Estimation of the Optimal Iteration Number for Minimal Image Discrepancy

Gengsheng L. Zeng, Fellow, IEEE

Abstract — Due to noise, the iterative image reconstruction algorithms must stop early before reaching the convergence. There is an optimal stopping point, at which the discrepancy of the reconstruction to the true image reaches minimum. It is still an open problem to find this optimal stopping point. This paper establishes two approximate relationships towards solving this open problem. The first approximate relationship is between the iterative Landweber algorithm and an iteration-number-emulated filtered backprojection (FBP) algorithm. The second approximate relationship is between the optimal iteration-number-emulated FBP reconstruction and the optimal projection-domain filtered data. These two relationships can help us to estimate the optimal stopping point.

Keywords: image reconstruction, tomography, iterative reconstruction algorithm

I. INTRODUCTION

In medical imaging, iterative image reconstruction methods are now more popular than the analytical FBP (filtered backprojection) algorithm. One main motivation of using an iterative image reconstruction algorithm is its better performance in noise control.

Iterative algorithms are used to reconstruct an image by minimizing an objective function (or, equivalently, maximizing a likelihood function), which is equivalent to solving a system of linear equations. The objective function is set up in the projection domain. In transmission tomography, Gaussian noise model is commonly assumed in the sinogram domain. The maximum likelihood solutions for transmission tomography are equivalent to least-square solutions. When the objective function is minimized, the maximum likelihood solution is reached. Due to noise, this maximum likelihood solution is too noisy to be useful in clinical applications. The iterative algorithm minimizes the objective function step by step. An early stopping rule must be used to obtain a less-noisy solution.

If we evaluate the discrepancy between the $k$th iteration solution of an iterative algorithm and the true image, a typical discrepancy versus iteration number curve is shown in Fig. 1. It is observed from Fig. 1 that there is an initial converging trend, a minimum discrepancy point, and a later diverging trend. The minimum discrepancy solution is not the maximum likelihood solution due to the noise. It is beneficial to stop the iteration at the minimum discrepancy point and obtain a less-noisy image. Iterating too long means that the noise component dominates the solution whereas performing too few iterations means loss of resolution. This phenomenon has been termed semi-convergence by Natterer [1].

In 1990, Herbert proposed the use of Pearson’s chi-squared statistic to determine whether the reconstructed image followed the Poisson noise distribution. If the reconstruction followed the Poisson noise distribution the iteration could be stopped for emission tomography [2]. In 1991, Liacer and Veklerov proposed to stop the iteration when the reconstruction enters a region of feasibility in which the images represent a good compromise between sharpness and smooth regions of high activity, without the “noise artifact” having yet set in [3]. In 2007, Elfving and Nikazad suggested some stopping rules for Landweber-type iteration [4]. The rules presented in [4] required a training process and were not easy to implement. In 2008, Gaitanis et al studied the histograms of the solutions at each iteration and found that the tail distribution was related to the image quality [5]. Guo and Renaut in 2011 suggested a stopping rule by testing the probability of the sinogram being generated with the current reconstruction [6]. In a 2013 paper, Ben Boulallegue et al developed a heuristic method, in which the mean of the quadratic error between the sinogram and the forward projections normalized with the forward projections of the mean of the image was calculated. If this calculated value is close to 1, the iteration could be stopped [7]. In other words, the measurement is modeled as $P = AX + e$, where $e$ is Poisson noise. They calculated the ratio $\left(\sum (p_i - (AX^{(i)}))^2) / \sum (AX^{(i)})\right)$ and checked whether this ratio was close to one. Here $A$ is the projection matrix, $X^{(i)}$ is the reconstructed image at the $i$th iteration, and the subscript $i$ indicates the $i$th ray sum. Another method is to use the truncated SVD (singular value decomposition) as suggested by Borges et al in 2014 [8]. Morozov’s discrepancy principle was yet another stopping criterion [9]. This method calculated the $\ell^2$ discrepancy between the data and the reprojected image $\|AX^{(i)} - P\|^2$ estimate. When this discrepancy was equal to the $\ell^2$ norm of the noise, the iteration is terminated. The problem of Morozov’s method is the assumption that the power of $AX^{(i)} - P$ can equal to the power of noise when $X^{(i)}$ is not the true solution. Up to present time, it is an open problem to find the optimal stopping point, at which the reconstruction has the minimum discrepancy from the true solution.

This paper is organized as follows. In section II, two approximate relationships are introduced.
The first approximate relationship was established by us over the last seven years. This relationship extended the conventional FBP algorithm to an iteration-number-emulated FBP algorithm. This iteration-number-emulated FBP algorithm contained an “iteration number” so that is can be used to approximate the iterative Landweber algorithm.

The second approximate relationship developed in this paper estimates the optimal emulated “iteration-number” in the iteration-number-emulated FBP algorithm using the noisy measurements. The relation is expressed in the frequency-domain power spectra. To express anything in power spectra, one must assume that the associated signals are stationary. This is a strong assumption because the signals and noise in medical imaging measurements are not stationary.

Based on these two approximate relationships, this paper suggests a new method to estimate the optimal stopping point for the iterative Landweber algorithm. Computer simulations are provided in the Numerical Results section to evaluate the proposed two approximation methods.

\[ X^{(k)} = X^{(k-1)} + \alpha A^T (P - AX^{(k-1)}) , \]

where \( X^{(k)} \) is the result of \( X \) at the \( k \)th iteration and the relaxation parameter \( \alpha \) controls the step size. The Landweber algorithm is just a gradient decent algorithm with a constant step size (also known as the relaxation parameter) \( \alpha > 0 \). A gradient descent method with a non-constant optimal step size converges faster than the Landweber algorithm that uses a constant step size. Convergence rate is not a concern here; our purpose is to establish a relationship to the iteration-number-emulated FBP algorithm. For a quadratic objective function (1), the Landweber algorithm will converge if the relaxation parameter \( \alpha > 0 \) is small enough. It is well known that the convergence condition for the Landweber algorithm is \( 0 < \alpha < 2/\lambda_{\text{max}} \) where \( \lambda_{\text{max}} \) is the largest eigenvalue of \( A^T A \).

The recursive expression (2) can be transformed into a non-recursive form as follows.

\[
X^{(k)} = \alpha A^T P + (I - \alpha A^T A)X^{(k-1)} \\
= \alpha A^T P + (I - \alpha A^T A)[\alpha A^T P + (I - \alpha A^T A)X^{(k-2)}] \\
= \sum_{n=0}^{k-1} (I - \alpha A^T A)^n \alpha A^T P .
\]

In this paper, the initial condition \( X^{(0)} \) is assumed to be zero. If the square matrix \( (I - M) \) is non-singular, we have the identity:

\[
\sum_{n=0}^{k-1} M^n = (I - M)^{-1}(I - M^k) ,
\]

which can be readily verified by pre-multiplying \( (I - M) \) on both sides. Thus the non-recursive form (3) can be further written as a closed form without the \( \Sigma \) sign:

\[
X^{(k)} = (A^T A)^{-1}[(I - (I - \alpha A^T A)^k)A^T P .
\]

An alternative development of (5) is to introduce the residual

\[
R^{(k)} = A^T P - A^T AX^{(k)} \\
= A^T P - A^T AX^{(k-1)} - \alpha A^T AR^{(k-1)} \\
= R^{(k-1)} - \alpha A^T AR^{(k-1)} \\
= (I - \alpha A^T A)R^{(k-1)} \\
= (I - \alpha A^T A)^k R^{(0)} .
\]

If \( X^{(0)} \) is zero, then \( R^{(0)} = A^T P \) and (6) yields

\[
A^T P - A^T AX^{(k)} = (I - \alpha A^T A)^k A^T P .
\]

If \( A^T A \) exists, we have the same expression as (5).

In tomography, \( A^T A \) is a projection-backprojection operator in the matrix form. When it operates upon an image \( X \), it approximates the convolution of the image \( X \) and a two-dimensional (2D) 1/r kernel, where \( r \) is the distance to the origin [10, 11]. In the Fourier domain, this 1/r
convolution kernel corresponds to the Fourier-domain function 1/\|\vec{\omega}\| [10, 11], where \(\vec{\omega}\) is the frequency vector in the 2D Fourier domain.

Expression (5) is discrete; a better approximation of \(A^T A\) by 1/\|\vec{\omega}\| in the frequency domain requires denser sampling in the spatial domain and finer spatial domain representation of the image.

We thus obtain the approximate equivalence between the matrix \(A^T A\) and the Fourier-domain function 1/\|\vec{\omega}\|. We also notice that the identity matrix \(I\) corresponds to the transfer function 1 in the Fourier domain. Multiplication with a matrix \(k\) times is equivalent to multiplying the filter transfer function \(k\) times in the frequency domain. In the frequency domain, (5) approximates the application of the following filter to the backprojected image \(A^T P\) [11, 12]:

\[
F(\|\vec{\omega}\|) = \|\vec{\omega}\| [1 - (1 - \alpha / \|\vec{\omega}\|)^k], \quad \text{if } \|\vec{\omega}\| \neq 0.
\]

We define \(F(0) = 0\). In the discrete frequency domain, we require that \(1 - \alpha / \|\vec{\omega}\| < 1\), which is the same as \(0 < \alpha < 2\|\vec{\omega}\|\) for all non-zero discrete frequencies. If one assumes that the discrete frequencies are between -0.5 and 0.5 with a frequency sampling interval \(\Delta\), the lowest non-zero frequency is \(\Delta\) and our requirement becomes \(0 < \alpha < 2\Delta\). For example, if the 512-point Discrete Fourier Transform is used, \(\Delta = 1/512\). In this case, parameter \(\alpha\) can be chosen anywhere in \(0 < \alpha < 2/512\).

Using the Central Slice Theorem [10, 11], (5) is also equivalent to an FBP algorithm with the ramp filter replaced by

\[
F(\omega) = |\omega| [1 - |H(\omega)|^k], \quad \text{with } H(\omega) = 1 - \alpha / |\omega|.
\]

In (8), \(\vec{\omega}\) is the frequency vector in the frequency domain of the two-dimensional image, \(\omega\) in (9) is the frequency (scalar) in the frequency domain along the one-dimensional detector. They are related by the central slice theorem.

In the iteration-number-emulated FBP algorithm, the selection of the value of \(\alpha\) is \(0 < \alpha < 2\Delta = 2/(1/\Delta)\). Notice that 1/\(\Delta\) is the largest value of the filter 1/|\omega| . On the other hand, in the iterative Landweber algorithm, the selection of the value of \(\alpha\) is \(0 < \alpha < 2/\lambda_{\text{max}}(A^T A)\), where \(\lambda_{\text{max}}(A^T A)\) is the maximum eigenvalue of the matrix \((A^T A)\) and is also the \(\ell_2\) norm of the matrix \((A^T A)\) [12].

Our first relationship can be stated as follows. The \(k\)th iteration result \(X^{(k)}\) of the iterative Landweber algorithm can be approximately estimated by an iteration-number-emulated FBP result with the ramp filter replaced by (9).

This first relationship is not exact, but an approximation. In the derivation of this relationship, continuous “sampling” for the view angle and on the detector was assumed. It was also assumed that the image array for the iterative Landweber algorithm had an infinite size with each pixel being infinitesimal.

### B. Approximate relationship between the image-domain discrepancy and the spectrum-domain measure

In this part, we estimate the optimal “iteration number” \(k\) in the iteration-number-emulated FBP algorithm so that the resultant image is close to the true image.

In the matrix domain, \(P\) is a 1D vector containing all noisy measurements; \(P\) is made up of view-by-view measurements \(p_0(s)\), where \(s\) is the variable on the 1D detector. The noisy measurements \(p_0(s)\) can be decomposed into the noiseless part \(r_0(s)\) and the zero-mean noise \(n_0(s) : p_0(s) = r_0(s) + n_0(s)\).

Taking the 1D Fourier transform of \(p_0(s)\) with respect to variable \(s\), results in \(\hat{p}_0(\omega) = \hat{r}_0(\omega) + \hat{n}_0(\omega)\), where \(\hat{r}_0(\omega)\) and \(\hat{n}_0(\omega)\) are the Fourier transforms of \(r_0(s)\) and \(n_0(s)\) with respect to the variable \(s\), respectively.

Let the true image be \(X^{\text{true}}\). The discrepancy between the true image and the iteration-number-emulated FBP reconstruction \(X^{(k)}\) can be symbolically expressed as

\[
D(k) = E\|X^{\text{true}} - X^{(k)}\|^2
\]

\[
= E\|\mathbf{B}\mathbf{F}\mathbf{F}^{-1}([|\omega| \hat{r}_0(\omega) - |\omega| [1 - |H(\omega)|^k/\lambda_{\max}(\mathbf{A}^T\mathbf{A})]\{\hat{r}_0(\omega) + \hat{n}_0(\omega)\})\|^2
\]

where \(\mathbf{B}\) is the backprojection operator which is a summation over the view angle \(\theta\). \(\mathbf{F}\) is the 1D view-by-view inverse Fourier transform operator with respect to frequency \(\omega\), and \(H(\omega)\) is defined in (9). In (10), the noise is a random process; “E” represents the ensemble average.

In the following, we use the shorthand notation of \(\hat{r} = \hat{r}_0(\omega)\), \(\hat{n} = \hat{n}_0(\omega)\), \(\hat{p} = \hat{p}_0(\omega)\), and \(H = H(\omega)\). In discrete implementation, the combination of \(\mathbf{B}\) and \(\mathbf{F}\) can be expressed by a matrix \(\mathbf{B}\). Thus (10) can be expressed as

\[
D(k) = E\|\|\omega\| \hat{r} - |\omega| [1 - H^k](\hat{r} + \hat{n})\|^2
\]

\[
\leq E\|\|\omega\| \hat{r} - |\omega| [1 - H^k](\hat{r} + \hat{n})\|^2.
\]

Let the second factor on the right-hand-side of (11) be \(G(k)\) and

\[
G(k) = E\|\|\omega\| \hat{r} - |\omega| (1 - H^k)H^k(\hat{r} + \hat{n})\|^2
\]

\[
= E\sum_{\theta} \sum_{\omega} |\omega|^2 [\hat{r}^*H^k \hat{n} - \hat{n}^*H^k(1 - H^k)]\|^2
\]

\[
= E\sum_{\theta} \sum_{\omega} |\omega|^2 [\hat{r}^*H^k \hat{n} - \hat{n}^*H^k(1 - H^k)]^2,
\]

where \(\sum_{\omega}\sum_{\omega}\) sums over all discrete frequencies \(\omega\), \(\sum_{\theta}\sum_{\omega}\) sums over all detector view angles \(\theta\), and * denotes the complex conjugate. In (10), the image norm is defined for the 2D spatial-domain image. In (12), the norm is defined in the 2D “image” with variables (\(\omega, \theta\)).
In (12), the cross-terms such as \( E[\hat{n}_r \hat{n}_z] \), \( E[\hat{n}_r^2 \hat{n}_z] \), \( E[\hat{n}_r^2 \hat{n}_z] \), and \( E[\hat{n}_r \hat{n}_z^2] \) are set to zero. This is because the ensemble average of the noise is zero (in both spatial domain and frequency domain).

Next, we will find the power spectra. Let \( S_n = E[\hat{n}_r^2] \) be the noise spectrum, and \( S_p = E[\hat{p}_r^2] \) be the spectrum of the measured signal \( p = r + n \). Since \( S_p = S_r + S_n \), we can approximate the power spectrum of the noiseless signal by \( S_r = S_p - S_n \) even though we do not know the true signal \( r \). Therefore

\[
G(k) = \sum_\omega \sum_\theta \omega^2 [(S_p - S_n)H^{2k} + S_n(1 - H^{2k})^2]. \quad (13)
\]

Usually, the ensemble mean “\( E \)” requires multiple noise realizations. If the random process is stationary, the ensemble mean can be replaced by the spatial domain mean by using only one noise realization. For a stationary random process, the power spectrum is the Fourier transform of the spatial-domain auto-correlation function.

In (10), the image discrepancy function \( D(k) \) describes the squared “distance” between the iteration-number-emulated FBP result and the true image as the parameter \( k \) changes. The parameter \( k \) is the emulated iteration number in the iteration-number-emulated FBP algorithm. In reality, we are unable to calculate \( D(k) \) because we do not know the noiseless signal. On the other hand, the spectra \( S_n \) and \( S_r \) can be estimated from the measurements and the noise model.

The sinogram-domain discrepancy is measured by \( G(k) \). According to Parseval’s theorem [13], the sinogram-domain discrepancy can be evaluated in the frequency domain and thus (13).

We must point out that the proposed method of using minimum of \( G(k) \) to estimate the minimum of \( D(k) \) is not an exact method. The proposed method involves the Fourier spectrum calculation. Fourier spectrum calculation assumes that the random processes involved are stationary. However, the noise spectrum and signal spectrum in medical imaging are not stationary.

**C. Calculation of Power Spectra**

We can define the image discrepancy for a reconstruction algorithm as in the first line of (10). In real patient studies, this discrepancy function cannot be evaluated because the true image is unknown.

In (12) and (13), \( G(k) \) tries to estimate the closeness between the ramp-filtered true Radon transform and the iteration-number-emulated ramp-filtered measurements as a function of \( k \), using spectra. To estimate the power spectrum of a random process is easier than to estimate the original random process. For example, the power spectrum of the white noise is a constant and can be estimated by one realization, while the white noise itself is unpredictable. The power spectra for the signal and the noise can be estimated by one realization. Therefore \( G(k) \) defined in (12) and (13) is easier to obtain than the discrepancy \( D(k) \) in (10). An important question is whether the optimal \( k \) values for \( D(k) \) and \( G(k) \) are correlated so that we can use one to estimate the other. We will use computer simulations to answer this question in the next section.

The procedure for computing the measurement spectrum \( S_p \) is as follows.

Step 1: Obtain the measured sinogram, \( p_\theta(s) \), whose vector notation is \( P \) as in (1). Here \( \theta \) is the view angle and \( s \) is the detector coordinate.

Step 2: View-by-view find the 1D fast Fourier transform (FFT) \( \hat{p}_\theta(\omega) \) of the sinogram \( p_\theta(s) \), with respect to the detector linear coordinate \( s \), whose frequency is denoted by \( \omega \). The resulting frequency-domain signal \( \hat{p}_\theta(\omega) \) has variables \( \theta \) and \( \omega \).

Step 3: Find \( S_p(\omega, \theta) = \hat{p}_\theta(\omega) \hat{p}_\omega(\omega) \), which is the squared norm of the result \( \hat{p}_\theta(\omega) \) from Step 2. In other words, each point in \( S_p(\omega, \theta) \) is the square of the real part of \( \hat{p}_\theta(\omega) \) plus the square of the imaginary part of \( \hat{p}_\theta(\omega) \).

The evaluation of the noise spectrum \( S_n \) depends on the noise model in a medical imaging system. In this paper, the computer simulations assume X-ray CT transmission imaging. Let the incoming X-ray flux be \( I_0 \), which is the number of photons incident onto the object that would be reaching the detector if there were no object. After the transmission rays penetrate the object and reach the detector, along a certain path, the X-ray flux can be estimated by Beer’s law as

\[
I = I_0 e^{-x(s)dt} \quad (14)
\]

where \( x \) is the linear attenuation coefficient and \( p = \int xdt \) is the sinogram measurement. The noise \( I \) is Poisson distributed with its mean as the expected value of \( I \). Since

\[
p = \int xdt = \ln I_0 - \ln I \quad , \quad (15)
\]

the variance of the sinogram \( p \) can be estimated by

\[
\text{var}(p) = \left[ \frac{d(-\ln I)}{dl} \right]^2 \text{var}(l) \approx \frac{1}{I^2} \times I = \frac{1}{I_0^2} \quad (16)
\]

The noise model for transmission sinogram \( p \) can be assumed to be Gaussian distributed with a zero mean.

In practice, for a given sinogram (which is the combination of the true Radon transform and random noise), the noise variance can be estimated according to (14) as

\[
\text{var}(p_\theta(s)) \approx \frac{1}{I_0 \exp(-p_\theta(s))} \quad (17)
\]

for the transmission data. Here \( I_0 \) is blank scan measurement. In practice, after the sinogram measurements are provided, one realization of the pseudo zero-mean Gaussian noise with
variance given by (17) is generated by computer and its corresponding noise spectrum $S_n$ is estimated.

The procedure for generating the noise spectrum $S_n$ is as follows.

Step 1: Obtain the measured sinogram, $p_0(s)$.

Step 2: Find the noise variance for each measurement according to a noise model, for example, (17).

Step 3: Generate pseudo zero-mean Gaussian noise $n_0(s)$ using the calculated noise variance from Step 2.

Step 4: View-by-view find the 1D fast Fourier transform (FFT) $\hat{n}_0(\omega)$ of the pseudo noise $n_0(s)$, with respect to the detector linear coordinate $s$, whose frequency is denoted by $\omega$. The resulting frequency-domain pseudo noise $\hat{n}_0(\omega)$ has variables $\theta$ and $\omega$.

Step 5: Find $S_n(\omega, \theta) = \hat{n}_0(\omega)\hat{n}_0(\omega)$, which is the squared norm of $\hat{n}_0(\omega)$ from Step 4. Each point in $S_n(\omega, \theta)$ is the square of the real part of $\hat{n}_0(\omega)$ plus the square of the imaginary part of $\hat{n}_0(\omega)$.

III. NUMERICAL RESULTS

A. Iterative Landweber reconstructions vs. iteration-number-emulated FBP reconstructions

A computer generated elliptical phantom was used to compare the reconstruction results using both the iterative Landweber algorithm and the iteration-number-emulated FBP algorithm, as shown in Fig. 2. The Radon transforms were generated analytically. The number of views was 427 over $180^\circ$. The detector size was 729 bins. The reconstructed image was stored in a $512 \times 512$ array. No noise was added to the projections in this part, and the purpose here is to compare the spatial resolution for some given $k$ values. The parameter $\alpha$ was 0.001. The line profiles indicate that both algorithms give almost the same spatial resolution when they use the same parameter $k$ with the parameter $\alpha$ fixed. The computer simulation results in Fig. 2 imply that we can use the Fourier-domain function (9) to study the iterative Landweber algorithm.

Results in Fig. 2 demonstrate that the first relationship presented in this paper is a pretty good approximation. It is difficult for human eye to tell the differences between the images reconstructed by both methods with the same parameters $\alpha$ and $k$.

<table>
<thead>
<tr>
<th>$k$</th>
<th>Iterative Landweber</th>
<th>Iteration-number-emulated FBP</th>
<th>Line Profiles</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td><img src="image1.png" alt="Iterative Landweber" /></td>
<td><img src="image2.png" alt="Iteration-number-emulated FBP" /></td>
<td><img src="image3.png" alt="Line Profiles" /></td>
</tr>
<tr>
<td>50</td>
<td><img src="image1.png" alt="Iterative Landweber" /></td>
<td><img src="image2.png" alt="Iteration-number-emulated FBP" /></td>
<td><img src="image3.png" alt="Line Profiles" /></td>
</tr>
<tr>
<td>100</td>
<td><img src="image1.png" alt="Iterative Landweber" /></td>
<td><img src="image2.png" alt="Iteration-number-emulated FBP" /></td>
<td><img src="image3.png" alt="Line Profiles" /></td>
</tr>
</tbody>
</table>

![Figure 2. Iterative Landweber and iteration-number-emulated FBP reconstructions with the same parameter $k$ = 10, 50, and 100, respectively. The line profiles are drawn horizontally passing through the centers of the smaller bright and dark circles. The solid lines are for the iteration-number-emulated FBP images and the dashed lines are for the iterative Landweber images.](image4.png)
For each phantom, the images were reconstructed using the iteration-number-emulated FBP algorithm. Since the true image was known, the image discrepancy was calculated at each emulated iteration. The optimal iteration number \( k_D \), was recorded for every test case. At the optimal iteration number, \( k_D \), the image discrepancy \( D(k) \) reached the minimum value for the given noise level, using the iteration-number-emulated FBP algorithm.

On the other hand, the noiseless Radon transform’s power spectrum \( (n_p S_S) \) was estimated for each phantom from noisy measurements. This estimation was achieved by the power spectrum of the noisy measurements minus the estimated noise power spectrum. One noise realization of the detector noise was generated for each noise level and its corresponding noise power spectrum \( nS \) was calculated, which was very similar to the white noise’s spectrum. The curve \( G(k) \) as defined in (13) was created for each noise level and for each phantom. The optimal \( k \) (i.e., the \( k \) making \( G(k) \) minimum) was recorded for each test case. Let the optimal \( k \) be \( k_G \) for each case.

### Table 1. Phantom 1 results

<table>
<thead>
<tr>
<th>Incident flux, ( I_0 )</th>
<th>Actual ( k_D )</th>
<th>Estimated ( k_G )</th>
<th>( D(k_D) )</th>
<th>( D(k_G) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>10000</td>
<td>150</td>
<td>105</td>
<td>0.1346</td>
<td>0.1659</td>
</tr>
<tr>
<td>25000</td>
<td>255</td>
<td>175</td>
<td>0.0828</td>
<td>0.0941</td>
</tr>
<tr>
<td>50000</td>
<td>338</td>
<td>259</td>
<td>0.0568</td>
<td>0.0606</td>
</tr>
<tr>
<td>75000</td>
<td>399</td>
<td>326</td>
<td>0.0449</td>
<td>0.0466</td>
</tr>
<tr>
<td>100000</td>
<td>451</td>
<td>384</td>
<td>0.0377</td>
<td>0.0386</td>
</tr>
<tr>
<td>250000</td>
<td>671</td>
<td>646</td>
<td>0.0207</td>
<td>0.0214</td>
</tr>
<tr>
<td>500000</td>
<td>915</td>
<td>914</td>
<td>0.0124</td>
<td>0.0124</td>
</tr>
<tr>
<td>750000</td>
<td>1098</td>
<td>1158</td>
<td>0.0090</td>
<td>0.0090</td>
</tr>
<tr>
<td>1000000</td>
<td>1248</td>
<td>1332</td>
<td>0.0071</td>
<td>0.0071</td>
</tr>
</tbody>
</table>

### Table 2. Phantom 2 results

<table>
<thead>
<tr>
<th>Incident flux, ( I_0 )</th>
<th>Actual ( k_D )</th>
<th>Estimated ( k_G )</th>
<th>( D(k_D) )</th>
<th>( D(k_G) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>10000</td>
<td>72</td>
<td>36</td>
<td>0.0186</td>
<td>0.0625</td>
</tr>
<tr>
<td>25000</td>
<td>85</td>
<td>46</td>
<td>0.0097</td>
<td>0.0306</td>
</tr>
<tr>
<td>50000</td>
<td>96</td>
<td>54</td>
<td>0.0058</td>
<td>0.0181</td>
</tr>
<tr>
<td>75000</td>
<td>103</td>
<td>59</td>
<td>0.0043</td>
<td>0.0135</td>
</tr>
<tr>
<td>100000</td>
<td>108</td>
<td>63</td>
<td>0.0034</td>
<td>0.0105</td>
</tr>
<tr>
<td>250000</td>
<td>124</td>
<td>76</td>
<td>0.0017</td>
<td>0.0050</td>
</tr>
<tr>
<td>500000</td>
<td>137</td>
<td>86</td>
<td>0.0010</td>
<td>0.0029</td>
</tr>
<tr>
<td>750000</td>
<td>144</td>
<td>93</td>
<td>0.0007</td>
<td>0.0020</td>
</tr>
<tr>
<td>1000000</td>
<td>150</td>
<td>98</td>
<td>0.0006</td>
<td>0.0016</td>
</tr>
</tbody>
</table>

### Table 3. Phantom 3 results

<table>
<thead>
<tr>
<th>Incident flux, ( I_0 )</th>
<th>Actual ( k_D )</th>
<th>Estimated ( k_G )</th>
<th>( D(k_D) )</th>
<th>( D(k_G) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>10000</td>
<td>733</td>
<td>608</td>
<td>0.0166</td>
<td>0.0169</td>
</tr>
<tr>
<td>25000</td>
<td>1294</td>
<td>1118</td>
<td>0.0083</td>
<td>0.0083</td>
</tr>
<tr>
<td>50000</td>
<td>1834</td>
<td>1619</td>
<td>0.0045</td>
<td>0.0045</td>
</tr>
<tr>
<td>75000</td>
<td>2181</td>
<td>1974</td>
<td>0.0031</td>
<td>0.0031</td>
</tr>
<tr>
<td>100000</td>
<td>2438</td>
<td>2191</td>
<td>0.0023</td>
<td>0.0023</td>
</tr>
<tr>
<td>250000</td>
<td>3289</td>
<td>3015</td>
<td>0.0010</td>
<td>0.0010</td>
</tr>
<tr>
<td>500000</td>
<td>3952</td>
<td>3667</td>
<td>0.0005</td>
<td>0.0005</td>
</tr>
<tr>
<td>750000</td>
<td>4344</td>
<td>4055</td>
<td>0.0003</td>
<td>0.0003</td>
</tr>
<tr>
<td>1000000</td>
<td>4622</td>
<td>4332</td>
<td>0.0002</td>
<td>0.0002</td>
</tr>
</tbody>
</table>
Three scatter plots of $k_D$ vs. $k_G$ are shown in Figs. 4-9 for the six computer generated phantoms. The scatter plots show close correlation between $k_G$ and $k_D$. We thus can somehow use $k_G$ to estimate $k_D$.

In an iterative algorithm, a larger iteration number, $k$, means a longer computation time. For the iterative Landweber algorithm, any $(\alpha, k)$ pair satisfying $k \times \alpha = \text{const}$ gives approximately the same results. For example, the combination of $\alpha = 0.001$ and $k = 100$ gives almost the same results as the combination of $\alpha = 0.001$ and $k = 1000$. Once we have one $(\alpha, k)$ pair that gives the optimal solution, we can scale up $\alpha$ and scale down $k$ as long as the algorithm converges.

IV. CONCLUSION

In medical imaging, the maximum likelihood solution of an image reconstruction problem is usually too noisy to be useful. An iterative algorithm must stop early to obtain a less-noisy image. The optimal stopping point is where the discrepancy between the reconstructed image and the true image reaches minimum. Since the true image is never available, the optimal stopping point is difficult to estimate. This paper develops a frequency-domain approach to estimate the optimal stopping point for the iterative Landweber algorithm.

This paper first establishes an approximate relationship between the iterative Landweber algorithm and the iteration-
number-emulated FBP algorithm that is an analytic algorithm. The iteration number \( k \) is associated with a parameter \( k \) of a weighted ramp filter in the frequency domain. This relationship enables us to use (9) to study the properties of the iterative Landweber algorithm.

The second approximate relationship is based on the following intuition. If two very similar images (characterized by the image discrepancy measure \( D(k) \)) are generated by an FBP algorithm, their ramp-filtered sinograms must be very similar (characterized by the frequency-domain measure \( G(k) \)). The image-domain discrepancy is intractable because the true image is never known. The frequency-domain measure \( G(k) \) can be readily estimated, because the power spectra can be estimated without knowing the original signal and the original noise. For example, the noise power spectrum can be re-created by using another noise realization, which is different from the original noise.

The optimal stopping point according to image discrepancy measure \( D(k) \) is \( k_D \). The optimal stopping according to spectrum-domain measure \( G(k) \) is \( k_G \). Correlation between \( k_D \) and \( k_G \) has been demonstrated by computer simulations with three different phantoms and with various noise levels. Our future work will be to extend this approach to the situations with noise weighting [14], irregular sinogram sampling [15], non-zero initial condition [16], and penalty constraints [17].

REFERENCES