RWLS iteration (9) and SQS-RWLS iteration (20) are identical for solving RWLS problem (1), except that SIRT uses $\tilde{C}$ and SQS uses $\hat{C}$ to precondition the relaxed gradient descent.

B. Step size for unregularized case

Without regularization, SIRT-WLS and SQS-WLS are identical. The latter can thus be relaxed using the same method developed for the former in Section III which typically results in $\alpha^* = 1.99$.

C. Step size for regularized case

When regularization is applied, the optimal step size is given by

$$\hat{\alpha}^* = \frac{2}{\Lambda(\hat{C}A'WA) + \beta\Lambda(\hat{C}Q'Q)} \quad (22)$$

Where

$$\Lambda(\hat{C}A'WA) \triangleq \lambda_{\max}(\hat{C}A'WA) + \lambda_{\min}(\hat{C}A'WA)\Lambda(\hat{C}Q'Q) \triangleq \lambda_{\max}(\hat{C}Q'Q) + \lambda_{\min}(\hat{C}Q'Q).$$

We develop the needed eigenvalue bounds next.

With respect to $\Lambda(\hat{C}A'WA)$, we apply the transformation $\hat{C} = \hat{S}\tilde{C}$ where $\hat{S} \triangleq \mathrm{diag}\{\hat{s}_j\}$ and $\hat{s}_j \triangleq c_j/\tilde{c}_j$ to write $\lambda_{\max}(\hat{C}A'WA) = \lambda_{\max}(\hat{S}\tilde{C}A'WA) \quad \lambda_{\max}(\hat{S})$. We thus have that,

$$\Lambda(\hat{C}A'WA) \leq \hat{s}_{\max} + \lambda_{\min}(\hat{C}A'WA)$$

where $\hat{s}_{\max} \triangleq \max_j \hat{s}_j$ and

$$\lambda_{\min}(\hat{C}A'WA) \leq \frac{\mathrm{tr}(\hat{C}A'WA)}{n}.$$

The bounds for $\Lambda(\hat{C}Q'Q)$ are established using derivations from the previous section. For minimum norm regularization,

$$\lambda_{\max}(\hat{C}Q'Q) = \lambda_{\max}(\hat{C}) = 1/\min \hat{C}_j \quad (23a)$$

$$\lambda_{\min}(\hat{C}Q'Q) = \lambda_{\min}(\hat{C}) = 1/\max \hat{C}_j \quad (23b)$$

while for finite difference regularization,

$$\lambda_{\max}(\hat{C}Q'Q) \leq \frac{4N}{\min_j \hat{c}_j} \quad (24a)$$

$$\lambda_{\min}(\hat{\boldsymbol{C}}\boldsymbol{Q}'\boldsymbol{Q})=0. \quad \text{(24b)}$$

Combining the above, we obtain the following near-optimal step size α^* expression:

$$\hat{\alpha}^* = \frac{2}{\hat{s}_{\max} + \frac{\text{tr}(\hat{\boldsymbol{C}}\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A})}{n} + \beta\left(\frac{w_1}{\min_j \hat{c}_j} + \frac{w_2}{\max_j \hat{c}_j}\right)}, \quad \text{(25)}$$

where $w_1 = w_2 = 1$ for minimum-norm regularization and $w_1 = 4N$ and $w_2 = 0$ for N-dimensional finite difference regularization.

The similarity of step size expressions (18) for SIRT-RWLS and (25) for SQS-RWLS is striking. Which of the two algorithms has the fastest convergence rate depends on the relation between $\tilde{\alpha}^*\tilde{\boldsymbol{C}}$ and $\hat{\alpha}^*\hat{\boldsymbol{C}}$ since these terms constitute the only difference between them. Note that $\tilde{\alpha}^*\tilde{\boldsymbol{C}} \approx \hat{\alpha}^*\hat{\boldsymbol{C}}$ when $\tilde{c}_j \gg w_1\beta$ since $\hat{c}_j = \tilde{c}_j + w_1\beta$. We expect this mild condition to be met for most applications.

V. Ordered subsets

Typically SIRT and SQS establish the low-frequency components of an image faster than the high-frequency components. Ordered subsets (OS) is a well-known, albeit heuristic technique for accelerating convergence in early iterations [9]. The idea is to partition the projection data and the corresponding system matrix rows and successively perform updates using these subsets. To simplify the development of OS-SIRT below, we assume partitioning is done in such a way that each subset includes a view of all pixels[2].

Although quite distinct from one another when originally presented, the terms SIRT [10] and SART [28] are now used interchangeably in the literature. SIRT updates the image using all projections whereas SART uses one projection at a time. In that sense, SART can be viewed as an OS version of SIRT where each subset consists of a single projection. Variants thereof where multiple projections are grouped to form subsets have been studied for both algorithms, e.g., [29], [30].

In this paper, we use the ordered subsets approach taken for SQS [23]. Each full iteration uses the gradients of M partial cost functions of the form

$$\Psi_m(\boldsymbol{x}) \triangleq \frac{1}{2}\|\boldsymbol{A}_m\boldsymbol{x} - \boldsymbol{y}_m\|^2_{\mathbf{W}_m} + \beta\frac{1}{2M}\|\boldsymbol{Q}\boldsymbol{x}\|^2_2$$

where the m-subscripts indicate that only data for the mth subset is used. Letting $k = nM + m$ denote the mth sub-iteration of the nth full-iteration, the relaxed, preconditioned (incremental) gradient descent algorithms considered can be written as

---

[2]This assumption may not always hold for OS with helical CT.

$$x^{\left(\frac{k+1}{M}\right)} = x^{\left(\frac{k}{M}\right)} - \alpha DM \nabla \Psi_m(x^{\left(\frac{k}{M}\right)}) \quad (26)$$

where OS-SIRT and OS-SQS correspond to $\alpha D = \alpha^* \tilde{C}$ and $\alpha D = \alpha^* \hat{C}$, respectively. The preconditioners are not altered as a result of the subsets.

In practice, the subset balance condition $\nabla \Psi(x) \approx M \nabla \Psi_m(x)$ will hold only approximately and the iterates approach a limit cycle rather than converging to a minimizer. When the subset imbalance is mild but non-negligible, OS-SIRT and OS-SQS may exhibit sub-optimal convergence rates due to step size estimates α^* and $\hat{\alpha}^*$ being too aggressive. To compensate, we introduce a scaling factor that reduces the step size based on a measure of the imbalance. Specifically, we quantify and use the degree to which the term $M\Lambda(DA_m' W_m A_m)$ associated with (26) deviates from $\Lambda(DA' W A)$ which plays a central role in the step size computations.

Consider (temporarily) computing subset specific step size estimates based on matrices for the individual subsets. That is,

$$\alpha_m^* = \frac{2}{M\Lambda(DA_m' W_m A_m) + \beta \Lambda(DQ' Q)} \quad (27)$$

where $\Lambda(DA_m' W_m A_m)$ is the only unknown entity.

In a manner similar to when developing relaxation for SQS-RWLS, let $D = S_m C_m$ where $S_m \triangleq \mathrm{diag}\left\{s_j^m\right\}$ and $s_j^m \triangleq c_j^m / d_j$. Furthermore, let $C_m \triangleq \mathrm{diag}\left\{c_j^m\right\}$ where c_j^m denotes column sums of A_m akin to (10). Then

$\lambda_{\max}(DA_m' W_m A_m) = \lambda_{\max}(S_m C_m A_m' W_m A_m) \leq s_{\max}^m$ where $s_{\max}^m \triangleq \max_j s_j^m$. Also, $\lambda_{\min}(DA_m' W_m A_m) \leq s_{\max}^m \lambda_{\min}(C_m A_m' W_m A_m)$. Rather than introducing a trace based bound for the right hand side eigenvalue, we use the approximation $\lambda_{\min}(C_m A_m' W_m A_m) \approx \lambda_{\min}(CA' W A)$. These considerations, along with the definition of $S_M \triangleq \max_m s_{\max}^m M$ to eliminate the subset dependency, lead us to define the following generic OS-PGD step size estimate

$$\alpha_{\mathrm{os}}^* = \frac{2}{S_M \Lambda(DA' W A) + \beta \Lambda(DQ' Q)}. \quad (28)$$

For OS-SIRT, replacing D by $\tilde{C}$ results in the step size estimate

$$\tilde{\alpha}_{\mathrm{os}}^* = \frac{2}{\tilde{S}_M \Lambda(\tilde{C}A' W A) + \beta \Lambda(\tilde{C}Q' Q)}. \quad (29)$$

where $\tilde{S}_M \triangleq \max_m \max_j (\tilde{c}_j^m / \tilde{c}_j) M$.

For OS-SQS, replacing D by $\hat{C}$ results in the step size estimate

$$\hat{\alpha}_{\text{os}}^* = \frac{2}{\hat{S}_M \Lambda(\hat{C}A'WA) + \beta\Lambda(\hat{C}Q'Q)}. \quad (30)$$

where $\hat{S}_M \triangleq \max_m \max_j (\hat{c}_j^m / \hat{c}_j) M$.

When the subsets are well-balanced, $S_M \approx 1$ and the step size is comparable to the one developed before considering ordered subsets. When the subsets are mildly imbalanced, the use of S_M results in a smaller value for the step size. For large enough M, the subsets will be imbalanced to the point where neither OS-SIRT nor OS-SQS will come sufficiently close to x^*. The only viable remedy for this case is to use fewer subsets.

We close this section by reminding the reader that ordered subsets do not guarantee convergence, unless one gradually decreases the relaxation parameter [31]. The proposed rescaling of the step sizes for OS-SIRT and OS-SQS is equally heuristic. We neither claim that the rescaling results in convergent algorithms nor do we claim that convergence will not occur in practice without it. The most important contribution of the above developments is perhaps scaling factor S_M and its use for indicating subset imbalance.

VI. Experimental Results

A. Comparing SIRT and SQS

We implemented SIRT-RWLS and SQS-RWLS as described above and empirically compared their convergence rates. The data consisted of luggage scans obtained as part of a U.S. Department of Homeland Security sponsored project managed by ALERT at Northeastern University (Boston, MA) [32]. Several suitcase-like containers were scanned using an Imatron C300 fifth-generation electron-beam X-ray CT scanner. We used two of these data sets denoted DS1 and DS2. After data rebinning, the system was modeled as a third-generation equiangular fan-beam geometry having a circular source trajectory covering 216 degrees over 864 view angles that were 0.25 degrees apart and 864 detectors spanning a range of 41.3 degrees corresponding to a pitch of approximately 0.048 degrees. The source-to-isocenter distance was 675 mm while the isocenter-to-detector distance was 900 mm yielding $2.33 \times$ magnification. Polynomial beam hardening correction was applied. We reconstructed 512×512 images using isotropic 0.928 mm wide pixels. Pixels outside a circular support region having radius 475 mm were excluded from consideration. Modeled on area intersection the resulting $746,496 \times 204,836$ system matrix A contained approximately one billion non-zero elements. Weight matrix W was formed by partly reversing the log-normalization of the projection data, i.e., $w_i = \exp(-y_i)$. The beam intensity was unknown and assumed constant which may be inaccurate.

SIRT and SQS were both initialized using filtered back-projection for $x^{(0)}$. We ran 256 iterations of each algorithm for each experiment. Comparisons were based on values of $\Psi(x^{(k)})$. We performed OS reconstructions for $M = 1, 2, 4, 8, 16,$ and 32. Non-negativity was imposed on the image after each subiteration using gradient projection. Reconstructions were computed for a wide range of β values. We report results for $\beta = 1.0$ as that value produced images that were clearly regularized but not overly so.

Figures 2 and 3 show $\beta = 1.0$ and $\beta = 5.0$ reconstructions respectively for DS1 using SIRT and minimum norm regularization and DS2 using SQS and finite difference regularization. The intensity values represent modified Hounsfield units (MHU) for which air has a value of 0 while water has a value of 1000. The images were cropped to 350×430 to save space. DS1 is corrupted by a metal streak artifact that cuts into the water container. DS2 suffers even more from both metal streak and metal shading artifacts. These types of corruptions are not uncommon for luggage data although the particular data used here was purposely chosen to be challenging. More importantly, *the images produced by SIRT and SQS are indistinguishable from one another*. For the minimum norm regularized images of DS1, the numerical difference is within ±0.01 MHU. For the finite difference regularized images of DS2, the numerical difference is within ±0.10 MHU. Similar results were obtained for DS1 using finite difference regularization and DS2 using minimum norm regularization.

Table I lists minimum and maximum column sum values for $C \tilde{\ }$ as used by SIRT to compute LS and WLS solutions. Data weighting is seen to lower especially the minimum values. With the maximum values remaining large, $\tilde{a}^* \approx \hat{a}^*$. Due to the minimum values being lowered quite substantially, $\hat{C} \tilde{C}$ meaning some pixels will be updated more slowly by SQS than by SIRT. The degree to which that creates a difference in convergence rates is studied below.

Table II provides eigenvalue bounds for the data and regularizer terms associated with SIRT reconstruction based on minimum norm (MN) and finite difference (FD) regularization for SIRT. Table III provides the similar numbers for SQS. The largest eigenvalue of the data term equals 1 for SIRT by design and approximately does so for SQS because $\hat{s}_{\max} = \max c_j/\hat{c}_j \approx 1$. We can therefore infer that the conservative upper bound used for the smallest eigenvalue of the data term in both cases is at least three orders of magnitude smaller than the largest eigenvalue indicating SIRT and SQS are tasked with solving very poorly conditioned problems. We also see that the eigenvalue bounds for the regularizer terms are substantially smaller than those for the data terms.

The above observations indicate that near-optimal step size values, i.e., $a \approx 2/(1 + \varepsilon)$, should be achievable. Tables IV and V list the actual step sizes computed for OS-SIRT and OS-SQS along with the $\tilde{S}_M$ and $\hat{S}_M$ scaling factors. For $M = 1$, near-optimal step sizes were indeed produced for both SIRT and SQS. Increasingly smaller step sizes resulted for larger values of M due to data dependent subset imbalances becoming more significant. SIRT and SQS behaved the same in this regard.

We computed step size estimates for non-regularized WLS reconstructions and found them to be comparable to those listed in Tables IV and V. Regularization thus had negligible impact on $\tilde{a}^*$ and $\hat{a}^*$. We also computed step size estimates for pure LS reconstructions. These were all closer to 2 than the RWLS estimates but did decrease for larger value of M. This does not indicate a problem with the system model. Rather, it is related to physical characteristics of the system geometry, such as the rays being much wider at one end than at the other, causing pixel coverage to be both location and view dependent. Combined with a shortscan source trajectory, the larger angular strides associated with larger values of M

invariably induce subset column sum differences. This in turn results in a larger value of the S_M scaling factor which causes the step size estimate to decrease.

While all OS-SIRT and OS-SQS configurations were found to be convergent, $\Psi(\boldsymbol{x}^{(k)})$ values for $M = 16$ and 32 were larger than those for $M = 8$ indicating convergence to a limit cycle. We suspect the reason for this behavior is related to the aforementioned data variations. The statistical weighting considered in this paper is not intended to handle such systematic data inconsistencies.

Figures 4 and 5 show $\Psi(\boldsymbol{x}^{(k)})$ for OS-SIRT applied to DS1 and OS-SQS applied to DS2, respectively, using $M = 1$ and $M = 8$ as well $a = 1.00$ and $a = a^*$. The axes are scaled to best show configuration differences. This includes the x-axes being linear and the y-axes logarithmic. The plots for $M \quad 8$ show similar convergence behavior, while the plots for M 16 confirm the previously mentioned fast convergence to a sub-optimal Ψ value. Data set DS1 is noisy but cleaner and thus likely more internally consistent than DS2. This is reflected in the final Ψ value for DS1 being substantially lower than that of DS2.

Figures 6 and 7 provide scatter plots of OS-SIRT versus OS-SQS for the combinations of M and a mentioned above. A correlation coefficient of 1.00 was computed for all plots indicating that applying OS-SQS to DS1 and OS-SIRT to DS2 would produce identical plots to those shown in Figs. 4 and 5.

There is in other words, no difference between the convergence rates of SIRT and SQS for the data tested—To more clearly illustrate the speed-up produced by relaxation, Figs. 8 and 9 plot the ratio of the interpolated number of iterations of the slower curve relative to the number of iterations needed to reach the same value of the faster curve. The two top plots show that the relaxed versions of OS-SIRT and OS-SQS decrease Ψ close to $\tilde{a^*}$ and $\hat{a^*}$ times faster than when using $a = 1.00$ for the first 32 iterations for which speed-up factors were computed. This speed-up clearly indicates that the proposed step size estimates for SIRT, namely, (18) and (29), and SQS, namely, (25) and (30), are very effective. The two bottom plots reconfirm that subset imbalances cause both OS-SIRT and OS-SQS to quickly approach a near-final value of Ψ whereafter only slow progress is made in reducing it. Ideally, the speed-up would be constant at 8 corresponding to the number of subsets used.

B. Comparison with Conjugate Gradient

SIRT-RWLS and SQS-RWLS are stationary iterative methods that repeatedly apply a gradient based "correction" to the current estimate of the solution. As shown above, convergence is governed by the largest and smallest eigenvalues of the underlying iteration matrix. Conjugate gradient (CG), on the other hand, is a Krylov subspace method [17] that iteratively forms an increasingly larger basis for which an approximation to the solution is found that minimizes the cost function in that subspace. Convergence depends on the full eigenvalue spectrum and is proportional to the square root of the number of iterations needed by SIRT [14]. Preconditioning (PCG) may accelerate the convergence rate further.

To compare the behavior of the proposed relaxed versions of SIRT and SQS with PCG when applied to RWLS problem (1), we implemented the Polak-Ribiére CG algorithm along with a Jacobi preconditioner [33], [34]. Enforcing non-negativity on the solution was not considered as such a constraint is difficult to efficiently incorporate into a CG algorithm due to the need for conjugacy preservation among the basis vectors, c.f. [35]. All algorithms were regularized using $\beta = 1.0$. OS-SIRT and OS-SQS were executed for $M = 8$.

Figure 10 shows PCG-RWLS reconstructions for DS1 using minimum norm regularization and DS2 using finite difference regularization. The corresponding SIRT-RWLS and SQS-RWLS reconstructions look similar to those in Figs. 2 and 3. The difference images reveal that high density objects and edges exhibit the greatest degree of disparity. As indicated by Figs. 11 and 12, cost function $\Psi(x^{(k)})$ has converged to a lower value for PCG-RWLS than for SIRT-RWLS and SQS-RWLS. While the images of the former kind thus are "more optimal" than those of the latter, one could argue that they are qualitatively less desirable due to their noisier appearance and the Gibbs-like ringing which can be seen to accentuate object edges as well as the metal streak artifacts.

Figures 11 and 12 show that OS-SIRT and OS-SQS converge faster than both CG and PCG during the early iterations, only to be overtaken when limit cycle behavior sets in. CG is seen to converge more slowly than PCG, but eventually reaches the same limit value due to data inconsistencies causing PCG to stall out. The cost function limit values are lower than those in Figs. 4 and 5 due to non-negativity not being enforced.

For reconstructions based on 256 iterations, a speed-up factor of 16 should be expected. After the first 32 iterations when limit cycle behavior has not yet become an issue, OS-SIRT and OS-SQS are seen to have converged to about the same value of $\Psi(x^{(k)})$ as the CG algorithm. This correlates with the speed-up factors of 2 and 8 shown to be achieved for the proposed near-optimal relaxation and the use of eight ordered subsets. For cleaner data, OS-SIRT and OS-SQS could be run for more subsets in which case they would converge faster than shown here, possibly to the point where they converge as fast as the PCG algorithm. While the convergence rate for PCG could be improved using a more sophisticated preconditioner, the design and implementation thereof is by no means easy. Combined with their ability to effortlessly incorporate non-negativity, OS-SIRT and OS-SQS thus appear to form viable alternatives to PCG for solving RWLS problems.

VII. Conclusion

We have shown that SIRT can be modified to solve a true WLS problem, i.e., without the otherwise inherent geometric weighting. We have also shown that such a version of SIRT and a commonly used version of SQS solve the same Tikhonov regularized WLS problem using the same gradient descent approach except for their (diagonal) preconditioning and step size. We developed practical methods for selecting the step sizes for both algorithms. Empirical results suggest these step sizes to achieve near-optimal relaxation across a wide range of data. We proposed a heuristic adjustment to the step size estimates that accounts for imbalances when using ordered subsets. The convergence rates for SIRT and SQS were found to be indistinguishable.

We compared the proposed relaxed OS-SIRT and OS-SQS algorithms with an implementation of PCG. Non-negativity was not applied due to associated implementation complications for PCG. We found OS-SIRT and OS-SQS to converge faster than PCG during the early iterations. However, OS-SIRT and OS-SQS eventually succumbed to limit cycle behavior due to data inconsistencies, resulting in PCG achieving a lower final value of the cost function. A qualitative comparison of the associated images did not support a lower cost function value corresponding to a more desirable image. OS-SIRT and OS-SQS thus appear to form viable alternatives to PCG for solving RWLS problems.

While SIRT cannot be extended to handle non-quadratic regularizers, SQS was designed with these in mind. Compelling future work includes the possibility of introducing relaxation of such an extension. Comparison with acceleration based on Nesterov's momentum [15] would be interesting in that context.

Acknowledgments

References

1. Sauer K, Bouman C. A local update strategy for iterative reconstruction from projections. IEEE Trans Signal Proc. 1993; 41:534–548.

2. Erdogan H, Fessler JA. Ordered subsets algorithms for transmission tomography. Physics in Medicine and Biology. 1999; 44:2835–2851. [PubMed: 10588288]

3. Elbakri IA, Fessler JA. Statistical image reconstruction for polyenergetic x-ray computed tomography. IEEE Trans Medical Imaging. 2002; 21:89–99. [PubMed: 11929108]

4. Kak, A.; Slaney, M. Principles of CT Imaging. SIAM; 2001.

5. Shi D, Zou Y, Zamyatin AA. Weighted simultaneous algebraic econstruction technique. Proc Intl Mtg on Fully 3D Image Recon in Rad and Nuc Med. 2011:160–2.

6. Shi D. Unified interpretations of variants of simultaneous algebraic reconstruction technique (SART). Proc Intl Mtg on Fully 3D Image Recon in Rad and Nuc Med. 2013:166–8.

7. Gregor J, Benson T. Computational analysis and improvement of SIRT. IEEE Trans Medical Imaging. 2008; 27:918–924. [PubMed: 18599397]

8. Gregor J. Algorithmic improvements to SIRT with application to X-ray CT of luggage. Proc 3rd Intl Mtg on Image Formation in X-ray CT. 2014:121–4.

9. Hudson HM, Larkin RS. Accelerated image reconstruction using ordered subsets of projection data. IEEE Trans Medical Imaging. 1994; 13:601–609. [PubMed: 18218538]

10. Gilbert P. Iterative methods for the three-dimensional reconstruction of an object from projections. Journal of Theoretical Biology. 1972; 36:105–117. [PubMed: 5070894]

11. Herman G, Lent A. Iterative reconstruction algorithms. Comput Biol Med. 1976; 6:273–294. [PubMed: 1000955]

12. Herman G, Lent A. Quadratic optimization for image reconstruction I. Computer Graphics and Image Processing. 1976; 5:319–332.

13. Lakshminarayanan A, Lent A. Methods of least squares and SIRT in reconstruction. Journal of Theoretical Biology. 1979; 76:267–295. [PubMed: 439905]

14. van der Sluis A, van der Horst HA. SIRT and CG-type methods for the iterative solution of sparse linear least-squares problems. Linear Algebra and its Applications. 1990; 130:257–302.

15. Kim D, Ramani S, Fessler HA. Combining ordered subsets and momentum for accelerated X-ray CT image reconstruction. IEEE Trans Medical Imaging. 2015 To appear.

16. Delaney AH, Bresler Y. Globally convergent edge-preserving regularized reconstruction: an application to limited-angle tomography. IEEE Trans Image Processing. 1998; 7:204–221.

17. Saad, Y. Iterative Methods for Sparse Linear Systems. PWS Publishing Co; Boston: 1996.

18. Horn, R.; Johnson, C. Matrix Analysis. Cambridge Univ. Press; 1985.

19. Golub, G.; van Loan, C. Matrix Computations. Johns Hopkins Univ. Press; 1996.

20. Jiang M, Wang G. Convergence of the simultaneous algebraic reconstruction technique (SART). IEEE Trans Image Processing. 2003; 12:957–961.

21. Jiang M, Wang G. Convergence studies on iterative algorithms for image reconstruction. IEEE Trans Medical Imaging. 2003; 22:569–579. [PubMed: 12846426]

22. Wang J, Zheng Y. On the convergence of generalized simultaneous iterative reconstruction algorithms. IEEE Trans Image Processing. 2007; 16:1–6.

23. Kim D, Pal D, Thibault JB, Fessler JA. Accelerating ordered subsets image reconstruction for X-ray CT using spatially non-uniform optimization transfer. IEEE Trans Medical Imaging. 2013; 32:1965–1978. [PubMed: 23751959]

24. De Pierro AR. On the relation between the ISRA and the EM algorithm for positron emission tomography. IEEE Trans Medical Imaging. 1993; 12:328–333. [PubMed: 18218422]

25. De Pierro AR. A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography. IEEE Trans Medical Imaging. 1995; 14:132–137. [PubMed: 18215817]

26. Censor Y, Elfving T. Block-iterative algorithms with diagonally scaled oblique projections for the linear feasibility problem. SIAM J Matrix Anal Appl. 2002; 24:40–58.

27. Jacobson MW, Fessler JA. An expanded theoretical treatment of iteration-dependent majorize-minimize algorithms. IEEE Trans Image Processing. 2007; 16:2411–2422.

28. Andersen AH, Kak AC. Simultaneous algebraic reconstruction technique (SART): A superior implementation of the ART algorithm. Ultrasonic Imaging. 1984; 6:81–94. [PubMed: 6548059]

29. Wang G, Jiang M. Ordered-subset simultaneous algebraic reconstruction technique (OS-SART). Journal of X-ray Science & Technology. 2004; 12:169–177.

30. Benson T, Gregor J. Distributed iterative image reconstruction for micro-CT with ordered-subsets and focus of attention problem reduction. Journal of X-ray Science & Technology. 2004; 12:231–240.

31. Ahn S, Fessler JA. Globally convergent image reconstruction for emission tomography using relaxed ordered subsets algorithms. IEEE Trans Medical Imaging. 2003; 22:613–626. [PubMed: 12846430]

32. http://www.northeastern.edu/alert/transitioning-technology/alert-datasets

33. Press, W.; Teukolsky, S.; Vetterling, W.; Flannery, B. Numerical Recipes. Cambridge Univ. Press; 2007.

34. Fessler JA, Booth SD. Conjugate-gradient preconditioning methods for shift-variant PET image reconstruction. IEEE Trans Image Processing. 1999; 8:688–699.

35. Gregor, J.; Rannou, F. SPIE Medical Imaging Conf. Vol. 4322. San Diego, CA: 2001. Least squares framework for projection MRI reconstruction; p. 888-898.

Biographies

Jens Gregor received the M.S.E.E. degree from Aalborg University, Aalborg, Denmark, in 1988, and earned a Ph.D. degree in electrical engineering in 1991, also from Aalborg University. He has worked at the University of Tennessee, Knoxville, since then. He is now a Professor in the Department of Electrical Engineering and Computer Science as well as an Adjunct Associate Professor of Radiology. His research interests are in pattern recognition and computational aspects of image reconstruction.

Jeffrey A. Fessler received the B.S.E.E. degree from Purdue University in 1985, the M.S.E.E. degree from Stanford University in 1986, and the M.S. degree in Statistics from Stanford University in 1989. From 1985 to 1988 he was a National Science Foundation Graduate Fellow at Stanford, where he earned a Ph.D. degree in electrical engineering in 1990. He has worked at the University of Michigan since then. From 1991 to 1992 he was a Department of Energy Alexander Hollaender Post-Doctoral Fellow in the Division of Nuclear Medicine. From 1993 to 1995 he was an Assistant Professor in Nuclear Medicine and the Bioengineering Program. He is now a Professor in the Departments of Electrical Engineering and Computer Science, Radiology, and Biomedical Engineering. He became a Fellow of the IEEE in 2006, for contributions to the theory and practice of image reconstruction. He received the Francois Erbsmann award for his IPMI93 presentation, and the Edward Hoffman Medical Imaging Scientist Award in 2013. He has served as an associate editor for IEEE Transactions on Medical Imaging, the IEEE Signal Processing Letters, and the IEEE Transactions on Image Processing, and is currently serving as an associate editor for the IEEE Transactions on Computational Imaging. He has chaired the IEEE T-MI Steering Committee and the ISBI Steering Committee. He was co-chair of the 1997 SPIE conference on Image Reconstruction and Restoration, technical program co-chair of the 2002 IEEE International Symposium on Biomedical Imaging (ISBI), and general chair of ISBI 2007. His research interests are in statistical aspects of imaging problems.
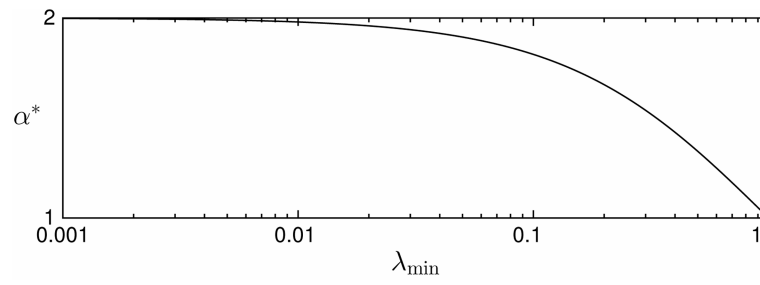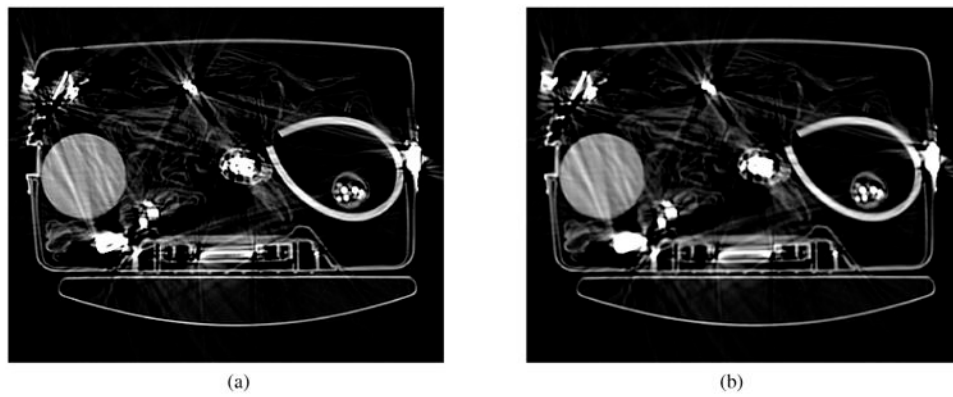
Fig. 1.
The value for optimal step size α^* quickly approaches 2 as eigenvalue $\lambda_{\min}$ of $C\tilde{A}'WA$ drops below 1.

Fig. 2.
DS1 reconstructions using minimum norm regularization: (a) SIRT-RWLS using $\beta = 1.0$ and (b) SIRT-RWLS using $\beta = 5.0$. Image intensity levels truncated to 0-1800 MHU.

Fig. 3.
DS2 reconstructions using finite difference regularization: (a) SQS-RWLS using $\beta = 1.0$ and (b) SQS-RWLS using $\beta = 5.0$. Image intensity levels truncated to 0-1800 MHU.
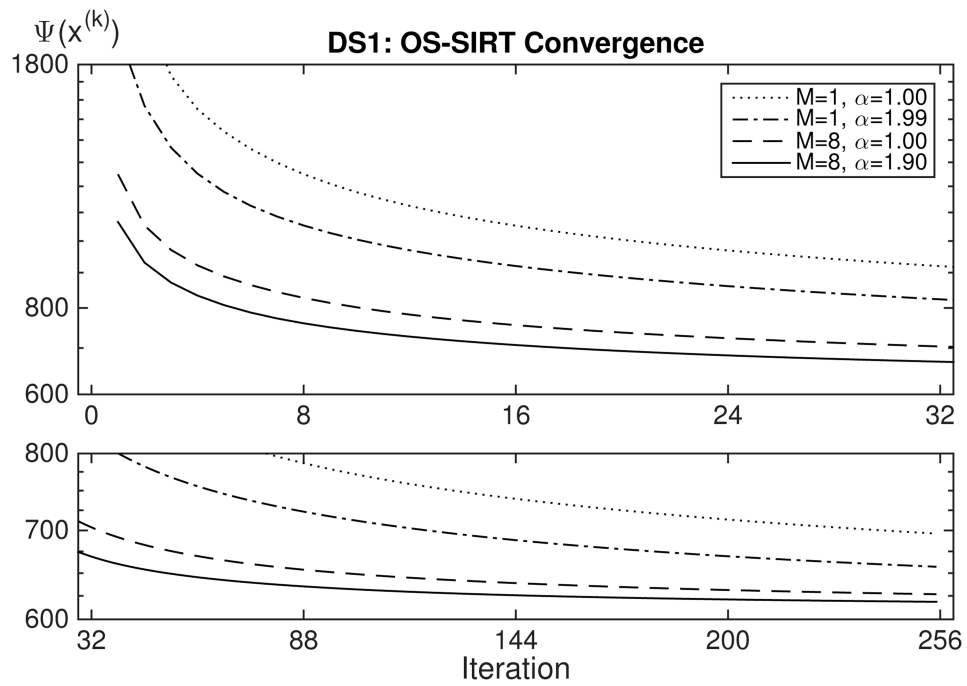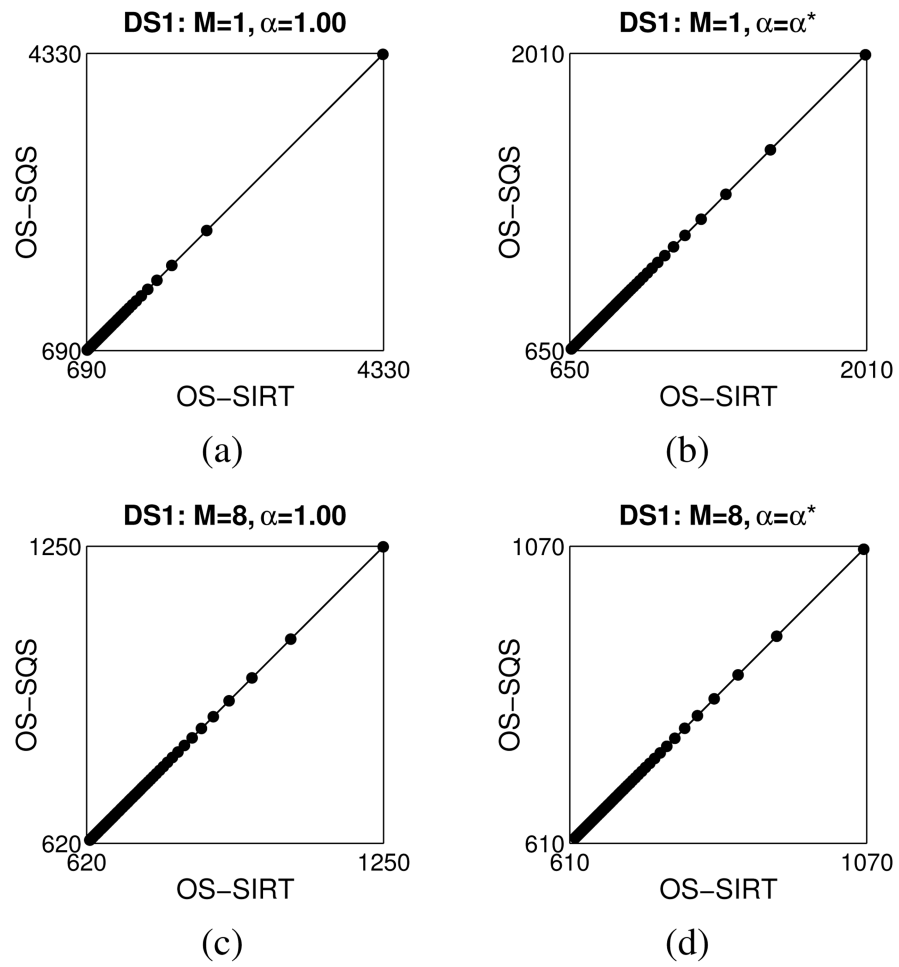
Fig. 4.
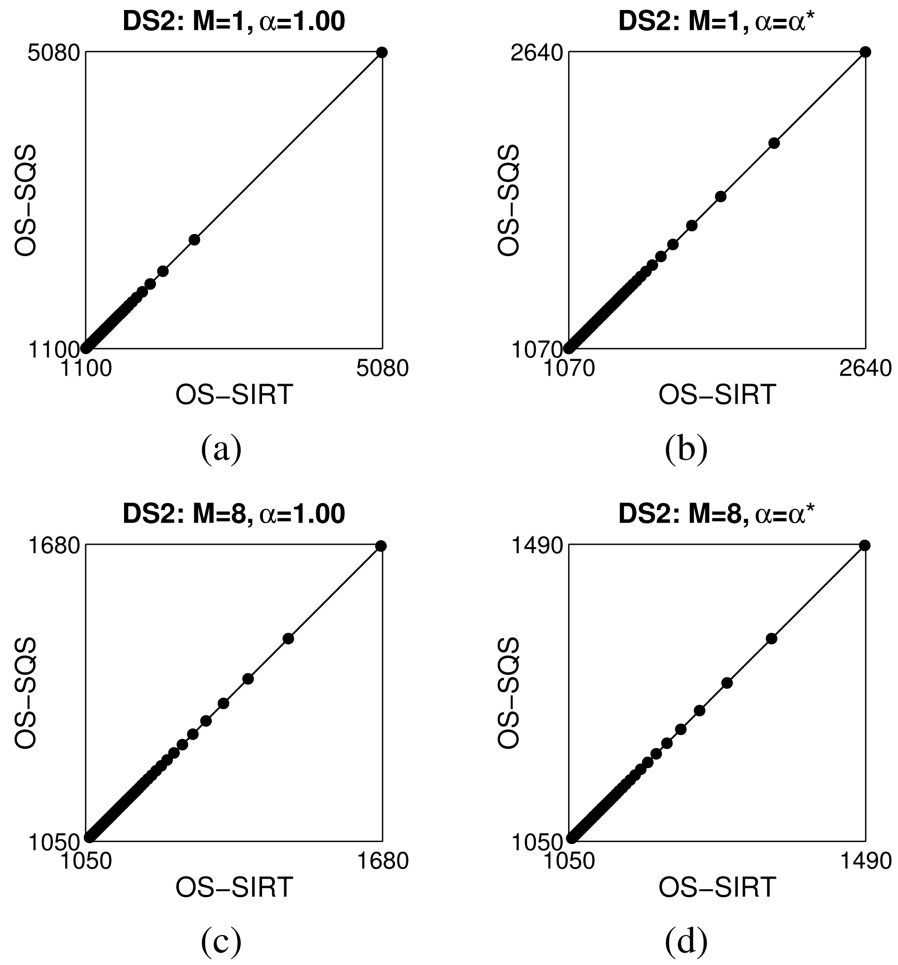Cost function $\Psi(\boldsymbol{x}^{(k)})$ plots for application of OS-SIRT to data set DS1 using minimum norm regularization.
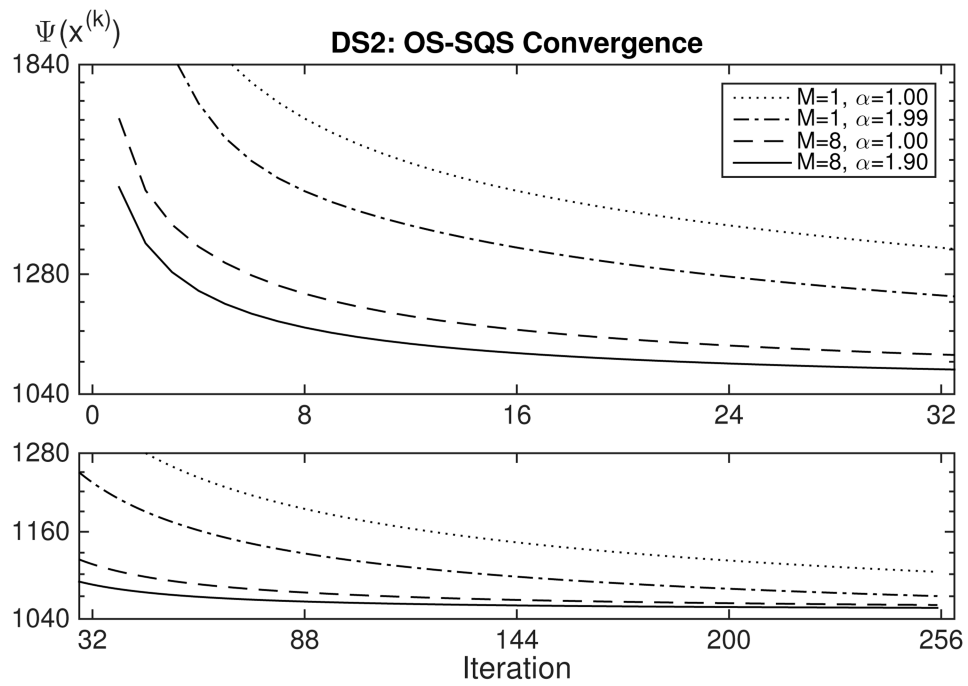
Fig. 5.
Cost function $\Psi(\boldsymbol{x}^{(k)})$ plots for application of OS-SQS to data set DS2 using finite difference regularization.

DS1: M=1, α=1.00

DS1: M=1, α=α*

(a)

(b)

DS1: M=8, α=1.00
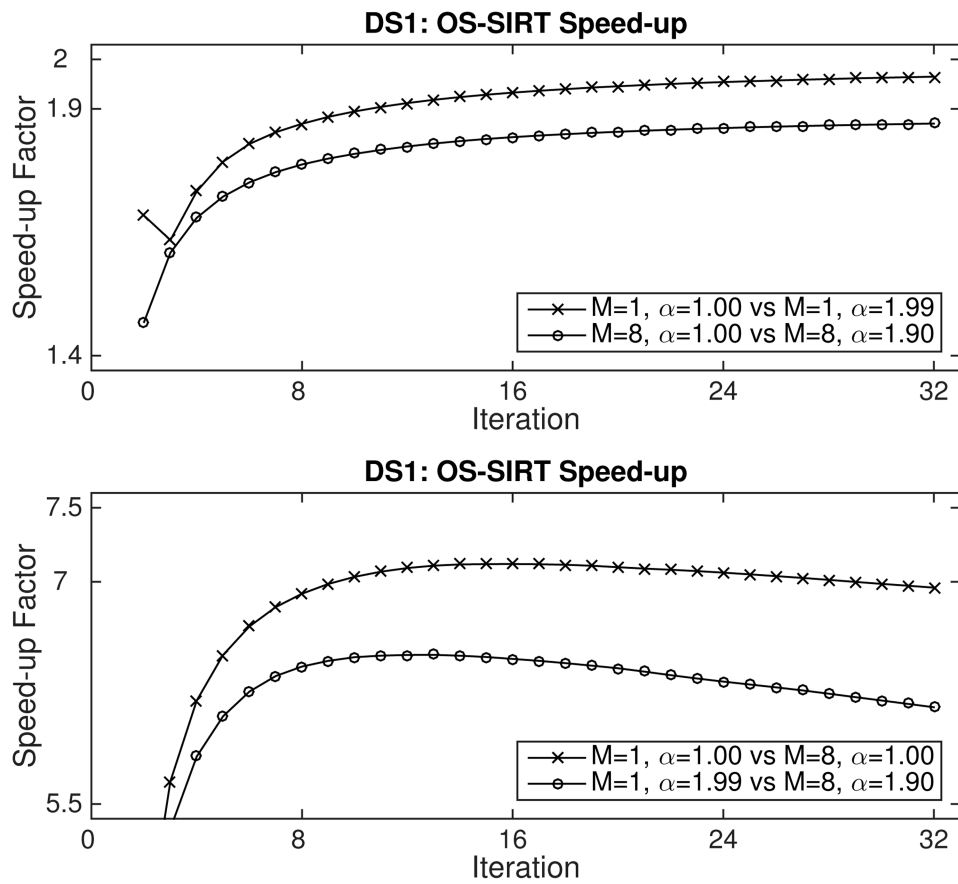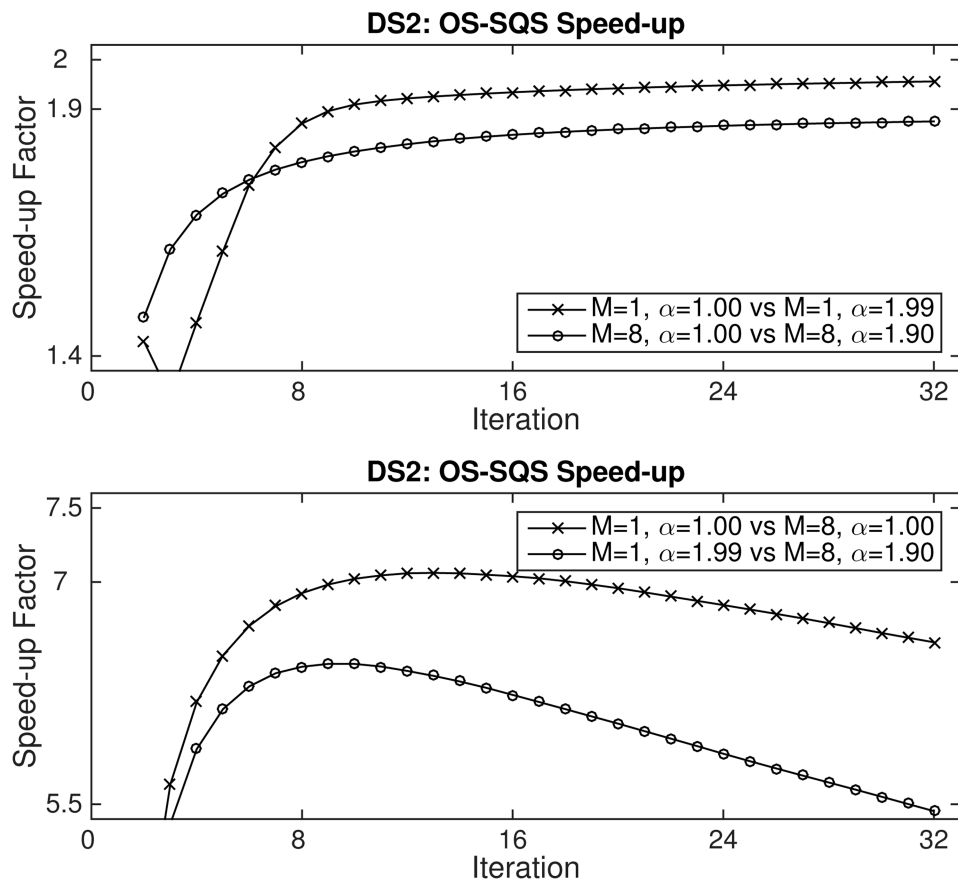
DS1: M=8, α=α*

(c)

(d)

Fig. 6.
Scatter plots of OS-SIRT versus OS-SQS for DS1 reconstructions using minimum norm regularization.
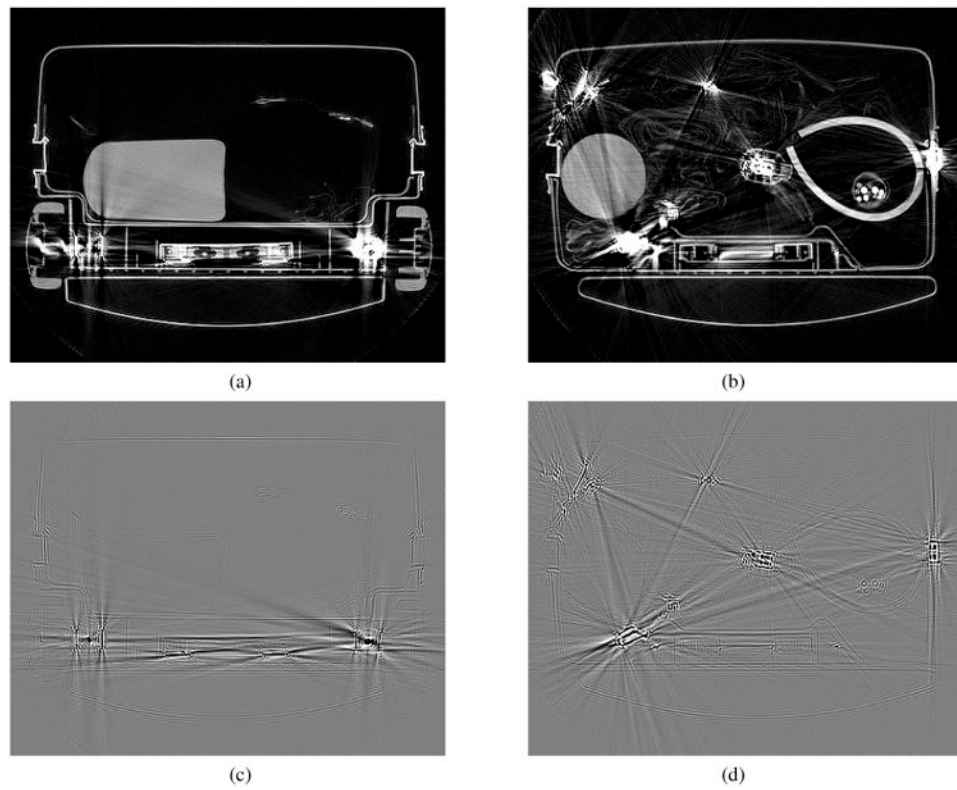
Fig. 7.
Scatter plots of OS-SIRT versus OS-SQS for DS2 reconstructions using finite difference regularization.

Fig. 8.
OS-SIRT speed-up factors for data set DS1 using minimum norm regularization. Top: $\alpha = 1$ versus $\alpha = \alpha^*$. Bottom: $M = 1$ versus $M = 8$.

Fig. 9.
OS-SQS speed-up factors for data set DS2 using finite difference regularization. Top: $\alpha = 1$ versus $\alpha = \alpha^*$. Bottom: $M = 1$ versus $M = 8$.

Fig. 10.
PCG-RWLS reconstructions: (a) DS1 using minimum norm regularization, (b) DS2 using finite difference regularization, (c) DS1 difference between PCG-RWLS and SIRT-RWLS and (d) DS2 difference between PCG-RWLS and SQS-RWLS. Image intensity levels truncated to 0-1800 MHU for (a) and (b). Difference image intensity levels truncated to ±1000 MHU for (c) and (d).
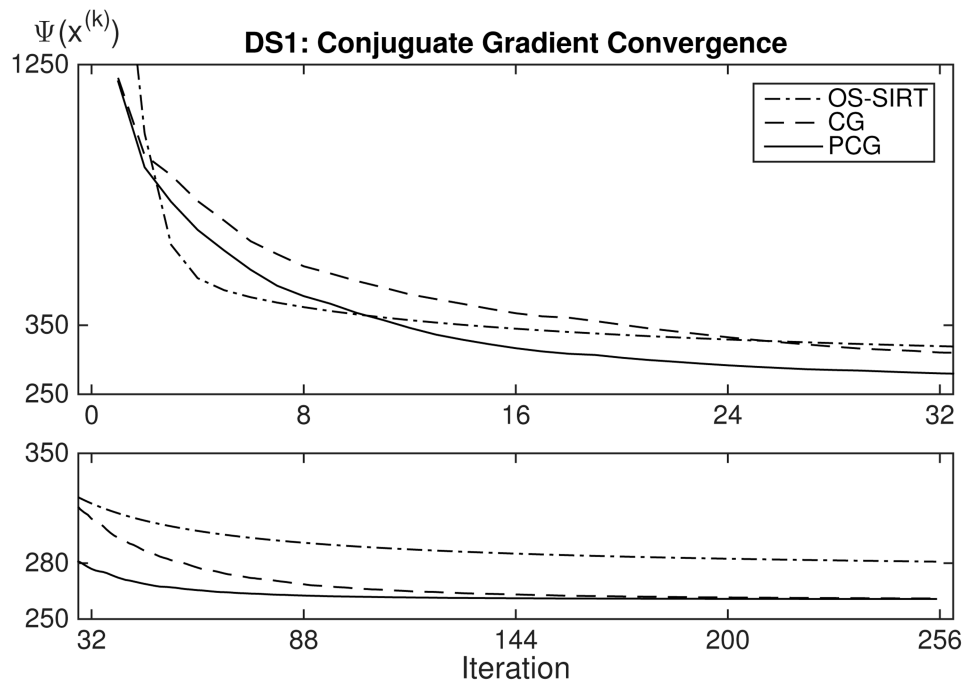
Fig. 11.
Cost function $\Psi(x^{(k)})$ plots for application of OS-SIRT, CG and PCG to data set DS1 using minimum norm regularization.
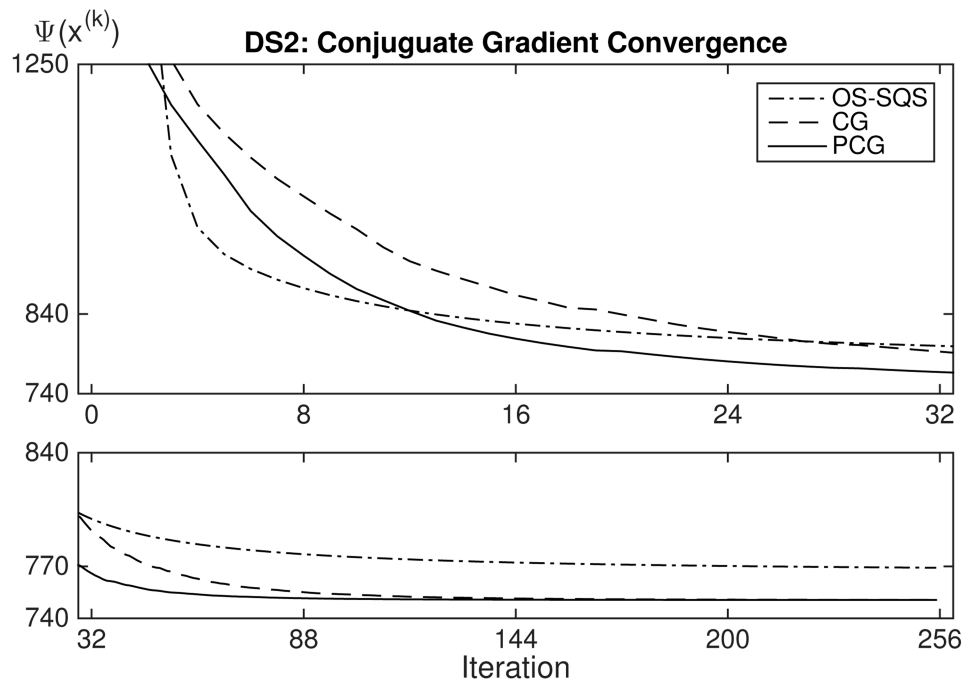
Fig. 12.
Cost function $\Psi(\boldsymbol{x}^{(k)})$ plots for application of OS-SQS, CG and PCG to data set DS2 using finite difference regularization.

Table I

SIRT Column Sum Statistics

	Both-LS	DS1-WLS	DS2-WLS
$\max_j \tilde{c}_j$	666,140	412,423	321,432
$\min_j \tilde{c}_j$	304,479	3,941	1,026

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table II

SIRT Eigenvalue Bounds

Data	Reg.	$\Lambda(C\tilde{A}'WA)$	$\Lambda(C\tilde{Q}'Q)$
DS1	MN	1.000903	0.000256
DS1	FD	1.000903	0.002030
DS2	MN	1.000933	0.000977
DS2	FD	1.000933	0.007794

Table III

SQS Eigenvalue Bounds

Data	Reg.	$\Lambda(\hat{C}A'WA)$	$\Lambda(\hat{C}Q'Q)$
DS1	MN	1.000901	0.000256
DS1	FD	1.000884	0.002026
DS2	MN	1.000930	0.000976
DS2	FD	1.000908	0.007733

Table IV

OS-SIRT Step Size $\tilde{\alpha^*}$ as Function of M.

	DS1 (MN)		DS2 (FD)	
M	$\tilde{S_M}$	$\tilde{\alpha^*}$	$\tilde{S_M}$	$\tilde{\alpha^*}$
1	1.0000	1.9977	1.0000	1.9977
2	1.0094	1.9792	1.0094	1.9792
4	1.0263	1.9466	1.0262	1.9467
8	1.0525	1.8982	1.0524	1.8983
16	1.1240	1.7776	1.1239	1.7777
32	1.3654	1.4636	1.3653	1.4636

Table V

OS-SQS Step Size $\hat{\alpha^*}$ as Function of M.

	DS1 (MN)		DS2 (FD)	
M	$\hat{S}_M$	$\hat{\alpha^*}$	$\hat{S}_M$	$\hat{\alpha^*}$
1	1.0000	1.9827	1.0000	1.9829
2	1.0053	1.9723	1.0052	1.9727
4	1.0158	1.9520	1.0154	1.9530
8	1.0433	1.9011	1.0428	1.9021
16	1.1106	1.7868	1.1103	1.7874
32	1.2859	1.5449	1.2854	1.5456