

ACCELERATED ORDERED-SUBSETS ALGORITHM BASED ON SEPARABLE QUADRATIC SURROGATES FOR REGULARIZED IMAGE RECONSTRUCTION IN X-RAY CT

Donghwan Kim and Jeffrey A. Fessler

The University of Michigan
Department of Electrical Engineering and Computer Science
1301 Beal Avenue, Ann Arbor, MI 48109-2122

ABSTRACT

Iterative algorithms for X-ray CT image reconstruction offer the possibility of reduced dose and/or improved image quality, but require substantial compute time. Reducing the time will likely require algorithms that can be massively parallelized. Ordered subsets (OS) algorithms update all voxels simultaneously and thus are amenable to such parallelization. We present a new monotonic algorithm for regularized image reconstruction that is derived using optimization transfer with separable quadratic surrogates (SQS). The new algorithm accelerates the convergence rate by adapting reduced curvature values for the regularizer that were proposed by Yu *et al.* [1] for coordinate descent algorithms. We further accelerate the algorithm using ordered subsets. Simulation results show that the proposed OS algorithm converges faster than the traditional OS algorithm for X-ray CT reconstruction from a limited number of projection views.

Index Terms— Regularized image reconstruction, optimization transfer, separable quadratic surrogate, ordered subsets, optimum curvature, iterative image reconstruction.

1. INTRODUCTION

X-ray CT image reconstruction using regularized cost functions requires iterative algorithms. Both coordinate descent algorithms [1, 2] and ordered-subsets (OS) algorithms [3, 4] have been investigated, among others. Computation time remains a significant impediment to practical use of such algorithms, and it is likely that practical algorithms will require massive parallelization. We focus here on OS methods because they update all voxels simultaneously, facilitating parallelization.

OS methods usually are derived using optimization transfer [4], where a complicated cost function is replaced by iterative minimization of simpler surrogate functions. Optimization transfer derivations often begin with separable surrogate

This work was supported in part by NIH grant 1-R01-HL-098686 and by a KLA-Tencor graduate fellowship. The first author acknowledges Jang Hwan Cho for discussions about the algorithm.

functions that facilitate simultaneous voxel updates. Examples include expectation-maximization (EM) algorithms [5, 6], the convex algorithm [7], and separable quadratic surrogates (SQS) [4]. Such algorithms can readily accommodate non-quadratic edge-preserving regularizers and can enforce nonnegativity constraints. However, generally they require more iterations to converge than coordinate descent methods, particularly for high spatial frequencies [8], so it is desirable to accelerate them.

We can improve the convergence rate of an optimization transfer method by reducing the curvature of the surrogate functions, provided the monotonicity conditions are retained. Recently, Yu *et al.* [1] proposed a new curvature for edge-preserving regularization and applied it in a coordinate descent algorithm. Here we show that this curvature accelerates the original OS algorithm described in [4], particularly for reconstruction from limited projection views.

2. PROBLEM

The image reconstruction problem is to estimate an image $x \in \mathbb{R}^N$ from noisy measured transmission data $Y \in \mathbb{R}^M$. A simplified model for the measurement statistics is $Y_i \sim \text{Poisson}\{b_i e^{-[Ax]_i} + r_i\}$ for $i = 1, \dots, M$, where b_i, r_i are known nonnegative constants, and $A = \{a_{ij}\}$ is a $M \times N$ system matrix corresponding to forward projection. For simplicity of presentation, we focus on the log data $y_i = \log(b_i/(Y_i - r_i))$, for which we can use the linear model $y = Ax + \epsilon$, where $\epsilon \in \mathbb{R}^M$ denotes additive noise. In limited view tomography, under-sampling artifacts often are a dominant factor of image degradation.

We estimate x by minimizing a regularized cost function:

$$\begin{aligned} \hat{x} &\triangleq \arg \min_x \Psi(x) \\ \Psi(x) &\triangleq Q(x) + \beta R(x) = \frac{1}{2} \|y - Ax\|_W^2 + \beta R(x) \\ &= \sum_{i=1}^M q_i([Ax]_i) + \beta \sum_{k=1}^K \psi_k([Cx]_k) \end{aligned} \quad (1)$$

where $q_i(t) \triangleq \frac{1}{2} w_i(t - y_i)^2$, $\psi_k(t) \triangleq \lambda_k \psi(t)$ and $\psi(t)$ is an

edge-preserving potential function. The $K \times N$ matrix $C = \{c_{kj}\}$ is a finite differencing matrix. The regularization parameter β balances between the data fitting term $Q(x)$ and the penalty term $R(x)$. The diagonal matrix $W = \text{diag}\{w_i\}$ provides statistical weighting, and λ_k provides optional weights for the regularizer.

3. ALGORITHM

This section derives separable quadratic surrogates (SQS) for both the data-fit term $Q(x)$ and the regularizer $R(x)$.

3.1. SQS for data-fit term $Q(x)$

For completeness, we repeat the arguments in [4, 9]. Using the method in [6], we rewrite $[Ax]_i$ as follows:

$$[Ax]_i = \sum_{j=1}^N a_{ij}x_j = \sum_{j=1}^N \pi_{ij} \left(\frac{a_{ij}}{\pi_{ij}}(x_j - x_j^{(n)}) + [Ax^{(n)}]_i \right), \quad (2)$$

where $\sum_{j=1}^N \pi_{ij} = 1$ and π_{ij} is nonnegative and zero only if a_{ij} is zero. Using the convexity inequality:

$$q_i([Ax]_i) \leq \sum_{j=1}^N \pi_{ij} q_i \left(\frac{a_{ij}}{\pi_{ij}}(x_j - x_j^{(n)}) + [Ax^{(n)}]_i \right). \quad (3)$$

Thus, a separable surrogate for $Q(x)$ is

$$Q(x) \leq \phi_Q^{(n)}(x) \triangleq \sum_{j=1}^N \phi_{Q,j}^{(n)}(x_j) \quad (4)$$

$$\phi_{Q,j}^{(n)}(x_j) \triangleq \sum_{i=1}^M \pi_{ij} q_i \left(\frac{a_{ij}}{\pi_{ij}}(x_j - x_j^{(n)}) + [Ax^{(n)}]_i \right). \quad (5)$$

We choose $\pi_{ij} \triangleq \frac{a_{ij}}{\sum_{i=1}^M a_{i*}} = \frac{a_{ij}}{a_{i*}}$, where $a_{ij} \geq 0$ for the projection matrix A . Then the surrogate $\phi_Q^{(n)}(x)$, which is both separable and quadratic, can be rewritten as follows:

$$\begin{aligned} \phi_Q^{(n)}(x) &= Q(x^{(n)}) + \nabla Q(x^{(n)})(x - x^{(n)}) \\ &\quad + \frac{1}{2}(x - x^{(n)})' \text{diag}\{d_j^Q\}(x - x^{(n)}) \end{aligned} \quad (6)$$

where $d_j^Q = \frac{\partial^2}{\partial x_j^2} \phi_Q^{(n)}(x) = \sum_{i=1}^M w_i a_{ij} a_{i*}$.

3.2. SQS for regularizer $R(x)$

3.2.1. Separable surrogate for $R(x)$

Similar to the derivation in Sect. 3.1, we first derive a separable surrogate for the regularizer $R(x)$ given in (1). We adapt (4) to an arbitrary matrix C by letting $\pi_{kj} = \frac{|c_{kj}|}{c_{k*}}$,

where $c_{k*} = \sum_{j=1}^N |c_{kj}|$. Assuming $\psi(t)$ is convex, we have

$$R(x) \leq \phi_R^{(n)}(x) \triangleq \sum_{j=1}^N \phi_{R,j}^{(n)}(x_j) \quad (7)$$

$$\phi_{R,j}^{(n)}(x_j) \triangleq \sum_{k=1}^K \lambda_k \rho_{kj} \left(x_j - r_{kj}^{(n)} \right), \quad (8)$$

where $\rho_{kj}(t) \triangleq \frac{|c_{kj}|}{c_{k*}} \psi \left(\frac{c_{kj}}{|c_{kj}|} c_{k*} t \right)$, $r_{kj}^{(n)} \triangleq x_j^{(n)} - \frac{|c_{kj}|}{c_{k*}} \frac{1}{c_{k*}} [Cx^{(n)}]_k$.

Typically the regularization matrix C is sparse, so we can save computation by using the set $N_j = \{k = 1, \dots, K : c_{kj} \neq 0\}$. Combining with the data-fit surrogate yields the overall separable surrogate

$$\phi_j^{(n)}(x_j) = \phi_{Q,j}^{(n)}(x_j) + \sum_{k \in N_j} \lambda_k \rho_{kj} (x_j - r_{kj}^{(n)}). \quad (9)$$

3.2.2. Separable quadratic surrogate for $R(x)$

The surrogate $\phi_R^{(n)}(x)$ is separable, but not quadratic. To simplify minimization, we design a quadratic surrogate for it next. The simplest design is to find an upper bound on the curvature of $\phi_{R,j}^{(n)}(x_j)$, called the maximum curvature. Much smaller curvatures are derived in [10, Lemma 8.3, p.184] that are optimal when minimizing over the entire real line (see (13) below). However, the minimizer of a separable surrogate always lies in a finite interval, and this property provides the opportunity to use the method in [1] that yields even smaller curvatures that can accelerate the convergence rate.

We assume that the potential function $\psi(t)$ satisfies the conditions in [1, Thm 1], then $\rho_{kj}(t)$ also satisfies these. Under the conditions (10) below, the quadratic surrogate function for $\rho_{kj}(t)$ can be defined as

$$\sigma_{kj}^{(n)}(t) = \rho_{kj}(\Delta_{kj}^{(n)}) + \dot{\rho}_{kj}(\Delta_{kj}^{(n)})(t - \Delta_{kj}^{(n)}) + \frac{1}{2} s_{kj}^{(n)} (t - \Delta_{kj}^{(n)})^2,$$

where $\Delta_{kj}^{(n)} \triangleq x_j^{(n)} - r_{kj}^{(n)}$. We design the regularizer surrogate curvature $s_{kj}^{(n)}$ so that the surrogate $\sigma_{kj}^{(n)}(t)$ tightly satisfies the following conditions:

$$\begin{aligned} \rho_{kj}(x_j^{(n)} - r_{kj}^{(n)}) &= \sigma_{kj}^{(n)}(x_j^{(n)} - r_{kj}^{(n)}) \\ \rho_{kj}(x_j - r_{kj}^{(n)}) &\leq \sigma_{kj}^{(n)}(x_j - r_{kj}^{(n)}), \quad \forall x_j \in U_j^{(n)}, \end{aligned} \quad (10)$$

where $U_j^{(n)} = [u_j^{\min(n)}, u_j^{\max(n)}]$ is an interval containing the minimizer of $\phi_j^{(n)}(x_j)$. This set is computed by finding the smallest and the largest minimizers of the functions $\{\phi_{Q,j}^{(n)}(x_j), \rho_{kj}(x_j - r_{kj}^{(n)}), k \in N_j\}$. The minimizer of $\phi_{Q,j}^{(n)}(x_j)$ is $q_j^{(n)} = x_j^{(n)} - \frac{1}{d_j^Q} \frac{\partial}{\partial x_j} Q(x^{(n)})$, and the minimizer of $\rho_{kj}(x_j - r_{kj}^{(n)})$ is $r_{kj}^{(n)}$. Then we set $u_j^{\min(n)}$ and $u_j^{\max(n)}$ to be the minimum and maximum respectively of

$\{q_j^{(n)}, r_{kj}^{(n)}, k \in N_j\}$. Having $U_j^{(n)}$, the optimum curvature $s_{kj}^{(n)}$ of $\sigma_{kj}^{(n)}(t)$ satisfying the conditions (10) is computed with the method in [1, Fig. 12]:

$$t_{kj}^{(n)} = \begin{cases} -\Delta_{kj}^{(n)}, & |\Delta_{kj}^{(n)}| \leq \min\{|\Delta_{kj}^{\min(n)}|, |\Delta_{kj}^{\max(n)}|\} \\ \Delta_{kj}^{\min(n)}, & |\Delta_{kj}^{\min(n)}| \leq \min\{|\Delta_{kj}^{(n)}|, |\Delta_{kj}^{\max(n)}|\} \\ \Delta_{kj}^{\max(n)}, & \text{otherwise,} \end{cases} \quad (11)$$

where $\Delta_{kj}^{\min(n)} \triangleq u_j^{\min(n)} - r_{kj}^{(n)}$, $\Delta_{kj}^{\max(n)} \triangleq u_j^{\max(n)} - r_{kj}^{(n)}$.

The optimum curvature $s_{kj}^{(n)}$, which is the smallest curvature of $\sigma_{kj}^{(n)}(t)$ satisfying the conditions (10), is

$$s_{kj}^{(n)} = \begin{cases} 2 \left(\frac{\rho_{kj}(t_{kj}^{(n)}) - \rho_{kj}(\Delta_{kj}^{(n)})}{(t_{kj}^{(n)} - \Delta_{kj}^{(n)})^2} - \frac{\dot{\rho}_{kj}(\Delta_{kj}^{(n)})}{t_{kj}^{(n)} - \Delta_{kj}^{(n)}} \right), & \Delta_{kj}^{(n)} \neq 0 \\ \ddot{\rho}_{kj}(0), & \text{otherwise.} \end{cases} \quad (12)$$

The traditional curvature in most previous work is found by computing $s_{kj}^{(n)}$ for $t_{kj}^{(n)} = -\Delta_{kj}^{(n)}$ where $U_j^{(n)} = [-\infty, \infty]$, and turns out to be

$$s_{kj}^{(n)} = \frac{\dot{\rho}_{kj}(\Delta_{kj}^{(n)})}{\Delta_{kj}^{(n)}}, \quad (13)$$

which is larger than the optimum curvature (12).

The quadratic surrogate function of the separable surrogate $\phi_{R,j}^{(n)}(x_j)$ in (8) can be defined as

$$\phi_{SQS,j}^{R(n)}(x_j) \triangleq \sum_{k \in N_j} \lambda_k \sigma_{kj}^{(n)}(x_j - r_{kj}^{(n)}) \quad (14)$$

Then the overall separable quadratic surrogate function for $R(x)$ can be written as follows:

$$\begin{aligned} \phi_{SQS}^{R(n)}(x) &= R(x^{(n)}) + \nabla R(x^{(n)})(x - x^{(n)}) \\ &\quad + \frac{1}{2}(x - x^{(n)})' \text{diag}\{d_j^{R(n)}\}(x - x^{(n)}). \end{aligned} \quad (15)$$

where $d_j^{R(n)} = \frac{\partial^2}{\partial x_j^2} \phi_{SQS}^{R(n)}(x) = \sum_{k \in N_j} \lambda_k s_{kj}^{(n)}$.

3.3. Iterative parallelized update for SQS

To summarize the surrogate derivations above:

$$\begin{aligned} \Psi(x) &\leq \phi_{SQS}^{(n)}(x) = \phi_Q^{(n)}(x) + \beta \phi_{SQS}^{R(n)}(x) \\ &= \Psi(x^{(n)}) + \nabla \Psi(x^{(n)})(x - x^{(n)}) \\ &\quad + \frac{1}{2}(x - x^{(n)})' \text{diag}\{d_j^{(n)}\}(x - x^{(n)}), \end{aligned} \quad (16)$$

where $d_j^{(n)} = d_j^Q + \beta d_j^{R(n)}$. Because $\phi_{SQS}^{(n)}(x)$ is quadratic, its unconstrained minimizer is easily derived to be:

$$\hat{u}_j^{(n)} = \arg \min_u \phi_{SQS,j}^{(n)}(u) = x_j^{(n)} - \frac{1}{d_j^{(n)}} \frac{\partial}{\partial x_j} \Psi(x^{(n)}). \quad (17)$$

However, $x_j^{(n+1)}$ must be in the set $U_j^{(n)}$ to maintain the monotonicity, so we revise (17) to be:

$$x_j^{(n+1)} = \arg \min_{u \in U_j^{(n)}} \phi_{SQS,j}^{(n)}(u) = \text{clip}\{\hat{u}_j^{(n)}, U_j^{(n)}\}, \quad (18)$$

where the clip function chooses a nearest value in the interval $U_j^{(n)}$ when $\hat{u}_j^{(n)}$ falls out of the $U_j^{(n)}$. All voxels can be updated simultaneously using (18).

3.4. Reducing the interval $U_j^{(n)}$

The algorithm can be further accelerated while preserving monotonicity by reducing the interval $U_j^{(n)}$ which in turn decreases the the optimum curvature (12). Define

$$\bar{U}_j^{(n)} = \begin{cases} [x_j^{(n)} - \eta \delta_j^{\min(n)}, x_j^{(n)} + \eta \delta_j^{\max(n)}], & x_j^{(n)} \in U_j^{(n)} \\ U_j^{(n)}, & \text{otherwise,} \end{cases}$$

where $\delta_j^{\min(n)} = x_j^{(n)} - u_j^{\min(n)}$, $\delta_j^{\max(n)} = u_j^{\max(n)} - x_j^{(n)}$ and $\eta \in [0, 1]$ is a reduction factor. However, too much reduction can confine the update to a small interval, possibly slowing convergence.

3.5. Ordered subsets

Further acceleration is possible using ordered subsets, while possibly losing monotonicity [4]. We call the final algorithm the accelerated OS-SQS algorithm (A-OS-SQS). The algorithm reduces to the accelerated SQS algorithm when the number of ordered subsets is one.

4. RESULTS

The algorithm is simulated with the phantom Fig. 1(a). The projection space is 444 radial bins and 20 angles, and the reconstructed image is of size 256×256 . The noisy sinogram data y was generated by the Poisson noise model. The FBP reconstructed image Fig. 1(b) was the initial guess $x^{(0)}$ for the iterative reconstruction. The regularization parameter $\beta = 0.25$ and an edge-preserving hyperbola potential function $\psi(t) = \frac{\delta^2}{3} \left(\sqrt{1 + 3(t/\delta)^2} - 1 \right)$ with $\delta = 0.005$ was empirically chosen to produce a good image. The differencing matrix C had 1st-order differences in 2D. The parameters were chosen as $w_i = \exp(-y_i) \propto \frac{1}{\text{var}(y_i)}$, $\lambda_k = 1$ for horizontal and vertical differences, $\lambda_k = \frac{1}{\sqrt{2}}$ for diagonal differences. We reconstructed an image for each $\eta = 0.5, 0.25, 0.125$ and 1, 2 and 4 ordered subsets. Image reconstruction included the nonnegativity constraint. The OS-SQS and A-OS-SQS reconstructed images in Fig. 1 suggest that iterative image reconstruction can produce better images than FBP.

The proposed method uses smaller curvatures than previous work, thus converging faster than the conventional OS.

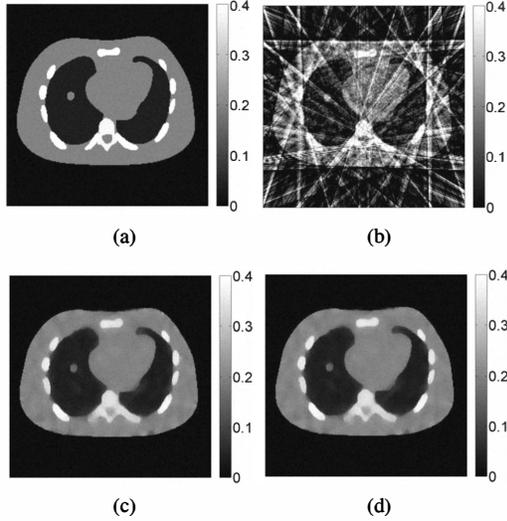


Fig. 1. (a) Phantom image, (b) FBP image $x^{(0)}$, (c) OS-SQS image $x^{(330)}$ and (d) A-OS-SQS image $x^{(290)}$ with $\eta = 0.25$ for 4 ordered subsets. The NRMS difference for both (c) and (d) is -30 [dB].

Fig. 2 plots NRMS difference [dB] between the current image $x^{(n)}$ and the converged image \hat{x} after 3000 iterations, versus iteration. The results show that A-OS-SQS is about 15% faster than the standard OS-SQS. This is a modest improvement, but the extra curvature computation required in (12) is small compared to forward projection. Fig. 2 illustrates that both reducing the interval $U_j^{(n)}$ and using ordered subsets accelerate the SQS algorithm, as expected. However, the results show that reducing the interval too much slows down the convergence speed. We found that for densely sampled view angles, the acceleration was less significant because the reduced curvatures in $R(x)$ are overwhelmed by the curvature of $Q(x)$.

5. CONCLUSION

We introduced the accelerated OS-SQS algorithm for regularized image reconstruction. The new algorithm converged somewhat faster than the previous OS method, with a small increase in computation per iteration for calculating the new curvatures (12). The new algorithm, like previous OS methods, is amenable to massive parallelization.

6. REFERENCES

[1] Z. Yu, J-B. Thibault, C. A. Bouman, K. D. Sauer, and J. Hsieh, "Fast model-based X-ray CT reconstruction using spatially non-homogeneous ICD optimization," *IEEE Trans. Im. Proc.*, 2010.

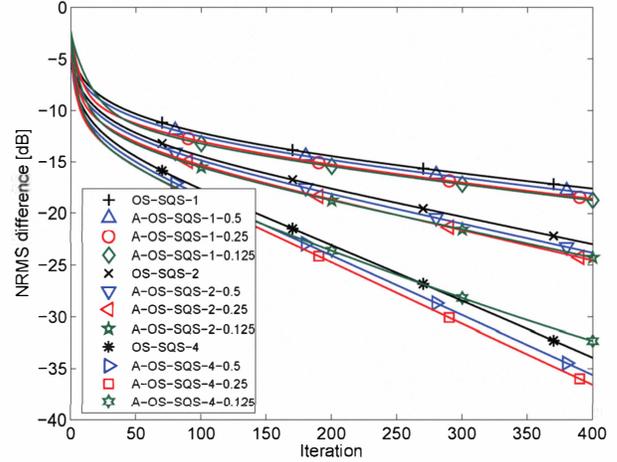


Fig. 2. NRMS difference [dB] versus iterations of OS-SQS and A-OS-SQS with $\eta = 0.5, 0.25, 0.125$ for 1, 2 and 4 ordered subsets

[2] J-B. Thibault, K. Sauer, C. Bouman, and J. Hsieh, "A three-dimensional statistical approach to improved image quality for multi-slice helical CT," *Med. Phys.*, vol. 34, no. 11, pp. 4526–44, Nov. 2007.

[3] J. Nuyts, B. De Man, P. Dupont, M. Defrise, P. Suetens, and L. Mortelmans, "Iterative reconstruction for helical CT: A simulation study," *Phys. Med. Biol.*, vol. 43, no. 4, pp. 729–37, Apr. 1998.

[4] H. Erdoğan and J. A. Fessler, "Ordered subsets algorithms for transmission tomography," *Phys. Med. Biol.*, vol. 44, no. 11, pp. 2835–51, Nov. 1999.

[5] K. Lange and R. Carson, "EM reconstruction algorithms for emission and transmission tomography," *J. Comp. Assisted Tomo.*, vol. 8, no. 2, pp. 306–16, Apr. 1984.

[6] A. R. De Pierro, "A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography," *IEEE Trans. Med. Imag.*, vol. 14, no. 1, pp. 132–7, Mar. 1995.

[7] K. Lange and J. A. Fessler, "Globally convergent algorithms for maximum a posteriori transmission tomography," *IEEE Trans. Im. Proc.*, vol. 4, no. 10, pp. 1430–8, Oct. 1995.

[8] K. Sauer and C. Bouman, "A local update strategy for iterative reconstruction from projections," *IEEE Trans. Sig. Proc.*, vol. 41, no. 2, pp. 534–48, Feb. 1993.

[9] H. Erdoğan and J. A. Fessler, "Monotonic algorithms for transmission tomography," *IEEE Trans. Med. Imag.*, vol. 18, no. 9, pp. 801–14, Sept. 1999.

[10] P. J. Huber, *Robust statistics*, Wiley, New York, 1981.