COMPARISON OF REGULARIZED AND SUPERIORIZED METHODS FOR TOMOGRAPHIC IMAGE RECONSTRUCTION

T. HUMPHRIES*, M. LORETO, B. HALTER, W. O'KEEFFE, AND L. RAMIREZ

Division of Engineering & Mathematics, School of STEM, University of Washington Bothell, Bothell, WA, USA

Abstract. We compare two approaches to image reconstruction in computed tomography (CT) which incorporate penalty functions to improve image quality in the presence of noisy data. The first approach adapts a previously proposed hybrid method for solving a regularized least squares problem, which simultaneously computes the regularization parameter and the corresponding solution. The second approach is based on the superiorization methodology, wherein the solution is perturbed between iterations of a feasibility-seeking algorithm to minimize a secondary objective. Numerical experiments indicate that while both approaches are able to significantly improve image quality, the heuristic applied to select the regularization parameter in the hybrid method does not generalize well to the CT reconstruction problem. The superiorization methodology is more effective, provided that a suitable stopping criterion can be determined.

Keywords. spectral projected gradient, regularization, superiorization, image reconstruction, tomography.

2010 Mathematics Subject Classification. 49M305, 90C52

1. INTRODUCTION

Computed tomography (CT) is a medical imaging modality in which X-ray measurements gathered from a number of views around the patient are used to reconstruct an image of the patient anatomy. Mathematically, the measurements are modeled as line integrals through an attenuation function that represents the distribution of body tissue; the set of all such line integrals represent the X-ray transform of the function. Techniques for reconstructing a CT image fall into two categories: analytical methods, which are based on directly inverting the X-ray transform, and iterative methods, which begin with an initial image estimate that is iteratively improved according to some optimization criteria. Analytical methods have the advantage of being computationally much less expensive than iterative methods; however, iterative methods offer more flexibility to accurately model the physics of CT acquisition, and to incorporate prior information about the image to be reconstructed [1].

In iterative methods, the image reconstruction problem is most often modeled as a large system of linear equations, $Ax = b$. As the system is ill-conditioned and the data are noisy, a solution to the corresponding least-squares problem,

$$x^* = \arg\min_x \ f_{LSQ}(x), \qquad f_{LSQ}(x) = \frac{1}{2}\|Ax-b\|_2^2 \tag{1}$$

is likely to be of poor quality. For this reason, iterative reconstruction methods often incorporate prior information, in the form of a function that penalizes undesirable image qualities. Classical choices include the ℓ_2 or ℓ_1 norms, or more commonly in image reconstruction, edge-preserving penalties such as total variation [2] or the Huber prior [3].

*Corresponding author.

E-mail address: thumphri@uw.edu

Received XXXXXX; Accepted XXXXXX

Prior information can be incorporated into the reconstruction algorithm in several ways. One is to solve a regularized least-squares problem [4, 5]:

$$x^* = \arg\min_x f_{RLSQ}(x;\lambda), \qquad f_{RLSQ}(x;\lambda) = f_{LSQ}(x) + \lambda\phi(x), \tag{2}$$

where $\phi(x)$ is a penalty function, and λ is a parameter controlling the weighting of the least squares term and the penalty. An important question in this problem is how to determine a value for the parameter λ; a value that is too small may not substantially improve image quality, while a value that is too large will over-regularize, resulting in an oversmoothed or otherwise inaccurate reconstruction. The optimal value of λ depends on the level of noise in the data, as well as the nature of the inverse problem.

A second approach is to *superiorize* [6] an algorithm for solving the least squares problem (1) by perturbing the solution between iterations to reduce the value of ϕ. Under certain conditions (discussed later in the paper), the superiorized version of the algorithm is guaranteed to eventually produce a solution which is of equivalent quality to the original algorithm with respect to solving (1); the expectation is that this solution should also be superior with respect to the penalty function. In place of the parameter λ in (2), superiorization requires specifying parameters which control the size and number of perturbations introduced between iterations.

In this paper, we compare a regularized and a superiorized approach to the CT image reconstruction problem. The first approach is based on a hybrid optimization algorithm previously proposed by Guerrero et al. [7] to solve the regularized problem. This method alternates between using the spectral projected gradient (SPG) method [8] to solve (2) for some choice of λ, then updating the choice of λ using simulated annealing [9]. The second method superiorizes the simultaneous algebraic reconstruction technique (SART) [10], a well-known iterative method in CT imaging. In comparing the approaches, we consider the quality of the images produced, the robustness of the algorithm with respect to different levels of noise, and the computational cost.

2. Materials and Methods

We first discuss the fundamentals of the CT imaging problem, followed by a description of the algorithms used in our study, as well as the the choice of penalty functions.

2.1. CT reconstruction.

The idealized physical model for X-ray imaging in two dimensions is given by Beer's Law:

$$I = I_0 \exp\left(-\int_j \mu(y)\,dy\right), \tag{3}$$

where I_0 is the initial intensity of the X-ray beam, I is the detected intensity on the other side of the object, $\mu(y)$ is the attenuation function which varies with position $y \in \mathbb{R}^2$, and $\int_j$ represents an integral along some line j through the object. After rearranging, (3) can be written as

$$\int_j \mu(y)\,dy = -\ln\frac{I}{I_0}, \tag{4}$$

expressing the line integral in terms of the log-transformed measurement.

In practice, the interaction of X-rays with matter is stochastic, and so the measured intensity is modeled as a Poisson random variable $\hat{I}$ with mean I. Since the standard deviation of a Poisson distribution is equal to the square root of the mean, the signal-to-noise ratio of the measurement is equal to $\sqrt{I}$. As a

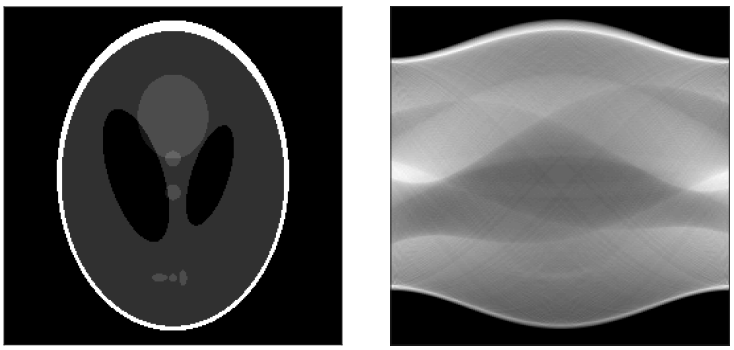

FIGURE 1. 256×256 pixel image of the Shepp-Logan phantom (left) and its sinogram (right). The sinogram is a sampling of the X-ray transform of the phantom, with the horizontal axis corresponding to different angles of view, and vertical axis corresponding to equally spaced line integrals through the phantom from that angle.

result, measurements are noisier when I is small, either because the initial beam intensity (I_0) is low, or if the value of the line integral is large; i.e., the beam passes through strongly attenuating material. The first case is particularly important in CT imaging, as one typically wishes to keep the dose to the patient as low as possible in clinical imaging.

The collection of all line integrals through μ over a 180° arc is known as its X-ray transform. An example of a commonly used numerical phantom and a sampling of its X-ray transform, or sinogram, are shown in Figure 1. The CT reconstruction problem is to recover the image of μ from its sinogram. Analytical methods, the most fundamental of which is filtered back projection (FBP) (see, e.g., [11]), are based on a mathematical formula to directly invert the X-ray transform. Filtered back projection is computationally fast, but unable to incorporate physical effects beyond the line integral model, or any prior information. Iterative methods [1] overcome this limitation by first representing the measurements in terms of a linear system of equations, $b = Ax$, where b is a vector consisting of the log-transformed measurements, x is the discretized image of μ to be recovered, and A is the system matrix modeling the contribution of each pixel in x to each measurement in b. This matrix is large, but also sparse, as each entry of b depends only on pixel intensities along one line through the image. The system of equations can then be solved in a least-squares or maximum likelihood sense [1].

2.2. **Spectral Projected Gradient (SPG).** The SPG method was introduced by Birgin et al. [8], and it has been widely studied (see [12] and references therein). The SPG solves problems of the form:

$$x^* = \arg\min_x f(x) \text{ subject to } x \in \Omega,$$

where Ω is a closed convex set in $\mathbb{R}^n$, and f is a smooth function. The method is attractive for its simplicity. It overcomes the traditional slowness of the gradient method by incorporating a spectral step length and a nonmonotone globalization strategy. Also, the method is very efficient for solving large-scale convex constrained problems in which projections onto Ω are easy to compute. Since its appearance, the SPG method has been useful for solving real applications in different areas, including optics, compressive sensing, geophysics, and image restoration, among others [12]. However, to our knowledge it has not been used to solve CT image reconstruction problems.

To obtain an optimal point x^*, SPG computes a trial point x^+ along the spectral projected gradient direction, $d_k = P_\Omega(x_k - \alpha_k g_k) - x_k$, where $g_k = \nabla f(x_k)$, P_Ω is defined as the projection of x onto Ω, and α_k is the spectral step. The point x^+ is tested until it satisfies a globalization condition, and a new iterate is obtained.

The basic idea to obtain the spectral step length is to regard the matrix $(1/\alpha_k)I$ as an approximation to the Hessian $\nabla^2 f(x_k)$ and impose a quasi-Newton secant equation

$$\frac{1}{\alpha_k} s_{k-1} = y_{k-1},$$

where $s_{k-1} = x_k - x_{k-1}$ and $y_{k-1} = \nabla f(x_k) - \nabla f(x_{k-1})$. In general, this equation cannot be satisfied exactly. However, accepting the least-squares solution that minimizes $\|\frac{1}{\alpha_k} s_{k-1} - y_{k-1}\|_2^2$, we obtain the so-called spectral step length:

$$\alpha_k = \frac{{s_{k-1}}^T s_{k-1}}{{s_{k-1}}^T y_{k-1}}. \tag{5}$$

By the Mean-Value Theorem of integral calculus, it follows that:

$$y_{k-1} = \left(\int_0^1 \nabla^2 f(x_{k-1} + t s_{k-1})\, dt \right) s_{k-1}.$$

Hence, α_k is the inverse of a Rayleigh quotient relative to the average Hessian matrix

$$\int_0^1 \nabla^2 f(x_{k-1} + t s_{k-1})\, dt,$$

and so it is bounded between the minimum and the maximum eigenvalue of the average Hessian, motivating the use of the word *spectral*.

The spectral step was originally proposed by Barzilai and Borwein [13], and further developed by Raydan et al. [14, 15]. Equation (5) is used to determine α_k, provided that for all iterations it is bounded away from zero on the positive side, and bounded away from infinity. As a safeguard, we define fixed parameters $0 < \alpha_{\min} < \alpha_{\max} < \infty$ and use at each iteration:

$$\alpha_k = \min\left\{ \alpha_{\max}, \max\left\{ \alpha_{\min}, \frac{{s_{k-1}}^T s_{k-1}}{{s_{k-1}}^T y_{k-1}} \right\} \right\}. \tag{6}$$

The SPG method is described in Algorithm 1. In every iteration, the method moves along the spectral projected direction with an initial step length α_k. To benefit from the use of the spectral step – which does not, in general, result in decrease in the objective – the method incorporates the nonmonotone globalization strategy of Grippo et al. [16], in which the line search parameter τ is made to satisfy:

$$f(x_k - \tau d_k) \leq \max_{0 \leq j \leq \min\{k, M\}} f(x^{(k-j)}) - \gamma \tau {d_k}^T g_k. \tag{7}$$

Here $M \geq 1$ is an integer specifying how many iterations to look back upon, and $0 < \gamma \ll 1$ is the sufficient decrease parameter for an Armijo-type criterion. The reduction of the line search parameter τ is based on a safeguarded quadratic interpolation (see [17] for more details).

2.3. **Hybrid method.** In the hybrid method of Guerrero et al. [7], simulated annealing (SA) [9] is used in conjunction with SPG to determine a suitable value of λ for the full regularization problem (2). In principle, one would of course like to find the value of λ which produces a regularized solution that is closest to the true solution; for example, in terms of the relative error,

$$RelErr(x) = \|x - x_{true}\|_2 \;/\; \|x_{true}\|_2. \tag{8}$$

Algorithm 1: Spectral Projected Gradient algorithm.

Given: x_0, $\alpha_0 \in [\alpha_{min}, \alpha_{max}]$, $maxiter$, tol, $\gamma \in (0,1)$ and $M \geq 1$.

$k = 0$;
$pgnorm = 1$;
$g_0 = \nabla f(x_0)$;
while $(pgnorm \geq tol)$ *and* $(k \leq maxiter)$ **do**
  $d_k = P_\Omega(x_k - \alpha_k g_k) - x_k$;
  $pgnorm = \| d_k \|_\infty$;
  $\tau = 1$;
  Set $x_+ = x_k + \tau d_k$;
  while $f(x^+) > \max\limits_{0 \leq j \leq \min\{k,M\}} f(x^{(k-j)}) - \gamma \tau {d_k}^T g_k$ **do**
    Reduce τ; $x_+ = x_k + \tau d_k$;
  end
  $x_{k+1} = x_+$;
  $s_k = x_{k+1} - x_k$;
  $g_{k+1} = \nabla f(x_{k+1})$;
  $y_k = g_{k+1} - g_k$;
  if ${s_k}^T y_k \leq 0$ **then**
    $\alpha_{k+1} = \alpha_{\max}$;
  else
    $\alpha_{k+1} = \min\left\{\alpha_{\max}, \max\left\{\alpha_{\min}, \dfrac{{s_k}^T s_k}{{s_k}^T y_k}\right\}\right\}$;
  end
  $k = k + 1$
end
return x_k, α_k, $pgnorm$

The relative error cannot be used to select λ, however, when the true solution x_{true} is not known. Thus one needs to select λ based on some other criterion. In some applications, methods such as the L-curve criterion, discrepancy principle, or generalized cross validation can be used [4, 5]; for example, if the level of noise in the data is known.

In [7], the authors observed that when λ is small, the SPG method is typically unable to converge, in the sense that the norm of the projected gradient direction (*pgnorm*) decreases very slowly (if at all) as the algorithm proceeds. This can be attributed to the poor conditioning of the Hessian matrix for the unregularized problem,

$$\nabla^2 f_{LSQ} = A^T A.$$

When the ℓ_2 norm, $\phi_2(x) = \frac{1}{2}\|x\|_2^2$, is used as a regularizer, the Hessian becomes

$$\nabla^2 f_{RLSQ} = A^T A + \lambda I,$$

whose condition number is $\kappa = (\sigma_1 + \lambda)/(\sigma_n + \lambda)$, where σ_1 and σ_n are the largest and smallest eigenvalues of $A^T A$. Since κ decreases as λ increases, SPG is eventually able to converge for some critical value of λ, as well as all larger values. Empirically, the authors observed that this critical value occurred close to the value of λ that minimized the relative error of the solution [7].

Based on this observation, the authors proposed a hybrid algorithm (Algorithm 2) in which the value of λ is selected based on monitoring the behaviour of *pgnorm* over several runs of SPG. The idea behind the algorithm is that every run of SPG terminates either when *pgnorm* reaches tol, or a maximum number of iterations is reached. The latter case indicates that the convergence criterion was not met, meaning that the problem is still poorly conditioned. The value of λ is then randomly perturbed and SPG is run again. The final solution and step size from the previous run are used as initialization parameters, in order to reduce the number of required SPG iterations.

If perturbing λ results in a *pgnorm* that is below tol, or less than the smallest *pgnorm* that has been attained so far, then the new value of λ is accepted. If not, it may still be accepted with some probability determined by the *Accept* function, according to the SA heuristic. The probability of acceptance in this case is given by

$$P = \exp\left(-\frac{\Delta E}{T_i}\right), \text{where } \Delta E = pgnorm_+ - pgnorm_i.$$

The variable *inside* keeps track of how many times (consecutively) the convergence criterion has been met. As *inside* increases, the size of the perturbations to λ shrink rapidly; additionally, the rate of reduction of T (controlled by the *Reduce* function) is accelerated (see [7] for details). The effect of these heuristics is that once a value of λ is attained that permits SPG to converge, the hybrid algorithm converges rapidly to a value of λ in its immediate neighbourhood.

2.4. **Simultaneous Algebraic Reconstruction Technique (SART).** The algebraic reconstruction technique (ART) for solving a linear system of equations was first proposed by Kaczmarz [18] and rediscovered several decades later in the context of CT imaging [19]. The ART iteration projects the current iterate towards the hyperplane defined by a single equation, using the formula

$$x_{k+1} = x_k - \omega_k \frac{\langle a_i, x_k \rangle - b_i}{\langle a_i, a_i \rangle} a_i^T, \tag{9}$$

where a_i is the ith row of A, b_i the ith element of b, $\langle \cdot \rangle$ denotes the inner product, and $\omega_k \in (0,2)$ is a relaxation parameter. The index, i, cycles sequentially over all rows of A in the classical version of the algorithm, but can also be chosen in other ways (see e.g. [20]) to accelerate convergence. The choice of $\omega_k = 1$ corresponds to the classical Kaczmarz method in which iterates are projected onto each hyperplane; taking $\omega_k < 1$ (underrelaxation) may reduce sensitivity to noise, while taking $\omega_k > 1$ (overrelaxation) may accelerate convergence if the data are consistent [21].

A weakness of ART in the context of CT imaging is its sensitivity to noise in the data [10]. To reduce this sensitivity, one can attempt to satisfy multiple equations simultaneously. The simultaneous algebraic reconstruction technique (SART) [10] is of the form

$$x_{k+1} = x_k - \omega_k D A^T M (A x_k - b), \tag{10}$$

Algorithm 2: Hybrid SPG algorithm for full regularization problem.

Given: x_0, $\alpha_0 \in [\alpha_{min}, \alpha_{max}]$, $maxiter$, tol, $\lambda_1 > 0$, $0 < \varepsilon < 1$, $T_1 > T_{min}$

$(x_1, \alpha_1, pgnorm_1) = SPG(x_0, \alpha_0; \lambda_1)$;
$i = 1$, $inside = 0$;
while $T_i > T_{min}$ **do**
  $\theta = rand(-1, 1)$;
  $\lambda_+ = \left|\lambda_i + \theta\left(\varepsilon/10^{inside}\right)\right|$;
  $(x_+, \alpha_+, pgnorm_+) = SPG(x_i, \alpha_i; \lambda_+)$;
  if $pgnorm_+ < tol$ **then**
    $inside = inside + 1$;
    $pgnorm_{i+1} = pgnorm_+$, $x_{i+1} = x_+$, $\lambda_{i+1} = \lambda_+$, $\alpha_{i+1} = \alpha_+$;
  else if $pgnorm_+ < pgnorm_i$ **then**
    $pgnorm_{i+1} = pgnorm_+$, $x_{i+1} = x_+$, $\lambda_{i+1} = \lambda_+$, $\alpha_{i+1} = \alpha_+$;
  else
    $inside = 0$;
    if *Accept*$(pgnorm_+, pgnorm_i, T_i)$ **then**
      $pgnorm_{i+1} = pgnorm_+$, $x_{i+1} = x_+$, $\lambda_{i+1} = \lambda_+$, $\alpha_{i+1} = \alpha_+$
    else
      $pgnorm_{i+1} = pgnorm_i$, $x_{i+1} = x_i$, $\lambda_{i+1} = \lambda_i$, $\alpha_{i+1} = \alpha_i$
    end
    $T_{i+1} = Reduce(T_i)$;
    $i = i + 1$;
  end
end
return x_i, λ_i

where D and M are diagonal matrices, with

$$D = \text{diag}\left\{1 \Big/ \sum_{k=1}^{m} |a_{kj}|,\ j = 1 \ldots n\right\}$$
$$M = \text{diag}\left\{1 \Big/ \sum_{k=1}^{n} |a_{ik}|,\ i = 1 \ldots m\right\}. \tag{11}$$

The SART algorithm converges to a weighted least squares solution of the linear system [22, 23]. To accelerate convergence, it is also possible to use a block-iterative approach [22], dividing measurements into subsets and cycling over the subsets sequentially. In this paper we only consider the case where all measurements are processed simultaneously[1]. We also use the constant value $\omega_k = 1.9/\rho(DA^TMA) \approx 1.9$ for the relaxation parameter (ρ denotes the spectral radius of the matrix, which is only slightly less than 1). This is the default choice of constant ω_k in the AIRTools software package used for our implementation of SART [24]; more details on this and other choices of ω_k can be found in [25].

[1]This case is sometimes referred to as SIRT (simultaneous iterative reconstruction technique), see e.g. [1]

The computational complexity for one iteration of SART is dominated by the multiplication by A and A^T. In comparison, one iteration of SPG applied to either (1) or (2) also requires at least one multiplication by A to compute $f_{LSQ}(x_k)$, and one multiplication by A^T when computing the gradient, $\nabla f_{LSQ}(x_k) = A^T(Ax_k - b)$. The iteration may also require additional multiplications by A to evaluate f_{LSQ} if backtracking is required during the line search. We found that this was rare in practice, however, since the nonmonotone globalization condition (7) does not require that the function decrease in every iteration. Thus it is reasonable to consider the cost per iteration of SART and SPG to be roughly equivalent.

2.5. **Superiorization Methodology.** Superiorization [6] is a recently proposed optimization heuristic in which the solution generated by an iterative algorithm such as SART is perturbed between iterations to produce a solution that is superior with respect to the secondary objective, $\phi(x)$. Specifically, if the update formula for the original iterative method is of the form $x_{k+1} = R(x_k)$, the superiorized algorithm takes the form

$$x_{k+1} = R(x_k + \beta_k v_k) \tag{12}$$

where $\{\beta_k\}_{k=0}^{\infty}$ is a summable sequence of non-negative real numbers, and $\{v_k\}_{k=0}^{\infty}$ is a sequence of bounded perturbation vectors. Typically, one chooses $\beta_k = \gamma^k$, $0 < \gamma < 1$, and v_k to be a nonascending direction of ϕ at x_k, e.g., $v_k = -\nabla\phi(x_k)/\|\nabla\phi(x_k)\|_2$.

The foundation of superiorization is the key result that if the original algorithm is *perturbation resilient*, then perturbations of the type described above can be introduced in every iteration without jeopardising convergence of the algorithm. More precisely, if the original algorithm produces a solution x^* such that the residual norm,

$$\|Ax^* - b\|_2 \leq \varepsilon \tag{13}$$

for some $\varepsilon \geq 0$ (defined as an ε-compatible solution), then the superiorized version of the algorithm eventually converges to a ε-compatible solution x^{sup}, which can be expected to satisfy $\phi(x^{sup}) < \phi(x^*)$, if the perturbations are nonascending directions of ϕ. SART is known to be perturbation resilient [26]; some sufficient conditions are discussed in [6]. More details on the methodology can be found in a recent special edition of *Inverse Problems* [27] and a continously updated online bibliography[2].

A superiorized version of SART is shown in Algorithm 3. The inputs to the algorithm are an initial guess x_0, a stopping tolerance ε, and two parameters N and γ. The form of the algorithm is slightly different than described in (12), as we allow for a total of N perturbations to be applied between every iteration of SART, to more effectively reduce the value of $\phi(x)$ between iterations. Mathematically, these are equivalent to a single bounded perturbation, but this perturbation cannot be determined *a priori*. Typically, computing the perturbations is inexpensive compared to the cost of multiplication by A and A^T, so the per-iteration cost of the superiorized algorithm is not significantly worse that of the basic version.

The parameter γ controls the rate at which the perturbation size decreases. The values of γ and N are chosen empirically and may significantly affect the performance of the algorithm [28]. Small values of γ cause the size of the perturbations to decrease rapidly, such that the perturbations may have little effect after the first few iterations, while a value of γ close to one results in larger perturbations that may delay convergence to an ε-compatible solution. Similarly, a large value of N will result in more rapid reduction

[2] http://math.haifa.ac.il/YAIR/bib-superiorization-censor.html

in perturbation size between iterations. The stopping criterion, ε, is also chosen empirically, depending on the application.

Algorithm 3: Superiorized SART Algorithm.

Given: x_0, $\gamma \in (0,1)$, N, $\varepsilon > 0$

$\ell = -1$;
$k = 0$;
while $\|Ax_k - b\|_2 \geq \varepsilon$ **do**
  $n = 0$;
  $x_{k,n} = x_k$;
  while $n < N$ **do**
    $v_{k,n} = -\nabla\phi(x_{k,n}) \;/\; \|\nabla\phi(x_{k,n})\|_2$;
    while true do
      $\ell = \ell + 1$;
      $\beta_{k,n} = \gamma^{\ell}$;
      $z = x_{k,n} + \beta_{k,n} v_{k,n}$;
      if $z \in \Omega$ *and* $\phi(z) \leq \phi(x_k)$ **then**
        $x_{k,n+1} = z$;
        break;
      end
    end
    $n = n + 1$;
  end
  $x_{k+1} = P_\Omega\left(x_{k,N} - \omega_k D A^T M\left[Ax_{k,N} - b\right]\right)$;
  $k = k + 1$;
end
return x_k

2.6. **Total variation and Huber penalties.** Two commonly-used regularization functions in image reconstruction are total variation [2] (TV) and the Huber penalty [3]. Both are intended to reduce noise in the image while preserving edges. For a 2-D image, the isotropic form of the TV function is

$$\phi_{TV}(x) = \sum_{(m,n)} \sqrt{\left(x_{(m+1,n)} - x_{(m,n)}\right)^2 + \left(x_{(m,n+1)} - x_{(m,n)}\right)^2 + \delta^2}, \tag{14}$$

where the image vector, x, has been reshaped into a 2-D array indexed by m and n. This function arises from summing the magnitude of a discrete gradient approximation over the entire image. A noisy image is characterized by many small fluctuations in the gradient approximation, resulting in a large TV value. The small parameter δ (which is sometimes omitted) ensures that the function is differentiable at points $x_{(m,n)}$ where the neighbouring pixels have the same value. The TV penalty has attracted a great deal of attention recently as it can be represented as the 1-norm of a sparsifying transform, making it a natural choice in many compressive sensing applications [29].

The Huber penalty has the form

$$\phi_{Huber}(x) = \sum_{(m,n)} \sum_{(p,q)\in\mathcal{N}_{(m,n)}} \psi\left(x_{(p,q)} - x_{(m,n)}\right), \tag{15}$$

where

$$\psi(z) = \begin{cases} \frac{z^2}{2\delta}, & |z| < \delta \\ |z| - \frac{\delta}{2}, & |z| \geq \delta \end{cases}. \tag{16}$$

Here, $\mathcal{N}_{m,n}$ is some neighbourhood of the (m,n)th pixel, and δ is a parameter discriminating between "small" variations in the image, which are subject to a quadratic penalty, and large deviations, which are penalized linearly. In this work, we choose $\mathcal{N}_{m,n} = \{(m+1,n),(m,n+1)\}$ to take into account the same pixel-wise differences as the TV function (14). In the limit as $\delta \to 0$, this gives a function that converges to the anistropic form of the TV [30],

$$\phi(x) = \sum_{(m,n)} \left|x_{(m+1,n)} - x_{(m,n)}\right| + \left|x_{(m,n+1)} - x_{(m,n)}\right|.$$

3. Numerical Experiments

In this section we compare the two approaches in several numerical experiments. Code for SPG and the hybrid algorithm were provided to us by the authors of [7]. We used the implementation of SART from the AIRTools II software package [24], and modified it to implement the superiorized version.

3.1. Simulation data.

We generated two datasets for use in comparing the two approaches. The first dataset is based on a 256×256 pixel image of the Shepp-Logan (SL) phantom (Figure 1), with a pixel size of 1.2 mm. We then simulated 180 parallel-beam views with a 1° increment using AIRTools II. Each view consisted of 362 line integrals,[3] resulting in a system matrix A of size $65,160 \times 65,536$. Poisson noise was added to the raw projection data proportional to initial count rates of $I_0 = 1.0\times10^4$, 2.5×10^4, 5.0×10^4, and 1.0×10^5 (see (3)), to assess the effect of noise on the effectiveness of the two approaches.

In the second set of experiments, we simulated data based on four slices of real CT images obtained from the Cancer Imaging Archive [31], representing slices through the abdomen, thorax, shoulder and head regions. The images were downsampled from 512×512 to 256×256 pixels, and the same pixel size and number of views were simulated as for the experiments with the SL phantom. For this set of experiments, only an initial count rate of 5.0×10^4 was simulated; this results in a lower signal to noise ratio for large or dense objects, such as the abdomen and shoulder images. The anatomical CT images are shown in Figure 2.

3.2. Unpenalized reconstructions.

To compare the baseline performance of the two underlying algorithms, we applied both SPG and SART to the reconstruction problem with no penalty functions. We chose the following stopping criterion for the two algorithms:

- SPG was run for a maximum of 250 function evaluations, or until $pgnorm < 10^{-6}$.
- SART was run until the residual norm, $\|Ax_k - b\|_2$, decreased by less than 0.25% in one iteration.

[3] This is equal to $\lfloor\sqrt{2}N\rfloor$ for $N = 256$, which ensures that every pixel is intersected by at least one line for every view.

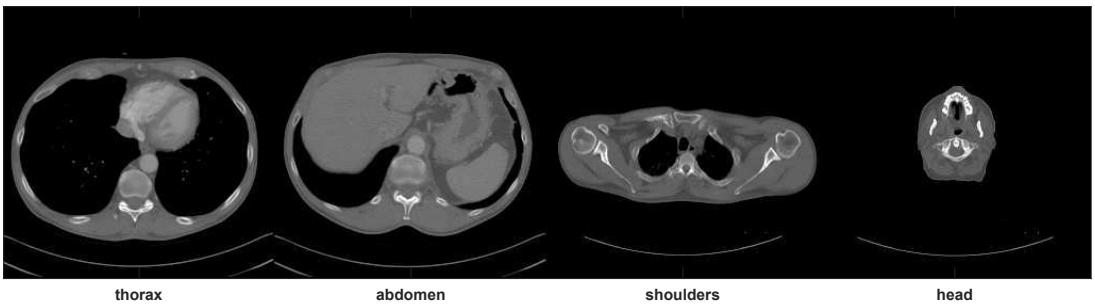

FIGURE 2. 256 x 256 pixel anatomical phantoms.

The stopping criterion for SART was chosen heuristically after some numerical experimentation. It is well known that both ART and SART exhibit semiconvergence [25, 32] when applied to problems with noisy data; i.e., image quality begins deteriorating after some number of iterations, even as the residual continues to decrease. This behaviour can be analyzed by bounding the error in the kth iterate as

$$\|x_k - \bar{x}\| \leq \|x_k - \bar{x}_k\| + \|\bar{x}_k - \bar{x}\|, \tag{17}$$

where $\bar{x}$ represents the weighted least-squares solution to the problem $Ax = \bar{b}$ with noiseless data $\bar{b}$, and x_k and $\bar{x}_k$ denote the iterates produced by SART (10) with noisy and noiseless data, respectively. In [25], the authors call the first term on the right of (17) the "noise error" and the second term the "iteration error", and show that the noise error grows with k, eventually dwarfing the iteration error.

For this reason, SART is typically stopped "early" after some number of iterations have been run; in general, the algorithm should be stopped earlier for noisier data. Alternatively, if the level of noise in the data is known, a stopping criterion based on the residual norm (13) can be applied, e.g., using the discrepancy principle [33]. We elected to use a rule based on the relative change in the residual, so that it could be applied consistently across all our test cases with varying noise levels.

The images produced by each algorithm for two test cases, as well as plots of the convergence behaviour, are shown in Figure 3. The nonmonotic convergence behaviour of SPG is apparent, as well as the semiconvergence phenomenon in both algorithms. The relative error (8), of the best solution found by each algorithm was roughly the same, but SPG converges more rapidly and exhibits semiconvergence at earlier iterations. This behaviour was typical of all eight experiments, as indicated in Table 1.

In every case, SPG terminated after 250 function evaluations; as discussed in [7], the ill-conditioning of the unregularized problem makes it impossible for the stopping criterion based on *pgnorm* to be attained. Table 1 indicates that the stopping criterion chosen for SART, on the other hand, terminated the algorithm prematurely (while the error was still decreasing) in all but one case.

3.3. **SPG with fixed regularization parameter.** Before applying the hybrid algorithm to the reconstruction problem, we first ran experiments to determine the optimal value of the regularization parameter λ for each test case. Specifically, we applied SPG to the regularized problem (2) for values of λ between 0 and 1, in increments of 0.02. We used both ϕ_{TV} (14) and ϕ_{Huber} (15) as regularization functions, with $\delta = 10^{-6}$ for ϕ_{TV} and $\delta = 10^{-3}$ for ϕ_{Huber}. For each value of λ, SPG was run for a maximum of 1000 iterations, or until $pgnorm < 10^{-6}$. We then computed the relative error to determine the optimal value of λ. While this is not possible to do in practice when the true image is not known, knowing the

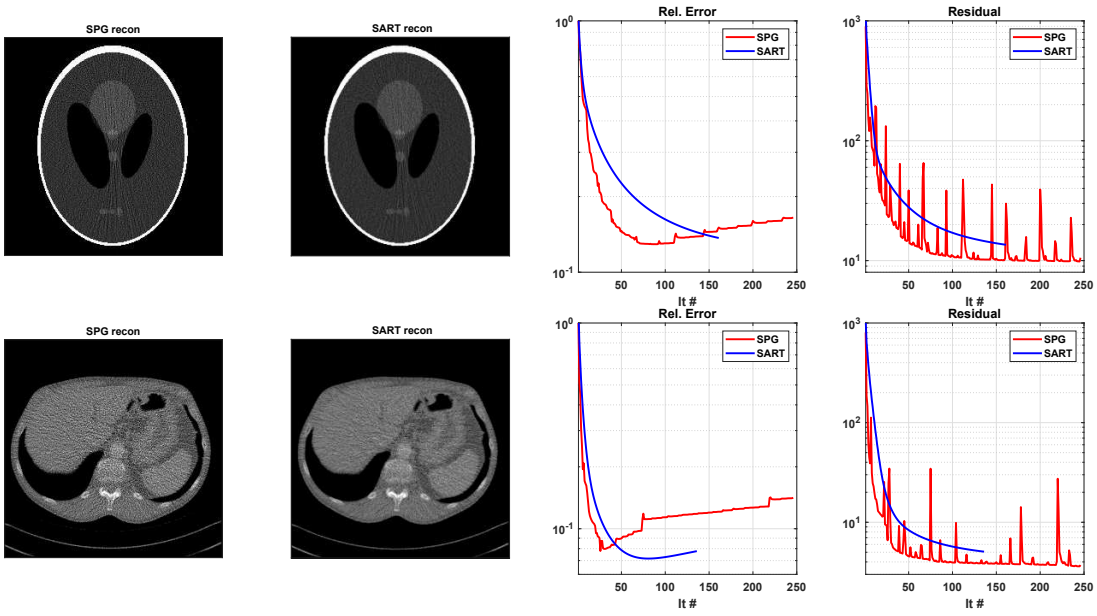

FIGURE 3. Results of SPG and SART reconstructions with no penalty. First row: Shepp-Logan phantom with $I_0 = 2.5 \times 10^4$; second row: abdomen phantom. Semilog plots show the relative error and residual norm as a function of the number of iterations.

TABLE 1. Results of SPG and SART reconstructions with no penalty. Relative error, number of iterations, and residual norm value (ε) are shown for each algorithm; first when the stopping criterion was reached (end) and then at the optimal error attained (opt). Dashed lines indicate that the two values are the same, i.e. the algorithm terminated while the error was still decreasing.

Experiment	SPG (end)			SPG (opt)			SART (end)			SART (opt)		
	Relerr	Its	ε	Relerr	Its	ε	Relerr	Its	ε	Relerr	Its	ε
SL 1.0e4	0.300	243	16.4	0.175	50	20.2	0.170	123	21.5	–	–	–
SL 2.5e4	0.165	246	10.6	0.129	90	11.0	0.137	160	13.5	–	–	–
SL 5.0e4	0.125	242	6.90	0.104	79	7.51	0.118	194	9.92	–	–	–
SL 1.0e5	0.088	243	4.83	0.081	153	5.10	0.101	235	7.28	–	–	–
Thorax	0.085	248	1.85	0.064	49	2.55	0.064	180	2.48	–	–	–
Abdomen	0.141	246	5.72	0.078	26	6.59	0.078	135	5.07	0.072	81	6.29
Shoulders	0.089	250	2.05	0.068	50	2.66	0.067	173	2.70	–	–	–
Head	0.071	240	1.17	0.063	119	1.27	0.072	226	1.61	–	–	–

optimal value of λ for each test case allowed us to assess the performance of the hybrid algorithm in the subsequent section.

The reconstructed images for two of the eight test cases are shown in Figure 4, along with plots showing the final relative error and *pgnorm* as a function of the parameter λ. From these plots, it is apparent that using ϕ_{TV} with $\delta = 10^{-6}$ does not allow the algorithm to converge (based on the size of *pgnorm*), for any value of λ. When the Huber penalty with $\delta = 10^{-3}$ is used, the algorithm is able to converge to the desired tolerance for some range of λ values, though not for all λ larger than some

TABLE 2. Results of SPG algorithm with fixed regularization parameter λ, using the TV and Huber penalties. Value of λ corresponding to the best solution found, the relative error of that solution, number of iterations required, and residual are shown.

Experiment	TV				Huber			
	λ	Relerr	Its	ε	λ	Relerr	Its	ε
SL 1.0e4	0.10	0.127	1000	20.6	0.44	0.052	1000	24.6
SL 2.5e4	0.06	0.089	998	12.4	0.28	0.033	1000	15.1
SL 5.0e4	0.04	0.068	999	8.57	0.20	0.026	903	10.6
SL 1.0e5	0.02	0.048	998	5.55	0.12	0.017	681	7.29
Thorax	0.02	0.051	1000	2.72	0.02	0.052	799	2.78
Abdomen	0.06	0.058	1000	6.78	0.10	0.055	733	7.62
Shoulders	0.02	0.056	1000	2.93	0.04	0.051	600	3.36
Head	0.02	0.070	999	2.12	0.02	0.044	572	1.70

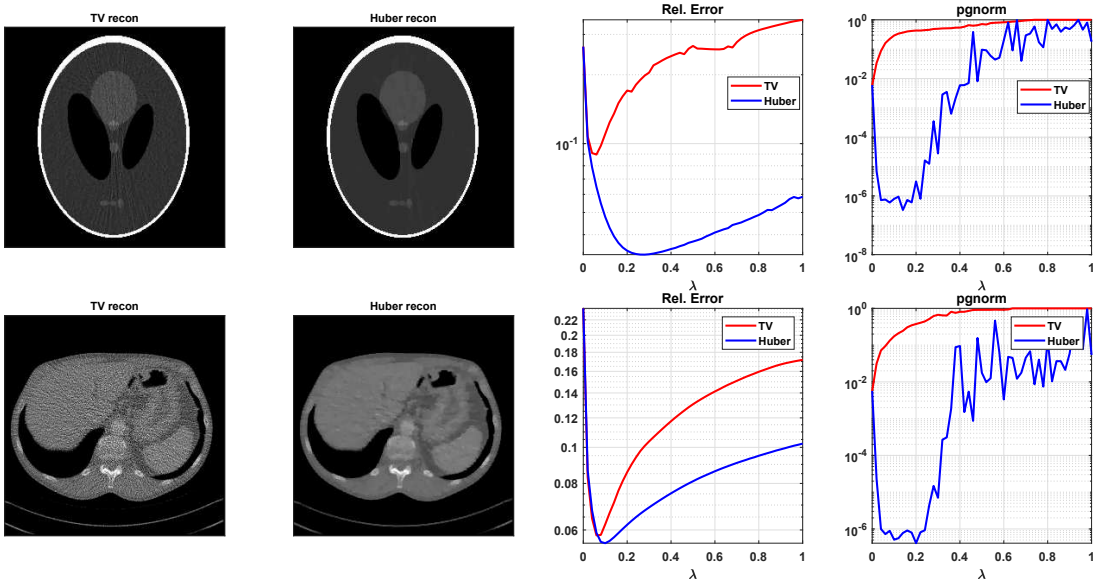

FIGURE 4. Results of SPG reconstructions with fixed regularization parameter λ, using TV and Huber prior as penalty functions. First row: Shepp-Logan phantom with $I_0 = 2.5 \times 10^4$; second row: abdomen phantom. Semilog plots show the final relative error and pgnorm attained by SPG as a function of λ.

critical value (which is the case when one uses the 2-norm). We discuss this phenomenon further in Section 4.

The results of all eight test cases are summarized in Table 2. We note that when the Huber prior was used, in six out of eight cases the optimal value of λ was one for which the algorithm was eventually able to converge, before reaching the maximum number of iterations. For the Shepp-Logan phantom, the optimal value of λ was larger for noisier data, as expected. For the anatomical phantoms, the optimal value of λ tended to be smaller than for the Shepp-Logan phantom. This is also intuitively sensible, as the anatomical images contain more fine detail than the Shepp-Logan phantom, making them more sensitive to over-regularization.

TABLE 3. Results of hybrid SPG algorithm using Huber penalty with two values of ε. Median values for final value of λ found by the algorithm, relative error, and total number of SPG iterations are shown.

Experiment	$\varepsilon = 0.05$			$\varepsilon = 0.2$		
	λ	Relerr	Its	λ	Relerr	Its
SL 1.0e4	0.066	0.127	4352	0.144	0.083	4612
SL 2.5e4	0.047	0.073	3190	0.128	0.041	2741
SL 5.0e4	0.036	0.050	2572	0.096	0.030	2485
SL 1.0e5	0.020	0.039	2754	0.068	0.020	2125
Lungs	0.021	0.053	2757	0.043	0.053	2912
Abdomen	0.039	0.068	2766	0.136	0.060	2892
Shoulders	0.047	0.052	2441	0.112	0.066	2439
Head	0.035	0.053	2071	0.174	0.114	2437

3.4. **Hybrid algorithm.** In light of the results of the previous section, we determined that the TV function with $\delta = 10^{-6}$ could not be used as a regularizer in conjunction with Algorithm 2, which requires that the convergence criterion based on *pgnorm* be attained near the optimal value of λ. We therefore considered only the Huber function with $\delta = 10^{-3}$. As with the experiments in the previous section, every iteration of SPG was run until $pgnorm \leq 10^{-6}$, up to a maximum of 1000 iterations; we used the default values of $T_1 = 40$ and $T_{min} = 10^{-1}$ from [7] as the parameters controlling the simulated annealing (SA) iterations.

Algorithm 2 includes a parameter ε controlling the size of the perturbation to λ between iterations of SPG. In [7], the authors tested several different different choices of ε ranging from 0.0005 to 0.05, and concluded that a larger value of ε is preferred when the data are noisier. In our experiments we compared results using $\varepsilon = 0.05$ and $\varepsilon = 0.2$. Since the SA component of the hybrid algorithm introduces a stochastic element to the optimization, we also ran the algorithm five times for each test dataset, to assess the average performance of the algorithm.

The results of the experiment are shown in Table 3 and Figure 5. The choice of ε has a significant effect on the value of λ to which the algorithm converges; with $\varepsilon = 0.2$, the final λ values were much larger than with $\varepsilon = 0.05$. This corresponded to improved image quality for the Shepp-Logan phantom and two of the anatomical phantoms, but not the shoulder and head phantoms. Even with $\varepsilon = 0.2$, however, the algorithm tended to underestimate the correct value of λ for the Shepp-Logan phantom experiments. While individual runs of the hybrid algorithm were able to obtain nearly optimal error in most of the test cases – at least for one of the two chosen values of ε – there was considerable variance in the final values of λ attained by the algorithm across different runs, resulting in inconsistent performance.

3.5. **Superiorized reconstructions.** The superiorized SART algorithm (Algorithm 3) was run on each of the eight test problems. In all cases, the algorithm was run until the residual norm (ε) was less than the value attained by the unsuperiorized SART algorithm (shown in the tenth column of Table 1). This choice was made to confirm that the superiorized algorithm was able to attain an ε value equal to the basic algorithm in all test cases. Based on numerical experimentation, we chose $N = 5$ and $\gamma = 0.9995$ as parameters to the algorithm, as this choice consistently gave the best results in a reasonable number of iterations. The same choices of δ as in the previous section were used for ϕ_{TV} and ϕ_{Huber}.

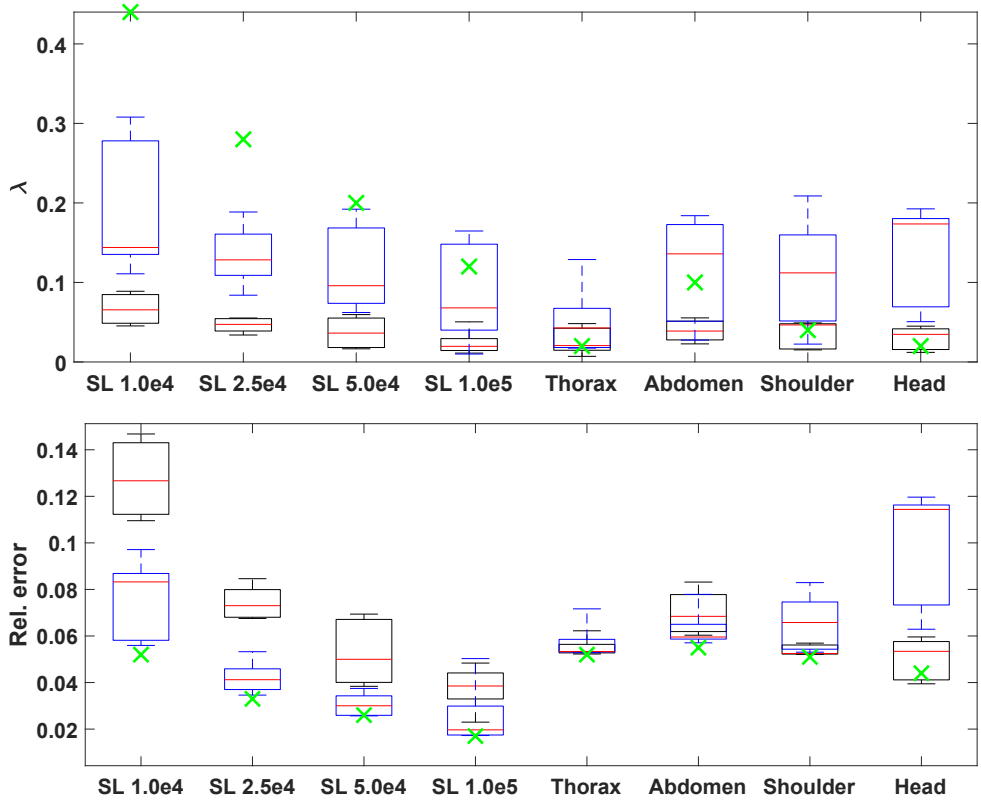

FIGURE 5. Box plots of final λ values (top) and relative errors (bottom) attained over five runs of the hybrid algorithm for each test case. Black boxes correspond to $\varepsilon = 0.05$, blue boxes to $\varepsilon = 0.2$. Red line indicates median value, box edges represent 25th and 75th percentiles, whiskers indicate most extreme datapoints. Optimal values (tabulated in Table 2) are indicated by green crosses.

Representative reconstructed images and convergence plots are shown in Figure 6. The most notable difference caused by the choice of penalty function is in the convergence behaviour. In particular, when the Huber prior is used as the penalty, the algorithm stagnates after some number of iterations, making little progress for several hundred iterations. Eventually the algorithm is able to make further progress and reach an error comparable to (or better than) that obtained when using TV as the penalty. Some investigation revealed that during those iterations where the algorithm stagnates, the Huber penalty appears to be oversmoothing the image, possibly because the perturbations are too large. Once the perturbation size becomes sufficiently small, the SART iteration is able to make progress once again. We provide an illustration in Figure 7. After 600 iterations (first image), the phantom is very blurry, and the image quality only begins to improve once the perturbation size is on the order of 10^{-2} (second image). The optimal image quality is obtained close to 1200 iterations (third image), after which point the image quality degrades due to semiconvergence (fourth image).

It appears that this behaviour is more a result of the choice of the small parameter δ in ϕ_{TV} and ϕ_{Huber}, rather than the functions themselves. Specifically, if the same values of δ are used in each, the performance of the superiorized algorithm is comparable whether ϕ_{TV} or ϕ_{Huber}, though we obtained slightly better results with ϕ_{Huber}. Additionally, when taking $\delta = 10^{-3}$, we found that reducing the initial

TABLE 4. Results of superiorized SART reconstructions with TV and Huber penalties. Relative error, number of iterations, and residual norm value are shown for each algorithm when the stopping criterion was reached (end) and at the optimal error attained (opt). Dashed lines indicate that the optimal error (to that point) was attained at the same point as the stopping criterion.

Experiment	TV (end)			TV (opt)			Huber (end)			Huber (opt)		
	Relerr	Its	ε	Relerr	Its	ε	Relerr	Its	ε	Relerr	Its	ε
SL 1.0e4	0.088	688	21.5	0.077	453	23.3	0.081	1117	21.5	0.053	837	24.5
SL 2.5e4	0.053	619	13.5	0.053	559	13.9	0.043	1133	13.5	0.034	966	14.8
SL 5.0e4	0.041	583	9.92	–	–	–	0.029	1162	9.92	0.027	1075	10.4
SL 1.0e5	0.033	551	7.28	–	–	–	0.019	1177	7.28	–	–	–
Thorax	0.052	1135	2.48	0.047	837	3.11	0.054	1692	2.48	0.051	1488	2.94
Abdomen	0.075	1219	5.07	0.048	625	7.71	0.076	1567	5.07	0.053	1157	7.40
Shoulders	0.054	811	2.70	0.050	533	3.21	0.055	1528	2.70	0.052	1365	3.09
Head	0.043	847	1.61	–	–	–	0.046	1630	1.61	–	–	–

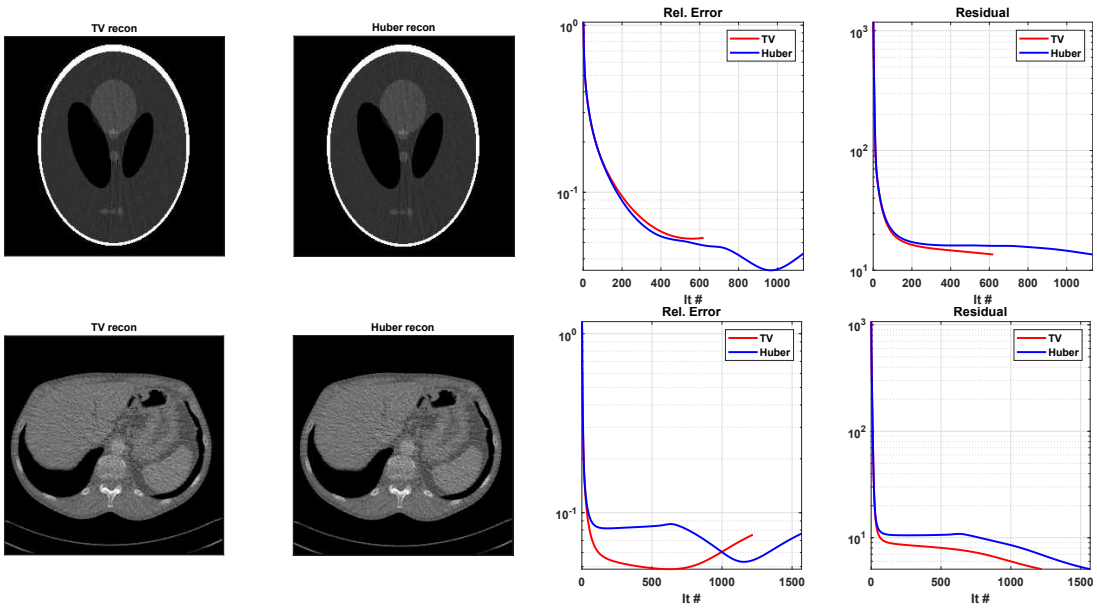

FIGURE 6. Results of superiorized SART reconstructions using TV and Huber prior as penalty functions. First row: Shepp-Logan phantom with $I_0 = 2.5 \times 10^4$; second row: abdomen phantom. Semilog plots show the relative error and residual norm as a function of the number of iterations.

perturbation size – that is, taking $\beta_{k,n} = \beta_0 \gamma^\ell$ for some $\beta_0 < 1$ in Algorithm 3 – eliminated the stagnating behaviour and allowed the algorithm to converge more quickly. This change also resulted in higher relative error overall, however, and so we have only shown the results with the original formulation of the algorithm.

The results of all eight test cases are summarized in Table 4. Overall, the image quality obtained using either penalty function (as measured by relative error) is comparable, with some exceptions. In particular, using ϕ_{Huber} with $\delta = 10^{-3}$ gave better results for the Shepp-Logan phantom, albeit after

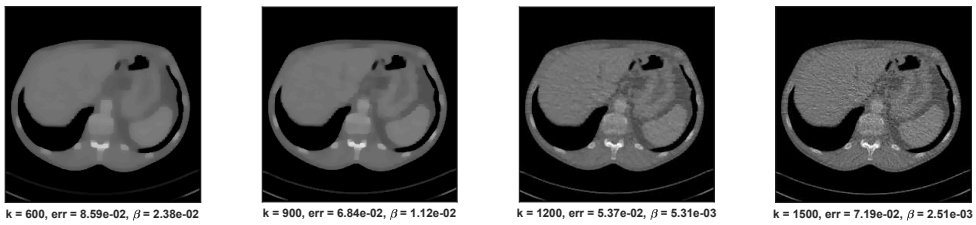

FIGURE 7. Figures showing the solution x_k for increasing k values in the abdomen phantom experiment, as well as the relative error and perturbation size, β. Compare with the relative error plot in Figure 6.

many more iterations. Compared to the unpenalized SART reconstructions (Table 1), the final relative error obtained by the superiorized method was considerably better in every case, with the exception of the abdomen phantom. We note that for this phantom, however, the *best* error obtained by the algorithm (4.8% using TV, 5.3% using Huber) was still much better than that of unpenalized reconstruction (7.2%); unfortunately, the algorithm did not terminate until the image quality had been significantly worsened due to semiconvergence, as shown in Figure 6.

4. DISCUSSION

The results of our numerical experiments provide some useful data in comparing the superiorized SART and hybrid SPG algorithms. Broadly, the regularized and superiorized approaches appear to have the *potential* to produce solutions of roughly the same quality. In Table 5 we have aggregated some of the results from the previous section; the table shows the overall lowest relative error achieved by each approach. This corresponds to the minimum error achieved by SART, superiorized SART, and SPG across the entire iteration (i.e. not just when the stopping criterion was attained) and the solution corresponding to the optimal value of λ for when SPG was applied with a fixed regularization parameter. We can observe that the best relative error attained by superiorized SART and SPG is virtually the same for all eight test cases. Both algorithms also substantially improve the image quality compared to the corresponding unpenalized algorithms (SART and SPG applied to (1)). For the Shepp-Logan phantom, the error is reduced drastically (by 65–80%); for the anatomical phantoms, the improvement was smaller (roughly 20–30%). This is to be expected as the Shepp-Logan phantom is piecewise constant, making it an ideal case for penalty functions which act on differences between neighbouring pixels.

The main challenge in applying each approach is determining appropriate parameters to achieve nearly optimal error. In the case of superiorization (Algorithm 3), the main parameters to determine are γ and N, which control the size and number of perturbations applied in every iteration, and ε, which determines when the algorithm stops. In this work we held the first two parameters fixed at $N = 5$ and $\gamma = 0.9995$. While it is possible that our results could be improved upon with different choices for these two parameters, in practice it appears difficult to determine optimal values without extensive experimentation. A recent paper [28] suggests taking $3 \leq N \leq 6$ and $\gamma = 0.75$ for a problem in proton computed tomography[4], but our own experiments indicated that it was necessary to take γ much closer to 1.0 to obtain good results in our experiments. Values as large as $\gamma = 0.99995$ have been used other studies [6].

[4]The parameter γ is denoted by α in the referenced paper.

TABLE 5. Comparison of best relative errors achieved by the unpenalized SART, superiorized SART, unregularized SPG, and regularized SPG algorithms. Results for the superiorized SART and regularized SPG were obtained using ϕ_{Huber}.

Experiment	SART		Sup. SART		Unreg. SPG		Reg. SPG	
	Relerr	Its	Relerr	Its	Relerr	Its	Relerr	Its
SL 1.0e4	0.170	123	0.053	835	0.175	50	0.052	1000
SL 2.5e4	0.137	160	0.034	966	0.129	90	0.033	1000
SL 5.0e4	0.118	194	0.027	1075	0.104	79	0.026	903
SL 1.0e5	0.101	235	0.019	1176	0.081	153	0.017	681
Thorax	0.064	180	0.051	1474	0.064	49	0.052	799
Abdomen	0.072	81	0.053	1157	0.078	26	0.055	733
Shoulder	0.067	173	0.053	1338	0.068	50	0.051	600
Head	0.072	226	0.046	1630	0.063	119	0.044	572

The choice of parameter ε used to define the stopping criterion (13) for superiorization is also challenging. Since the size of the perturbations eventually go to zero (at which point one is essentially applying SART), the superiorized SART algorithm is also susceptible to semiconvergence, as is apparent from the error plots in Figure 6. If the chosen value of ε is too large, the algorithm terminates before the optimal error is attained, while if ε is too small, image quality will begin to degrade due to semiconvergence before the stopping criterion is reached. Our experiments indicate that the best choice of ε is problem-specific and may not be straightforward to determine, even with a consistently applied stopping rule. The rule that we applied – halting the unsuperiorized algorithm when the change in residual between iterations became relatively small – caused the superiorized algorithm to halt prematurely in some cases – particularly those where the data were less noisy – but after the point of optimal error in cases where the data were noisier. In some of these cases, the error was significantly larger when the algorithm was halted compared to its optimal value – for instance, from 5.3% to 8.1% for the Shepp-Logan phantom at the lowest count rate (Table 4).

It should be noted that no claim has been made in the literature that the superiorization methodology produces a solution that is *optimal* in any sense; rather, it is true that the superiorized version of a perturbation resilient algorithm produces solutions that are equally constraints compatible, and likely to be superior with respect to the penalty function. This was the case in every one of our experiments. Whether this results in significantly improved image quality depends on the choice of ϕ and the other algorithmic parameters. We note that while the superiorized algorithm is eventually able to attain the same value of ε as the basic algorithm, it may take many more iterations to do so, as highlighted in Table 5.

For SPG applied to the regularized problem, the key parameter to be determined is the regularization parameter λ. The hybrid algorithm (Algorithm 2) proposes to determine λ by using the value of the *pgnorm* when the algorithm terminates as a guideline; the smallest value of λ for which the *pgnorm* attains the specified stopping tolerance is assumed to be close to the optimal value. Our numerical experiments indicate, however,that this claim not appear to hold, in general, for the regularized problem (2). In Figure 4, for example, the optimal value of λ for the Shepp-Logan phantom with $I_0 = 2.5 \times 10^4$ is $\lambda = 0.28$, but the algorithm converges to a tolerance of $pgnorm < 10^{-6}$ for a range of λ values from 0.04 to 0.20.

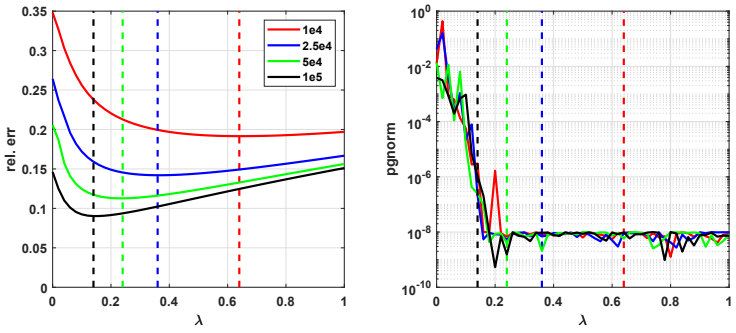

FIGURE 8. Plots of relative error and final *pgnorm* value for the Shepp-Logan phantom experiments, using $\phi_2(x) = \frac{1}{2}\|x\|_2^2$. Vertical dashed lines indicate λ value corresponding to optimal error.

In [7] this observation was made specifically for an image restoration problem where A represented a blurring operation, with $\phi_2(x) = \frac{1}{2}\|x\|_2^2$, and Gaussian noise was added to the blurred image. In our experiments, we have used a different matrix A (corresponding to a CT system matrix) as well as different regularization functions $\phi(x)$ and noise model. Even if $\phi_2(x)$ is used as a regularizer, however, we find that the *pgnorm* is not a reliable indicator of the optimal value of λ for the CT reconstruction problem. In Figure 8 we show the final relative error and *pgnorm* attained as a function of λ for the four Shepp-Logan phantom experiments, using $\phi_2(x)$. We can see that the optimal value of λ varies significantly with noise, but that the SPG algorithm is first able to converge for roughly the same value, $\lambda \approx 0.2$, across all experiments. This makes sense intuitively, because the failure of SPG to converge for small λ depends primarily on the poor conditioning of the Hessian, which is independent of the level of noise in the data.

The authors of [7] suggest that the parameter ε that controls the size of the perturbation to λ in every iteration Algorithm 2 can be increased for noisier data to improve the results. In our own experiments we did find that setting $\varepsilon = 0.2$ provided better results than $\varepsilon = 0.05$ (Figure 5, Table 3) in many cases, particularly for noisier data. That said, increasing the value of ε also produces less consistent results, as the variance in value of λ to which the algorithm converges also increases. In effect, tuning the value of the parameter ε to improve performance does not appear to be any easier than determining the value of λ empirically, with the additional drawback that the hybrid method introduces a random component to the algorithm.

A second issue is that the hybrid method does not generalize well to the total variation and Huber penalty functions considered in this paper. As discussed in Section 2.3, when $\phi_2(x)$ is used as a regularizer, the conditioning of the Hessian $\nabla^2 f_{RLSQ} = A^T A + \lambda I$ improves monotonically with λ. For a general penalty function $\phi(x)$, we have $\nabla^2 f = A^T A + \lambda \nabla^2 \phi$, and thus the conditioning of the problem depends on the Hessian of ϕ as well. Using the final value of *pgnorm* attained by SPG as a proxy for the conditioning of the problem, we can observe in Figure 4 that the effect of λ on the conditioning is much less predictable for the TV and Huber penalties. For the Huber penalty with $\delta = 10^{-3}$, SPG is only able to converge for a finite range of λ values, while for the TV penalty with $\delta = 10^{-6}$, it is not able to converge for any value of λ.

As was the case for the stagnating behaviour observed with superiorized algorithm, the issues of conditioning appear to result from the choice of the parameter δ rather than the particular forms of the

Huber and TV functions. In particular, we found that when δ is chosen to be small (e.g. on the order of 10^{-6}) in either function, the spectral step α_k (5) is rapidly driven to zero, causing the algorithm to cease progressing while the *pgnorm* is still relatively large. This occurs as a consequence of the fact that as δ goes to zero, both the TV and Huber functions approach nondifferentiability at some points, producing relatively large terms in the denominator of (5). Specifically, nondifferentiability occurs when neighbouring pixel differences are zero, which is frequently the case for piecewise constant images. The superiorization methodology, on the other hand, does not require that the penalty function even be differentiable [34] and can therefore be straightforwardly applied to a wide variety of penalty functions.

Putting aside issues specific to the hybrid algorithm, it is useful to contrast regularization more generally with superiorization. One potential advantage of regularization is that the strength of the regularization term does not diminish as the iteration proceeds, unlike the size of the perturbations in the superiorized method. This means that noise is suppressed in later iterations as well as early ones, and so the semiconvergence phenomenon that we observed in our experiments with the superiorized algorithm is not an issue. On the other hand, because regularization combines the data fidelity and penalty functions into a single objective, there is no guarantee of achieving a specific level of data fidelity (ε) even if the overall objective is being reduced. Superiorization, on the other hand, provides rigorous guarantees that an ε-compatible solution can be obtained.

Our numerical experiments considered only the case where A is $m \times n$ with $m \approx n$. The underdetermined case where $m \ll n$ is also of considerable practical importance in CT imaging, for example, in the case of so-called sparse-view imaging. This case is studied in a number of papers on superiorization, e.g. [6, 35]. In the underdetermined case, it can be proven that the unique solution to the non-negative least squares problem can have at most $m-1$ non-negative entries [36], producing high-frequency oscillations in the resulting solution [37]. This undesirable behaviour is a property of the problem itself, thus any algorithm that solves the least-squares problem (e.g. our superiorized SART algorithm) must exhibit such behaviour in the limit. This phenomenon is separate from the semiconvergence behaviour described in this paper, but with the same outcome; the stopping criterion for the superiorized algorithm must be carefully chosen, to avoid the emergence of these oscillations once the perturbation size becomes small. A regularized approach in which such oscillations are penalized within the objective function avoids this issue by solving a modified problem; however, in this case the selection of the regularization strength may again prove challenging.

CONCLUSIONS

In this paper we have compared two approaches to solving the CT reconstruction problem that incorporate penalty functions to improve image quality. The first method takes the hybrid SPG algorithm of Guerrero et al. [7], initially proposed for image restoration, and attempts to adapt it to the CT reconstruction problem by incorporating a different system matrix A and different regularization functions. Ultimately, we find that the heuristics that were formulated in developing the hybrid method do not generalize well to the reconstruction problem. The difficulties arise due to the nonsmooth nature of the penalty functions that are commonly used in image reconstruction, and possibly due to the different nature of the restoration and reconstruction problems.

The second method applies the superiorized SART algorithm [6] to the reconstruction problem. Overall we find that the superiorized method is effective in improving image quality, with some caveats. The

superiorized method is prone to the same semiconvergence phenomenon as SART, and so a stopping criterion must be chosen carefully to prevent image quality from degrading. As well, if the size of the perturbations introduced by the superiorization methodology are large, the algorithm may require many iterations to achieve the same data fidelity as the original SART algorithm.

Our experiments indicate that SPG applied to the regularized least squares problem and the superiorized SART algorithm are both capable of achieving roughly the same improvements in image quality if the parameters are chosen optimally. For SPG, this corresponds to the regularization parameter λ, while for superiorized SART, the parameters N and γ control perturbation size, while ε determines the stopping criterion. Our experiments indicate that finding optimal values of these parameters is challenging, and warrants further study.

ACKNOWLEDGMENTS

This work was supported by a National Research Experience for Undergraduate Program (NREUP) from the National Science Foundation (grant DMS-1460699), Mathematical Association of America (grant DMS-1652506) with additional funding provided by University of Washington Bothell. The authors thank Marcos Raydan for providing code for the hybrid optimization method, and the two anonymous referees for their constructive suggestions.

REFERENCES

[1] M. Beister, D. Kolditz, and W. A. Kalender. Iterative reconstruction methods in X-ray CT. *Physica Medica*, 28(2):94–108, 2012.

[2] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena*, 60(1-4):259–268, 1992.

[3] P. J. Huber et al. Robust estimation of a location parameter. *The annals of mathematical statistics*, 35(1):73–101, 1964.

[4] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of inverse problems*. Kluwer Academic Publishers, 1996.

[5] P. C. Hansen. *Rank-Deficient and Discrete Ill-Posed Problems*. SIAM, 1998.

[6] G. T. Herman, E. Garduño, R. Davidi, and Y. Censor. Superiorization: An optimization heuristic for medical physics. *Med. Phys.*, 39:5532–5546, 2012.

[7] J. Guerrero, M. Raydan, and M. Rojas. A hybrid-optimization method for large-scale non-negative full regularization in image restoration. *Inverse Problems in Science and Engineering*, 21(5):741–766, 2013.

[8] E. G. Birgin, J. M. Martínez, and M. Raydan. Nonmonotone spectral project gradient methods on convex sets. *SIAM Journal on Optimization*, 10(4):1196–1211, 2000.

[9] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983.

[10] A. H. Andersen and A. C. Kak. Simultaneous algebraic reconstruction technique (SART): a superior implementation of the ART algorithm. *Ultrasonic Imaging*, 6(1):81–94, 1984.

[11] F. Natterer. *The mathematics of computerized tomography*, volume 32. SIAM, 1986.

[12] E. G. Birgin, J. M. Martínez, and M. Raydan. Spectral projected gradient methods: Review and perspectives. *Journal of Statistical Software*, 60(3):1–21, 2014.

[13] J. Barzilai and J. M. Borwein. Two point step size gradient methods. *IMA Journal of Numerical Analysis*, 8:141–148, 1988.

[14] M. Raydan. On the Barzilai and Borwein choice of steplength for the gradient method. *IMA Journal of Numerical Analysis*, 13:321–326, 1993.

[15] M. Raydan. The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem. *SIAM Journal on Optimization*, 7:26–33, 1997.

[16] L. Grippo, F. Lampariello, and S. Lucidi. A nonmonotone line search technique for Newton's method. *SIAM Journal on Numerical Analysis*, 4(23):707–716, 1986.

[17] E. G. Birgin, J. M. Martínez, and M. Raydan. Algorithm 813: Spg-software for convex-constrained optimization. *ACM Transactions on Mathematical Software*, 27:340–349, 2001.

[18] S. Kaczmarz. Approximate solution of systems of linear equations (trans. P.C. Parks). *International Journal of Control*, 57(3):1269–1271, 1993. (Originally published as: Angenäherte Auflösung von Systemen linearer Gleichunger. Bulletin International de l'Academie Polonaise des Sciences. Lett A, 355-357. 1937).

[19] R. Gordon, R. Bender, and G. T. Herman. Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography. *Journal of theoretical biology*, 29(3):471–481, 1970.

[20] T. Strohmer and R. Vershynin. A randomized Kaczmarz algorithm with exponential convergence. *Journal of Fourier Analysis and Applications*, 15(2):262, 2009.

[21] G. T. Herman. *Fundamentals of computerized tomography: image reconstruction from projections*. Springer Science & Business Media, 2009.

[22] Y. Censor and T. Elfving. Block-iterative algorithms with diagonally scaled oblique projections for the linear feasibility problem. *SIAM Journal on Matrix Analysis and Applications*, 24(1):40–58, 2002.

[23] M. Jiang and G. Wang. Convergence of the simultaneous algebraic reconstruction technique (SART). *IEEE Transactions on Image Processing*, 12(8):957–961, 2003.

[24] P. C. Hansen and J. S. Jørgensen. AIR Tools II: algebraic iterative reconstruction methods, improved implementation. *Numerical Algorithms*, 79(1):107–137, 2018.

[25] T. Elfving, P. C. Hansen, and T. Nikazad. Semiconvergence and relaxation parameters for projected SIRT algorithms. *SIAM Journal on Scientific Computing*, 34(4):A2000–A2017, 2012.

[26] D. Butnariu, R. Davidi, G. T. Herman, and I. G. Kazantsev. Stable convergence behavior under summable perturbations of a class of projection methods for convex feasibility and optimization problems. *IEEE Journal of Selected Topics in Signal Processing*, 1(4):540–547, 2007.

[27] Y. Censor, G. T. Herman, and M. Jiang. Special issue on superiorization: theory and applications. *Inverse Problems*, 33(4), 2017.

[28] B. Schultze, Y. Censor, P. Karbasi, K. Schubert, and R. Schulte. An improved method of total variation superiorization applied to reconstruction in proton computed tomography. *IEEE Trans. Med. Imag.*, 2019.

[29] D. Needell and R. Ward. Stable image reconstruction using total variation minimization. *SIAM Journal on Imaging Sciences*, 6(2):1035–1058, 2013.

[30] V. Haase, K. Stierstorfer, K. Hahn, H. Schöndube, A. Maier, and F. Noo. Exploring the space between smoothed and non-smooth total variation for 3d iterative CT reconstruction. In *Medical Imaging 2019: Physics of Medical Imaging*, volume 10948, pages 1094820–1 – 1094820–7. SPIE, 2019.

[31] K. Clark et al. The cancer imaging archive (TCIA): maintaining and operating a public information repository. *Journal of digital imaging*, 26(6):1045–1057, 2013.

[32] T. Elfving, P. C. Hansen, and T. Nikazad. Semi-convergence properties of Kaczmarz's method. *Inverse Problems*, 30(5):055007, 2014.

[33] V.A. Morozov. *Methods for Solving Incorrectly Posed Problems*. Springer-Verlag New York, 1984.

[34] Y. Censor, H. Heaton, and R. Schulte. Derivative-free superiorization with component-wise perturbations. *Numerical Algorithms*, 80:1219, 2019.

[35] G. T. Herman and R. Davidi. Image reconstruction from a small number of projections. *Inverse problems*, 24(4):045011, 2008.

[36] C. L. Byrne. Iterative image reconstruction algorithms based on cross-entropy minimization. *IEEE Transactions on image processing*, 2(1):96–103, 1993.

[37] C. L. Byrne. Thoughts on superiorization I – V. *ResearchGate*, 2019.