

## APPLICATIONS OF A NONNEGATIVELY CONSTRAINED ITERATIVE METHOD WITH STATISTICALLY BASED STOPPING RULES TO CT, PET, AND SPECT IMAGING \*

JOHNATHAN M. BARDSLEY†

**Abstract.** In this paper, we extend a nonnegatively constrained iterative method and three stopping rules for its iteration to the medical imaging inverse problems of computed tomography (CT), positron emission tomography (PET), and single photon emission computed tomography (SPECT); the iterative method and stopping rules were introduced for the use in astronomical imaging. The paper begins with a brief introduction to the CT, PET, and SPECT mathematical and statistical models.

**Key words.** nonnegatively constrained iterative methods, inverse problems, statistical methods, medical imaging

**AMS subject classifications.** 65J22, 65K10, 65F22

**1. Introductory material.** The medical imaging modalities of computed tomography (CT), positron emission tomography (PET), and single photon emission computed tomography (SPECT) are widely used in the medical professions. Because of the unique set of strengths and weakness of each of these methods, they are utilized in different settings. However, mathematically they are closely related.

In this introductory section, we set the stage for the algorithmic discussion that constitutes the main result of the paper by briefly introducing the mathematical models for CT, PET, and SPECT imaging, their numerical discretization, and the associated statistical models. The mathematical discussion is included because it is accessible and will be of interest to the unfamiliar reader, while the statistical models are integral to the development of the computational methods that are our focus.

**1.1. Mathematical models.** CT is the most widely used of the three methods and has the simplest mathematical model. In the two-dimensional case, which is our focus in this paper, CT involves the reconstruction of the mass absorption function  $\mu$  of a body from one-dimensional projections of that body. A particular one-dimensional projection is obtained by integrating  $\mu$  along all parallel lines making a given angle  $\omega$  with an axis in a fixed coordinate system. Each line  $L$  can be uniquely represented in this coordinate system by  $\omega$  together with its perpendicular distance  $y$  to the origin.

Suppose  $L(\omega, y) = \{x(s) \mid 0 \leq s \leq S\}$ , with an X-ray source located at  $s = 0$  and a sensor at  $s = S$ . The standard assumption is that the intensity  $I$  of the X-ray along a line segment  $ds$  is attenuated via the model [6]

$$dI = -\mu(x(s))I ds.$$

The resulting ordinary differential equation can be solved using the method of separation of variables to obtain

$$I(S) = I(0)e^{-\int_0^S \mu(x(s)) ds},$$

\* Received August 25, 2010. Accepted for publication October 18, 2010. Published online February 21, 2011.  
Recommended by R. Plemmons. This work was supported by the NSF under grant DMS-0915107, and by the Physics Department at the University of Otago, New Zealand, who housed the author during his 2010-11 sabbatical year.

† Department of Mathematical Sciences, University of Montana, Missoula, MT, 59812  
(johnathan.bardsley@umontana.edu).

where  $I(0)$  is the intensity at the source and  $I(S)$  is the intensity at the receiver. Setting  $z = -\ln(I(S)/I(0))$ , we obtain the Radon transform model for CT:

$$(1.1) \quad z(\omega, y) = \int_{L(\omega, y)} \mu(x(s)) ds.$$

A discretized version of (1.1) is what is solved in the CT inverse problem, where  $z$  corresponds to collected data, and  $\mu$  is the unknown.

In both PET and SPECT, a somewhat different problem is solved. In both cases, a radioactive tracer element is injected into the body. The tracer then exhibits radioactive decay, resulting in photon emission. The emitted photons that leave the body are, in theory, all recorded by a photon detector, which also determines the line of response (LOR)  $L(\omega, y)$ . In PET, two photons are emitted, and if they reach the detector ring at the same time (i.e., within a few nanoseconds of each other), an event is registered along the connecting line, which is also the LOR. In SPECT, single photons are emitted and detected. The LOR is determined using a method known as columnation. In both cases, the task is to reconstruct the tracer density distribution  $u$  within the subject given the collected photon count data.

In both PET and SPECT, the data  $z(\omega, y)$  for the line  $L(\omega, y)$  correspond to the number of detected incidents along that line. The model relating the tracer density  $u$  to the data is similar to (1.1) and is given by

$$z(\omega, y) = \int_{L(\omega, y)} g_{\omega, y}(x(s)) u(x(s)) ds,$$

where the impulse response function  $g_{\omega, y}(x(r))$  can be viewed as the probability that an emission event located at  $x(r)$  along  $L(\omega, y)$  is detected.

For SPECT, we assume that the detector is located at  $x(S)$ . Then, making assumptions for probabilities analogous to those made for intensities in the derivation of the CT model above, and assuming probability of 1 at  $x(r)$ , one can obtain  $g_{\omega, y}(x(r)) = e^{-\int_r^S \mu(x(t)) dt}$ , so that the full model becomes

$$(1.2) \quad z(\omega, y) = \int_{L(\omega, y)} e^{-\int_s^S \mu(x(t)) dt} u(x(s)) ds,$$

where the interior integral is along the line  $L(\omega, y)$ .

For PET, since there are two photons that have to reach the respective detectors at  $x(0)$  and  $x(S)$ , the impulse response is the product of the probabilities,

$$g_{\omega, y}(x(r)) = e^{-\int_0^r \mu(x(t)) dt} e^{-\int_r^S \mu(x(t)) dt} = e^{-\int_{L(\omega, y)} \mu(x(t)) dt},$$

which does not depend on  $r$  and hence we have the somewhat simpler mathematical model

$$(1.3) \quad z(\omega, y) = e^{-\int_{L(\omega, y)} \mu(t) dt} \int_{L(\omega, y)} u(x(s)) ds.$$

A discussion of the PET mathematical and statistical models is given in [22].

Note the appearance of the absorption density  $\mu$  in both (1.2) and (1.3), which must be known beforehand in order to solve the PET and SPECT inverse problems. Estimates of  $\mu$  can be obtained using, for example, CT. Note also that brief derivations of all of these models can be found in [14], whereas [6] focuses on CT, [15] on PET, and [4] on SPECT.

**1.2. Numerical discretization.** After discretization, each of (1.1), (1.2), and (1.3) can be written as a system of linear equations. The discretization occurs both in the spatial domain where  $\mu$  and  $u$  are defined, as well as in the Radon transform domain where the independent variables are  $\omega$  and  $y$ . We will use a uniform  $n \times n$  spatial grid, and a grid for the transform domain with  $n_\omega$  angles and  $n_s$  sensors, both uniformly spaced. Then, after column-stacking the resulting two-dimensional grids, we obtain a matrix-vector system

$$(1.4) \quad \mathbf{z} = \mathbf{A}\mathbf{u},$$

with the data vector  $\mathbf{z} \in \mathbb{R}^{n_\theta n_s}$ , the unknown vector  $\mathbf{u} \in \mathbb{R}^{n^2}$ , and the forward model matrix  $\mathbf{A} \in \mathbb{R}^{n^2 \times (n_\theta n_s)}$ .

In the case of CT, for the  $i$ -th line  $L_i$ , we have the discrete model

$$z_i = \sum_{j=1}^{n^2} a_{ij}^{\text{Radon}} \mu_j,$$

where  $a_{ij}^{\text{Radon}}$  is the intersection length of  $L_i$  with pixel  $j$ . Written as (1.4), we have  $[\mathbf{A}]_{ij} = a_{ij}^{\text{Radon}}$  and  $u_j = \mu_j$ .

For both PET and SPECT, the system of linear equations instead has the form

$$z_i = \sum_{j=1}^{n^2} g_{ij} a_{ij}^{\text{Radon}} u_j,$$

where  $g_{ij}$  is the discrete impulse response function, which for PET is

$$g_{ij} = \exp \left( - \sum_{k=1}^{n^2} a_{ik}^{\text{Radon}} \mu_k \right),$$

(note no dependence on  $j$ ) and for SPECT is

$$g_{ij} = \exp \left( - \sum_{k=j}^{n^2} a_{ik}^{\text{Radon}} \mu_k \right).$$

Written as (1.4), we have  $[\mathbf{A}]_{ij} = \exp \left( - \sum_{k=1}^{n^2} a_{ik}^{\text{Radon}} \mu_k \right) a_{ij}^{\text{Radon}}$  for PET and  $[\mathbf{A}]_{ij} = \exp \left( - \sum_{k=j}^{n^2} a_{ik}^{\text{Radon}} \right) a_{ij}^{\text{Radon}}$  for SPECT.

**1.3. Statistical models.** The character of the noise in the data  $\mathbf{z}$  is an important consideration, particularly in the cases of PET and SPECT.

In CT, because intensities are typically high, a statistical model of the form

$$(1.5) \quad \mathbf{z} = \mathbf{A}\mathbf{u} + \mathbf{n},$$

where  $\mathbf{n}$  is a zero mean, independent and identically distributed normal distribution with variance  $\sigma^2$ , is not uncommon. The resulting likelihood function takes the form

$$(1.6) \quad p_{\mathbf{z}}(\mathbf{z}; \mathbf{u}) \propto \exp \left( - \frac{1}{2\sigma^2} \|\mathbf{A}\mathbf{u} - \mathbf{z}\|^2 \right).$$

PET and SPECT data are typically much more noisy, and hence it is important to accurately model noise statistics in these cases. We follow [4, 15], and use a Poisson statistical model

$$(1.7) \quad \mathbf{z} = \text{Poiss}(\mathbf{A}\mathbf{u} + \boldsymbol{\gamma}).$$

In this case, the likelihood function takes the form

$$(1.8) \quad p_{\mathbf{z}}(\mathbf{z}; \mathbf{u}) = \prod_{i=1}^n \frac{([\mathbf{A}\mathbf{u}]_i + \gamma_i)^{z_i} \exp[-([\mathbf{A}\mathbf{u}]_i + \gamma_i)]}{z_i!}.$$

Given image data  $\mathbf{z}$  arising from model (1.5) or (1.7), the maximum likelihood estimate of  $\mathbf{u}$  is obtained by maximizing  $p_{\mathbf{z}}(\mathbf{z}; \mathbf{u})$  with respect to  $\mathbf{u} \geq \mathbf{0}$ , or equivalently by solving

$$(1.9) \quad \mathbf{u}_{\text{ML}} = \arg \min_{\mathbf{u} \geq \mathbf{0}} T(\mathbf{u}; \mathbf{z}),$$

where for likelihood (1.6),

$$(1.10) \quad T(\mathbf{u}; \mathbf{z}) = \frac{1}{2\sigma^2} \|\mathbf{A}\mathbf{u} - \mathbf{z}\|^2,$$

whereas for likelihood (1.8),

$$(1.11) \quad T(\mathbf{u}; \mathbf{z}) = \sum_{i=1}^n \{([\mathbf{A}\mathbf{u}]_i + \gamma_i) - z_i \ln([\mathbf{A}\mathbf{u}]_i + \gamma_i)\}.$$

Solutions of (1.9) tend to be noise-corrupted due both to random errors in  $\mathbf{z}$  and ill-conditioning of the matrix  $\mathbf{A}$ , which in all cases is the discretization of a compact operator [13]. Thus, some form of regularization is needed. One way that this can be accomplished is by truncating an iterative method applied to (1.9) [5]. When this approach is taken, the choice of stopping the iteration becomes extremely important and is akin to the choice of a regularization parameter in the standard Tikhonov approach to regularization [5, 23].

In the context of PET imaging, the Richardson-Lucy algorithm (RL) is widely used. RL is an iterative method for solving (1.9), (1.11) [10, 19], and stopping rules for its iterations have been given [9, 16, 17, 18]. Much effort has gone into the development of efficient iterative methods for PET; see, e.g., the review paper [15] and the references therein and also [8]. In this paper, we apply the iterative method and stopping rules of [1, 3], developed in the context of astronomical imaging, to (1.9) for both (1.10) and (1.11). The iterative method, called modified residual norm steepest descent (MRNSD) in [12], was originally introduced in [7] in the context of PET.

The remainder of the paper is organized as follows. In Section 2, we present the iterative method and stopping rules that we will use for the image reconstruction step, and then in Section 3 we test our methods on some synthetically generated examples. Finally, we end with conclusions in Section 4.

**2. An iterative method and stopping rules.** It is now time to present our numerical method and iteration stopping rules for approximately solving (1.9). The methods require a least-squares formulation of the nonnegatively constrained optimization problem (1.9). For noise model (1.5) and  $T$  defined by (1.10), this is already the case. However, for noise model (1.7) an approximation of (1.11) must be computed. For this, we follow [1, 2], where a Taylor series argument is used to motivate the following quadratic approximation of  $T$ :

$$T(\mathbf{u}; \mathbf{z}) \approx T(\mathbf{u}_e; \mathbf{z}) + T_{\text{wls}}(\mathbf{u}; \mathbf{z}),$$

where

$$(2.1) \quad T_{\text{wls}}(\mathbf{u}; \mathbf{z}) \stackrel{\text{def}}{=} \frac{1}{2} \|\mathbf{C}^{-1/2}(\mathbf{A}\mathbf{u} - (\mathbf{z} - \boldsymbol{\gamma}))\|^2,$$

with  $\mathbf{C} = \text{diag}(\mathbf{z})$ . Here we assume that  $\mathbf{z} > \mathbf{0}$ . This leads to the following approximation of (1.9):

$$(2.2) \quad \mathbf{u}_{\text{wls}} = \arg \min_{\mathbf{u} \geq \mathbf{0}} T_{\text{wls}}(\mathbf{u}; \mathbf{z}).$$

Before continuing, we note that  $T$  defined by (1.10) also can be written as (2.1), with  $\mathbf{C} = \sigma^2 \mathbf{I}$  and  $\boldsymbol{\gamma} = \mathbf{0}$ . In what follows we will use this convention in order to unify notation. However, we note that this implies that the user knows the noise level  $\|\mathbf{n}\|^2$  in (1.5) a priori.

**2.1. The iterative method.** The algorithm of [3] for the numerical solution of (2.2) has the form

$$(2.3) \quad \mathbf{u}_{k+1} = \mathbf{u}_k - \tau_k \mathbf{u}_k \odot \mathbf{A}^T \mathbf{C}^{-1}(\mathbf{A}\mathbf{u}_k - (\mathbf{z} - \boldsymbol{\gamma})),$$

where “ $\odot$ ” denotes Hadamard (component-wise) multiplication, and the line search parameter  $\tau_k$  in (2.3) is given by

$$\tau_k = \min\{\tau_{\text{uc}}, \tau_{\text{bd}}\}.$$

Here,  $\mathbf{v}_k = \mathbf{u}_k \odot \nabla T_{\text{wls}}(\mathbf{u}_k)$  and

$$\begin{aligned} \tau_{\text{uc}} &= \frac{\langle \mathbf{v}_k, \nabla T_{\text{wls}}(\mathbf{u}_k) \rangle_2}{\langle \mathbf{v}_k, \mathbf{A}^T \mathbf{C}^{-1} \mathbf{A} \mathbf{v}_k \rangle_2}, \\ \tau_{\text{bd}} &= \min \{ [\mathbf{u}_k]_i / [\mathbf{v}_k]_i \mid [\mathbf{v}_k]_i > 0, [\mathbf{u}_k]_i \neq 0 \}. \end{aligned}$$

This method has been shown to be effective on several astronomical imaging examples. For more details, see [1, 3]. In this paper, we will refer to this technique as weighted MRNSD, or WMRNSD.

**2.2. The stopping rules.** Next, we introduce three stopping rules for (2.3). Each can be motivated from the assumption that our statistical model has the form

$$(2.4) \quad \mathbf{z} - \boldsymbol{\gamma} = \mathbf{A}\mathbf{u} + \mathbf{n},$$

where  $\mathbf{n}$  is distributed as a zero-mean Gaussian random vector with covariance matrix  $\mathbf{C}$ . For (1.5),  $\boldsymbol{\gamma} = \mathbf{0}$ ,  $\mathbf{C} = \sigma^2 \mathbf{I}$ , and model (2.4) is exact. For (1.7), on the other hand, we take  $\mathbf{C} = \text{diag}(\mathbf{z})$ , and the negative-log likelihood function (2.1) results. Note that alternative to the mathematical derivation of (2.1) above, one can use the statistical approximation (2.4) to derive the approximate maximum likelihood problem (2.1). More detailed derivations of these stopping rules can be found in [1].

**2.2.1. The discrepancy principle.** Assuming that (2.4) holds,  $\mathbf{C}^{-1/2}(\mathbf{A}\mathbf{u}_e - (\mathbf{z} - \boldsymbol{\gamma}))$ , where  $\mathbf{u}_e$  is the true image, is approximately normally distributed with covariance  $\mathbf{I}$ . Then by a standard result

$$2T_{\text{wls}}(\mathbf{u}_e) \sim \chi^2(n),$$

where  $\chi^2(n)$  denotes the chi-squared distribution with  $n$  degrees of freedom, and hence

$$E(T_{\text{wls}}(\mathbf{u}_e)) \approx \frac{1}{2} E(\chi^2(n)) = \frac{n}{2}.$$

Since in early iterations  $\frac{2}{n} T_{\text{wls}}(\mathbf{u}_k)$  is typically much larger than 1, a stopping rule of the form

$$(2.5) \quad \frac{2}{n} T_{\text{wls}}(\mathbf{u}_k; \mathbf{z}) \leq 1 + \epsilon_n$$

is therefore reasonable. We note that  $\epsilon_n = 0$  corresponds to Morozov's discrepancy principle [11, 23], and we recommend its use unless (2.5) isn't satisfied in a feasible number of iterations or if it uniformly yields over-regularized reconstructed images. Otherwise,  $\epsilon_n$  can be taken to be, for example,  $\pm\sqrt{2n}/n$ , or  $\pm 2\sqrt{2n}/n$ , i.e.,  $\pm$  one or two times the standard deviation of  $\chi^2(n)$ .

We note that for model (1.5), this stopping rule is equivalent to that presented in [21].

**2.2.2. Generalized cross validation.** In [16, 18] it is shown that the generalized cross validation (GCV) method [5, 23] can be used to develop iteration dependent stopping rules for the steepest descent and RL iterations. Following their approach, in [1] a GCV method is developed for (2.3).

The GCV function for (2.1) at iteration  $k$  is defined by

$$(2.6) \quad \text{GCV}(k) = n \left\| \mathbf{C}^{-1/2} (\mathbf{A}\mathbf{u}_k - (\mathbf{z} - \boldsymbol{\gamma})) \right\|^2 / \text{trace}(\mathbf{I}_n - \mathbf{C}^{-1/2} \mathbf{A}\mathbf{A}_k)^2,$$

where  $\mathbf{A}_k$  is the iterative regularization matrix satisfying

$$\mathbf{A}_k(\mathbf{z} - \boldsymbol{\gamma}) = \mathbf{u}_k,$$

with  $\mathbf{u}_k$  the  $k$ th WMRNSD iterate. The idea is then to stop WMRNSD at the iteration  $k$  that results in an increase in the GCV function.

In order to evaluate  $\text{GCV}(k)$  in practice,  $\text{trace}(\mathbf{I}_n - \mathbf{C}^{-1/2} \mathbf{A}\mathbf{A}_k)$  must be approximated. This can be accomplished using the Trace Lemma [23]: given  $\mathbf{B} \in \mathbb{R}^{n \times n}$  and  $\mathbf{v}$  a discrete white noise vector,  $E(\mathbf{v}^T \mathbf{B} \mathbf{v}) = \text{trace}(\mathbf{B})$ . Thus given a realization  $\mathbf{v}$  from a white noise process,

$$\text{trace}(\mathbf{I}_n - \mathbf{C}^{-1/2} \mathbf{A}\mathbf{A}_k) \approx \mathbf{v}^T \mathbf{v} - \mathbf{v}^T \mathbf{C}^{-1/2} \mathbf{A}\mathbf{A}_k \mathbf{v}.$$

Thus, if we know  $\mathbf{A}_k$ , we can efficiently estimate  $\text{GCV}(k)$  at each iteration. However, we don't know  $\mathbf{A}_k$ , nor do we want to compute it. Instead, we follow [16] and define

$$\mathbf{w}_k = \mathbf{A}_k \mathbf{v},$$

which yields the following approximation of (2.6):

$$(2.7) \quad \widetilde{\text{GCV}}(k) = n \left\| \mathbf{C}^{-1/2} (\mathbf{A}\mathbf{u}_k - (\mathbf{z} - \boldsymbol{\gamma})) \right\|^2 / (\mathbf{v}^T \mathbf{v} - \mathbf{v}^T \mathbf{C}^{-1/2} \mathbf{A}\mathbf{w}_k).$$

An iteration for  $\mathbf{w}_k$  can be derived, following [1], and is given by

$$(2.8) \quad \begin{aligned} \mathbf{w}_{k+1} &= \mathbf{w}_k - \tau_k [\mathbf{w}_k \odot \mathbf{A}^T \mathbf{C}^{-1} (\mathbf{A}\mathbf{u}_k - (\mathbf{z} - \boldsymbol{\gamma})) \\ &\quad + \mathbf{u}_k \odot \mathbf{A}^T \mathbf{C}^{-1} (\mathbf{A}\mathbf{w}_k - \mathbf{v}) ], \end{aligned}$$

where  $\mathbf{w}_0 = (\mathbf{u}_0 \odot \mathbf{v}) / (\mathbf{z} - \boldsymbol{\gamma})$ .

Thus, we immediately have a stopping rule for WMRNSD iterations, namely, iterate (2.3) and (2.8) simultaneously and stop the iterations if  $\widetilde{\text{GCV}}(k) > \widetilde{\text{GCV}}(k-1)$ , where  $\widetilde{\text{GCV}}(k)$  is defined in (2.7). As it is standard [23], we choose  $\mathbf{v}$  so that its components are either -1 or 1 with equal probability.

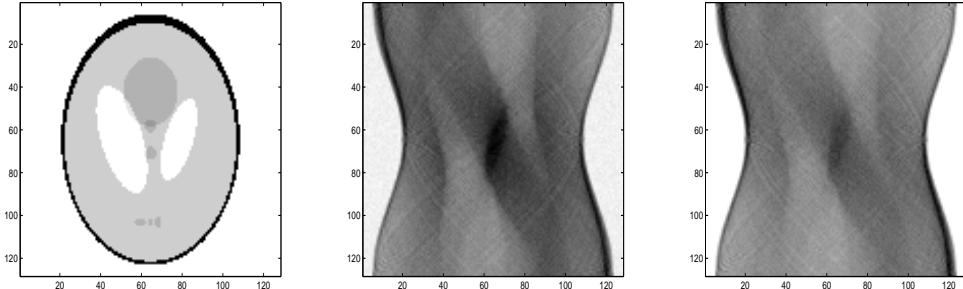


FIG. 3.1. The Shepp-Vardi phantom is pictured in the upper-left. The simulated sinogram for CT and SPECT are given, respectively, in the middle and right.

**2.2.3. The unbiased predictive risk estimator.** Similar to GCV is the unbiased predictive risk estimator (UPRE) [23]. With UPRE, the goal is to choose the iteration  $k$  for which the predictive risk  $E\left(\frac{2}{n}T_{wls}(\mathbf{u}_k; \mathbf{z}_e)\right)$ , where  $\mathbf{z}_e = \mathbf{A}\mathbf{u}_e + \boldsymbol{\gamma}$ , is smallest. The UPRE function for (2.1) is

$$(2.9) \quad \text{UPRE}(k) = \frac{2}{n}T_{wls}(\mathbf{u}_k; \mathbf{z}) + \frac{2}{n} \text{trace}\left(\mathbf{C}^{-1/2}\mathbf{A}\mathbf{A}_k\right) - 1.$$

It is derived as in [23] for regular least-squares problems and is an unbiased estimator of the predictive risk, hence its name. Following the same approach as for GCV, we can approximate (2.9) by

$$(2.10) \quad \widetilde{\text{UPRE}}(k) = \frac{2}{n}T_{wls}(\mathbf{u}_k; \mathbf{z}) + \frac{2}{n}\mathbf{v}^T\mathbf{C}^{-1/2}\mathbf{A}\mathbf{w}_k - 1,$$

where  $\mathbf{v}$  and  $\mathbf{w}_k$  are as above.

Thus, we have our third stopping rule for WMRNSD iterations; namely, iterate (2.3) and (2.8) simultaneously, and stop the iterations if  $\widetilde{\text{UPRE}}(k) > \widetilde{\text{UPRE}}(k-1)$ , where  $\widetilde{\text{UPRE}}(k)$  is defined in (2.10).

**3. Numerical results.** We now demonstrate the effectiveness of our methods on synthetically generated examples for each of the CT, PET, and SPECT imaging problems. For each of our tests we use the Shepp-Vardi phantom [20], given on the left in Figure 3.1, as the true image, generated by using MATLAB's phantom function. For data generation in CT, PET, and SPECT, we use the discretization of models (1.1), (1.3), and (1.2), respectively, described in Section 1.2. For PET, we ignore attenuation corresponding to  $\mu = 0$  in (1.3), which is common in synthetic numerical experiments done in the PET literature. For SPECT, we follow [4] and take  $\mu = 1$  where the absorption density of the body is positive and  $\mu = 0$  otherwise. The statistical model used for generating the CT data is (1.5), while for PET and SPECT it is (1.7). The noisy sinogram data for CT and SPECT are given in the middle and on the left in Figure 3.1. The sinogram for PET looks structurally very similar to the CT sinogram with noise more characteristic of the SPECT sinogram, thus we do not include it. The signal-to-noise level for these examples is approximately 30.

The inverse problem is to reconstruct the Shepp-Vardi phantom given the respective sinogram data  $\mathbf{z}$  and the forward model matrix  $\mathbf{A}$ . We do this by using the iterative method and stopping rule presented in Section 2 with an initial guess of  $\mathbf{u}_0 = \mathbf{1}$  and the three stopping rules, with  $\epsilon_n = 0$  in (2.5). The computations were done in MATLAB.

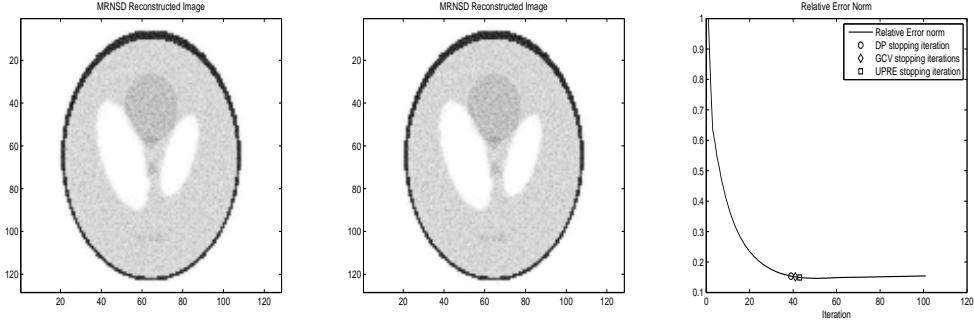


FIG. 3.2. *CT Example.* On the left and in the middle are the reconstructions obtained using the DP and UPRE stopping rules, respectively, and on the right is a plot of the relative error with the stopping iterations labeled.

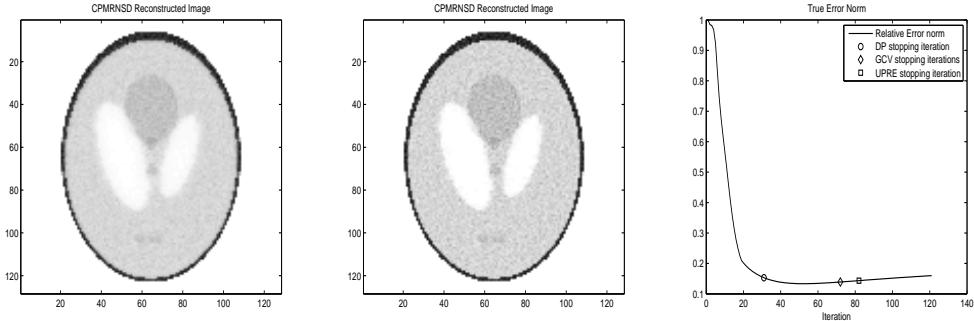


FIG. 3.3. *PET Example.* On the left and in the middle are the reconstructions obtained using the DP and GCV stopping rules, respectively, and on the right is a plot of the relative error with the stopping iterations labeled.

For the CT example, the reconstruction given by the DP and UPRE stopping rules are given on the left and middle, respectively, in Figure 3.2. On the right is a plot of the relative error  $\|\mathbf{u}_k - \mathbf{u}_{\text{true}}\|/\|\mathbf{u}_{\text{true}}\|$  with the stopping iterations for DP, GCV, and UPRE labeled. In this example, the reconstructions for the three methods are visually very similar.

Before continuing, we note that GCV and UPRE depend, through iteration (2.8), upon the particular randomly drawn vector  $\mathbf{v}$ . We have observed that different  $\mathbf{v}$  yield different stopping iterations. Though the methods seem to be relatively stable with respect to the choice of  $\mathbf{v}$ , additional stability can be obtained through averaging if multiple values of  $\mathbf{v}$  are chosen, and multiple iterations of the form (2.8) are implemented.

Additionally, we have found that regularly restarting the iteration (2.8) within the WM-RNSD iterations yields more effective stopping rules. In the results shown in Figure 3.2, as well as those that follow, we restart (2.8) every tenth WMRNSD iterations by setting  $\mathbf{w}_k = (\mathbf{u}_k \odot \mathbf{v}) / (\mathbf{z} - \boldsymbol{\gamma})$ .

For the PET example, the reconstruction given by the DP and GCV stopping rules are given on the left and middle in Figure 3.3. On the right is a plot of the relative error with the stopping iterations for DP, GCV, and UPRE labeled. In this case, the GCV and UPRE reconstructions appear slightly sharper than the DP reconstruction and have a slightly smaller relative error.

For the SPECT example, the reconstruction given by the DP and GCV stopping rules are given on the left and center in Figure 3.4. On the right is a plot of the relative error with the

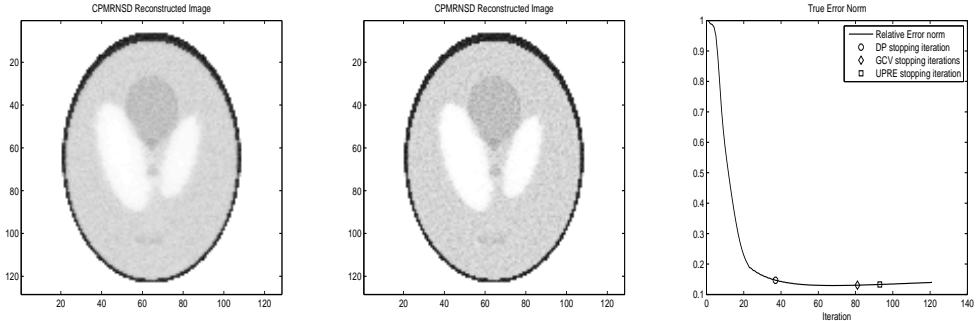


FIG. 3.4. *SPECT Example.* On the left and in the middle are the reconstructions obtained using the DP and GCV stopping rules, respectively, and on the right is a plot of the relative error with the stopping iterations labeled.

stopping iterations for DP, GCV, and UPRE labeled. In this case, once again, the GCV and UPRE reconstructions appear slightly sharper than the DP reconstruction, and have a slightly smaller relative error.

As it is the case for the discrepancy principle for regularization parameter selection in the case of Tikhonov regularization, the discrepancy principle consistently underestimates the stopping iteration. On the other hand, wherever we've chosen  $\epsilon_n = 0$  in (2.5), a negative value could reasonably be chosen as well, e.g., one standard deviation below the mean of the  $\chi^2(n)$ , or  $\epsilon_n = -\sqrt{2n}/n$ .

**4. Conclusions.** We have presented an iterative method and three statistically motivated stopping rules for solving inverse problems arising in the applications of CT, PET, and SPECT imaging. For PET and SPECT, a quadratic approximation of the negative-log Poisson likelihood function is needed in order to motivate the use of (2.1), whereas in the CT example Gaussian data, and hence a least-squares likelihood, are assumed. This allows for the implementation of the simple nonnegatively constrained WMRNSD algorithm of [1, 3] to be applied in all cases.

Stopping rules for WMRNSD follow from a Gaussian approximation of the data-noise model. For CT, such a model is assumed, whereas for PET and SPECT, an approximation (2.4) of a Poisson statistical model of the form (1.7) is used. The Gaussian/least-squares assumptions allow for the development of three stopping rules: the discrepancy principle, generalized cross validation, and unbiased predictive risk estimation. All three of these methods work in practice, as the numerical experiments show.

The codes used to perform the numerical experiments in this paper can be obtained from the author.

**Acknowledgment.** The author would like to thank Daniela Calvetti and Erkki Somersalo for their discussions of the PET problem and the National Science Foundation under grant DMS-0915107, for their support of this research.

#### REFERENCES

- [1] J. M. BARDSLEY, *Stopping rules for a nonnegatively constrained iterative method for ill-posed Poisson imaging problems*, BIT, 48 (2008), pp. 651–664.
- [2] J. M. BARDSLEY AND J. GOLDES, *Regularization parameter selection methods for ill-posed Poisson maximum likelihood estimation*, Inverse Problems, 25 (2009), 095005 (18 pages).
- [3] J. M. BARDSLEY AND J. G. NAGY, *Covariance-preconditioned iterative methods for nonnegatively constrained astronomical imaging*, SIAM J. Matrix Anal. Appl., 27 (2006), pp. 1184–1198.

- [4] P. GREEN, *Bayesian reconstructions from emission tomography data using a modified EM algorithm*, IEEE Trans. Med. Imag., 9 (1990), pp. 84–93.
- [5] P. C. HANSEN, *Rank-Deficient and Discrete Ill-Posed Problems*, SIAM, Philadelphia, 1997.
- [6] J. KAIPIO AND E. SOMERSALO, *Statistical and Computational Inverse Problems*, Springer, New York, 2005.
- [7] L. KAUFMAN, *Maximum likelihood, least squares, and penalized least squares for PET*, IEEE Trans. Med. Imag., 12 (1993), pp. 200–214.
- [8] L. KAUFMAN AND A. NEUMAIER, *PET regularization by envelope guided conjugate gradients*, IEEE Trans. Med. Imag., 15 (1996), pp. 385–389.
- [9] J. LLACER AND E. VEKLEROV, *Feasible images and practical stopping rules for iterative algorithms in emission tomography*, IEEE Trans. Med. Imag., 8 (1989), pp. 186–193.
- [10] L. B. LUCY, *An iterative technique for the rectification of observed distributions*, Astron. J., 79 (1974), pp. 745–754.
- [11] V. A. MOROZOV, *On the solution of functional equations by the method of regularization*, Soviet Math. Dokl., 7 (1966), pp. 414–417.
- [12] J. NAGY AND Z. STRAKOŠ, *Enforcing nonnegativity in image reconstruction algorithms*, in Mathematical Modeling, Estimation, and Imaging, D. C. Wilson et al., eds., Proc. of SPIE, 4121, SPIE, 2000, pp. 182–190.
- [13] F. NATTERER, *The Mathematics of Computerized Tomography*, SIAM, Philadelphia, 2001.
- [14] F. NATTERER AND F. WÜBBELING, *Mathematical Methods in Image Reconstruction*, SIAM, Philadelphia, 2001.
- [15] J. M. OLLINGER AND J. A. FESSLER, *Positron emission tomography*, IEEE Signal Proc. Mag., 14 (1997), pp. 43–55.
- [16] K. M. PERRY AND S. J. REEVES, *Generalized cross-validation as a stopping rule for the Richardson-Lucy algorithm*, in The Restoration of HST Images and Spectra II, R. J. Hanisch and R. L. White, eds., Space Telescope Science Institute, Baltimore, 1994, pp. 97–103.
- [17] S. PRASAD, *Statistical-information-based performance criteria for Richardson-Lucy image deblurring*, J. Opt. Soc. Amer. A, 19 (2002), pp. 1286–1296.
- [18] S. J. REEVES, *Generalized cross-validation as a stopping rule for the Richardson-Lucy algorithm*, Int. J. Imag. Syst. Tech., 6 (1995), pp. 387–391.
- [19] W. H. RICHARDSON, *Bayesian-based iterative method of image restoration*, J. Opt. Soc. Amer., 62 (1972), pp. 55–59.
- [20] L. A. SHEPP AND Y. VARDI, *Maximum likelihood reconstruction for emission tomography*, IEEE Trans. Med. Imag., 1 (1982), pp. 113–122.
- [21] H. J. TRUSSELL, *Convergence criteria of iterative restoration methods*, IEEE Trans. Acoust., Speech, Signal Process., 31 (1983), pp. 129–136.
- [22] Y. VARDI, L. A. SHEPP, AND L. KAUFMAN, *A statistical model for positron emission tomography*, J. Amer. Statist. Assoc., 89 (1985), pp. 8–37.
- [23] C. R. VOGEL, *Computational Methods for Inverse Problems*, SIAM, Philadelphia, 2002.