2-D TOMOGRAPHY FROM NOISY PROJECTIONS TAKEN AT UNKNOWN RANDOM DIRECTIONS

A. SINGER[*] AND H.-T. WU[†]

Abstract. Computerized Tomography (CT) is a standard method for obtaining internal structure of objects from their projection images. While CT reconstruction requires the knowledge of the imaging directions, there are some situations in which the imaging directions are unknown, for example, when imaging a moving object. It is therefore desirable to design a reconstruction method from projection images taken at unknown directions. Recently, it was shown that the imaging directions can be obtained by the diffusion map framework. Another difficulty arises from the fact that projections are often contaminated by noise, practically limiting all current methods, including the diffusion map approach. In this paper, we introduce two denoising steps that allow reconstructions at much lower signal-to-noise ratios (SNR) when combined with the diffusion map framework. The first denoising step consists of using the singular value decomposition (SVD) in order to find an adaptive basis for the projection data set, leading to improved similarities between different projections. In the second step, we denoise the graph of similarities using the Jaccard index, which is a widely used measure in network analysis. Using this combination of SVD, Jaccard index and diffusion map, we are able to reconstruct the 2-D Shepp-Logan phantom from simulative noisy projections at SNRs well below their currently reported threshold values. Although the focus of this paper is the 2-D CT reconstruction problem, we believe that the combination of SVD, Jaccard index graph denoising and diffusion maps is potentially useful in other signal processing and image analysis applications.

Key words. Computerized Tomography (CT), Diffusion maps, manifold graph denoising, singular value decomposition (SVD), Jaccard index, small world graphs, Shepp-Logan phantom.

1. Introduction. Transmission Computerized Tomography (CT) nowadays is a standard method to obtain internal structures non-destructively, routinely used in medical imaging [7, 13, 18, 17]. The classical two-dimensional CT problem is the recovery of a function $f : \mathbb{R}^2 \to \mathbb{R}$ from its Radon transform. In the parallel beam model, the Radon transform of f is given by the line integral

$$R_\theta(f)(r) = \int_{x\cdot\theta=r} f(x)dx,$$

where $\theta \in S^1$ is perpendicular to the beaming direction (S^1 is the unit circle), and $r \in \mathbb{R}$. The reconstruction of f from its Radon transform $R_\theta(f)$ is made possible due to the Fourier-projection slice theorem that relates the 1-D Fourier transform $\widehat{R_\theta(f)}$ of the Radon transform with the 2-D Fourier transform $\hat{f}$ of the function [7, 13, 18, 17]:

$$\widehat{R_\theta(f)}(\xi) = \hat{f}(\xi\theta), \quad \text{for all } \xi \in \mathbb{R}.$$

In other words, the 1-D Fourier transform of each projection is the restriction of the 2-D Fourier transform to the central line in the θ direction. Thus, the collection of the discrete 1-D Fourier transforms of all projections corresponds to the Fourier transform of the function f sampled on a polar grid. Therefore, the function f can be recovered by a suitable 2-D Fourier inversion. This reconstruction requires the knowledge of the beaming direction θ of each and every projection $R_\theta(f)$.

There are cases, however, in which the beaming directions are unknown, for example, when imaging certain biological proteins or other moving objects. In such

---

[*]Department of Mathematics and PACM, Princeton University, Fine Hall, Washington Road, Princeton NJ 08544-1000 USA, email: amits@math.princeton.edu

[†]Department of Mathematics, Princeton University, Fine Hall, Washington Road, Princeton NJ 08544-1000 USA, email: hauwu@math.princeton.edu

cases, one is given samples of the Radon transform $R_{\theta_i} f(r)$ for a finite but unknown set of N directions $\{\theta_i\}_{i=1}^N$, and the problem at hand is to estimate the underlying function f without knowing the directions. The sampling set for the parameter r is usually known and dictated by the physical setting of the acquisition process; for example, if the detectors are equally spaced then the values of r correspond to the location of the detectors along the detectors line, while the origin may be set at the center of mass. An alternative method for estimating the shifts will be discussed in the Summary section.

In this paper we address the reconstruction problem for the 2D parallel-beam model with unknown acquisition directions. Formally, we consider the following problem: Given N projection vectors $(R_{\theta_i} f(r_1), R_{\theta_i} f(r_2), \ldots, R_{\theta_i} f(r_n))$ taken at unknown directions $\{\theta_i\}_{i=1}^N$ that were randomly drawn from the uniform distribution over S^1 and $r_1, r_2, \ldots, r_n$ are fixed n equally spaced pixels in r, find the underlying density function $f(x)$ of the object.

This 2D reconstruction problem from unknown directions was previously considered by Basu and Bresler in [2, 1]. In particular, [2] derives conditions for the existence of unique reconstruction from unknown directions and shifts. The recovery problem is formulated as a non-linear system using the Helgason-Ludwig consistency conditions, that is used to derive uniqueness conditions. Stability conditions for the angle recovery problem under deterministic and stochastic perturbation models are derived in [1], where Cramér-Rao lower bounds on the variance of direction estimators for noisy projections are also given. An algorithm for estimating the directions is introduced in [1], and it consists of three steps: 1) Initial direction estimation; 2) Direction ordering; 3) Joint maximum likelihood refinement of the directions and shifts. Step 2 uses a simple symmetric nearest neighbor algorithm for projection ordering. Once the ordering is determined, the projection directions are estimated to be equally spaced on the unit circle, as follows from the properties of the order statistics of the uniform distribution. Thus, the problem boils down to sorting the projections with respect to their directions.

A different approach to sorting the projections with respect to their directions was employed in [6], where the ordering was obtained by a proper application of the diffusion map framework [5, 14]. Specifically, the method of [6] consists of constructing an $N \times N$ matrix whose entries are obtained from similarities between pairs of projections, followed by a computation of the first few eigenvectors of the similarity matrix. This method was demonstrated to be successful at relatively low SNRs, especially when the projections were first denoised using wavelet spin-cycling [4].

In this paper, we combine the diffusion map approach of [6] with two other denoising techniques that together allow reconstructions at much lower SNRs. The first denoising step consists of using the singular value decomposition (SVD) in order to find a basis for the projection data set. The advantage of the basis found by the SVD over the wavelet basis that was used in [6] is in its adaptivity to the data. A few singular vectors capture most of the data variability, and projecting the data projections onto this basis diminish the noise while capturing most of the signal features. We use the denoised projections to construct a similarity matrix between denoised projection pairs. Note that the only difference between the SVD procedure that we apply here and principal component analysis (PCA) is that the projections are not being centered by the removal of their average prior to the computation of the SVD.

Our second denoising technique consists of further denoising the similarity matrix using the Jaccard index (see, e.g. [8], where the Jaccard index was used to denoise

protein interaction maps). When two projections share a similar beaming direction, then it is expected that not only the similarity between the two of them would be significant, but also their similarity with all other projections of nearby beaming directions would be large. Fixing a pair of projections, the Jaccard index is a way of measuring the number of projections that are similar to both of them. The similarity measure between projections is often sensitive to noise: when the SNR is too low, we may assign a large similarity to projection pairs of completely different beaming directions. We use the Jaccard index to identify such false matchings of projections, because we do not expect a pair of projections of different beaming directions to have many projections that are similar to both of them. The denoised similarity matrix obtained by Jaccard index thresholding is then used as an input to the diffusion map method.

We performed extensive numerical experiments testing this combination of SVD, Jaccard index and diffusion map, and were able to reconstruct the 2-D Shepp-Logan phantom from simulative noisy projections at SNRs well below their currently reported threshold values [6]. Although the focus of this paper is the 2-D CT reconstruction problem, we believe that such combination of SVD, Jaccard index graph denoising and diffusion maps has the potential to become a useful tool for other signal processing and image analysis applications.

2. Background. In this section we provide the mathematical framework and tools that we use to solve the reconstruction problem. We first discuss the underlying geometry of the data as a one-dimensional closed curve in a high dimensional ambient space. This discussion is followed by a brief introduction to the diffusion map method and the way it provides intrinsic coordinates for the data. We also review SVD and the Jaccard index that will be later used for denoising the projections and their pairwise similarity matrix.

2.1. Underlying geometry. Every projection vector $(R_\theta f(r_1), R_\theta f(r_2), \ldots, R_\theta f(r_n))$ can be viewed as a point in $\mathbb{R}^n$. When varying the beaming direction θ over S^1, the projection vectors traverse a closed curve in $\mathbb{R}^n$. This curve can intersect itself, for example, when the function f has some axis of symmetry. The collection of N projection vectors $(R_{\theta_i} f(r_1), \ldots, R_{\theta_i} f(r_n))$ $(i = 1, \ldots, N)$ are therefore N sampling points on this closed curve. Noise contamination leads to deviation of these points from their underlying curve.

In the limit of infinitely large number of discretization points $n \to \infty$, we get a closed curve $\mathcal{C}$ in $L^2(\mathbb{R})$ whenever $f \in L^2(\mathbb{R}^2)$ has a compact support. Indeed, from the dominant convergence theorem we have that

$$\int_{\mathbb{R}} |R_{(\sin(t+h), \cos(t+h))}(f) - R_{(\sin t, \cos t)}(f)|^2 ds$$

$$\leq \int_{\mathbb{R}} \int_{\mathbb{R}^2} |f(x)\delta(s - x \cdot (\sin(t+h), \cos(t+h))) - f(x)\delta(s - x \cdot (\sin t, \cos t))|^2 dx ds \to 0,$$

as $h \to 0$, where we have used the following parametrization of the unit circle $S^1 = \{(\cos t, \sin t) \, | \, t \in [0, 2\pi)\}$. This means that $R_\theta(f)$ is a continuous function from S^1 to $L^2(\mathbb{R})$, and its image is a compact and connected continuous curve in $L^2(\mathbb{R})$. Hereafter we will assume that the curve $\mathcal{C}$ does not intersect itself, in particular, this assumption implies that the object has no axis of symmetry.

2.2. Diffusion Map. Diffusion map is a non-linear dimensionality reduction technique [5, 14] whose application to the reconstruction problem at hand was studied

in [6]. As discussed above, although the sampled projections are points in a high dimensional Euclidean space, they are restricted to a one-dimensional closed curve. This curve may have a complicated non-linear structure that may not be captured by projecting it linearly onto a low-dimensional subspace. Unlike linear methods such as SVD, diffusion map successfully finds the correct parametrization of the non-linear curve. In this subsection, we give a brief description of the diffusion map technique and discuss some of its properties and limitations.

We now outline the steps of the diffusion map algorithm. Suppose $x_1, \ldots, x_N \in \mathbb{R}^n$ is a collection of N data points to be embedded in a lower dimensional space. The first step is to construct an $N \times N$ matrix $\boldsymbol{W}$ of similarities between the data points. The similarities are defined using the Euclidean distances between the data points and a kernel function $k : \mathbb{R} \to \mathbb{R}$ scaled by a parameter $\epsilon > 0$ in the following way

$$W_{ij} = k\left(\frac{\|x_i - x_j\|}{\sqrt{2\epsilon}}\right), \quad \text{for } i, j = 1, \ldots, N.$$

The second step is to normalize $\boldsymbol{W}$ into a probability transition matrix $\boldsymbol{A}$ of a random walk on the data points by letting

$$\boldsymbol{A} = \boldsymbol{D}^{-1}\boldsymbol{W},$$

where $\boldsymbol{D}$ is a diagonal matrix whose entries are given by

$$D_{ii} = \sum_{j=1}^{N} W_{ij}, \quad \text{for } i = 1, \ldots, N.$$

We then define the normalized graph Laplacian $\boldsymbol{L}$ as

$$\boldsymbol{L} = \boldsymbol{I} - \boldsymbol{A}.$$

where $\boldsymbol{I}$ is the $N \times N$ identity matrix. The matrix $\boldsymbol{A}$ has a complete set of eigenvectors $\phi_0, \phi_1, \ldots, \phi_{N-1}$ with corresponding eigenvalues $1 = \lambda_0 \geq \lambda_1 \geq \cdots \geq \lambda_{N-1} \geq 0$, where $\phi_0 = (1, 1 \ldots, 1)^T$. Moreover, for positive kernels such as the Gaussian kernel $k(u) = \exp\{-u^2/2\}$, all resulting eigenvalues are non-negative. In the last step, the data points are embedded in a k-dimensional Euclidean space using the diffusion map defined by

$$x_i \mapsto (\lambda_1^t \phi_1(i), \ldots, \lambda_k^t \phi_k(i)), \quad \text{for } i = 1, \ldots, N,$$

where $t > 0$ is a parameter.

Whenever the data points are sampled from a low dimensional Riemannian manifold, the discrete random walk over the data points converges to a continuous diffusion process over that manifold in the limit of $N \to \infty$ and $\epsilon \to 0$. This convergence implies that the eigenvectors of $\boldsymbol{L}$ converge to the eigenfunctions of the continuous Laplace-Beltrami operator on the manifold when the data points are uniformly sampled over the manifold [5, 19, 3, 10]. If the sampling is non-uniform then the limiting operator is the backward Fokker-Planck operator rather than the Laplace-Beltrami [16], but we can still get the Laplace-Beltrami operator by using a different normalization of the similarity matrix [5].

In our case, the data points are the projections which are restricted to the closed-curve $\mathcal{C}$. Although the beaming directions are uniformly distributed over S^1, the

projections are not necessarily uniformly distributed over $\mathcal{C}$, due to the non-trivial Jacobian of the transformation $\theta \mapsto R_\theta(f)$. The eigenvectors of $\boldsymbol{L}$ computed by the diffusion map will therefore be discrete approximations of the eigenfunctions of the Fokker-Planck operator over $\mathcal{C}$. If instead we apply the normalization that leads to the Laplace-Beltrami operator, then the computed eigenvectors will be discrete approximations of the eigenfunctions of the Laplace-Beltrami operator over $\mathcal{C}$ which are nothing but the trigonometric functions of the arclength s given by 1, $\sin(2\pi m s/L), \cos(2\pi m s/L)$, $m = 1, 2, \ldots$, where L is the length of $\mathcal{C}$. In particular, the first two non-trivial eigenfunctions $\sin(2\pi s/L)$ and $\cos(2\pi s/L)$ provide a good parametrization of $\mathcal{C}$, as they embed $\mathcal{C}$ onto the unit circle S^1 in $\mathbb{R}^2$.

This embedding solves the problem. In fact, after assigning to each projection its relative beaming direction θ_i by the monotonic function $\theta_i \leftarrow \arctan \frac{\phi_1(i)}{\phi_2(i)}$, we reconstruct the correct ordering of the projections.

However, like other methods, diffusion map also has its limitations: noise will cause the data points to deviate from the curve. The perturbation of the data points by noise may distort the topology of the data set from being a non-intersecting closed curve (see, for example, Figure 2.1). We show below that the application of SVD and Jaccard index denoising helps the diffusion map to overcome these limitations.
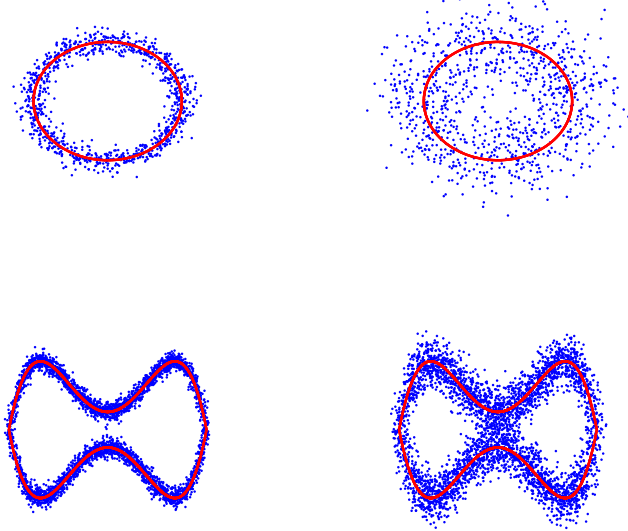

FIG. 2.1. *Upper row: The data (blue points) sampled from a circle (red line) at different noise levels. Lower row: The data (blue points) sampled from a closed curve (red line) at different noise levels.*

2.3. Denoising by SVD. Clearly, noise perturbs the topology through making the distances between projections (data points) less meaningful. It is therefore desirable to denoise the projections prior to computing the similarity matrix $\boldsymbol{W}$. A good denoising procedure will retain most characteristic features of the true signal while

diminishing the contribution of the noise. Thus, it is more beneficial to construct a similarity matrix $\boldsymbol{W}$ from the properly denoised projections and compute the eigenvectors of its corresponding random walk matrix. For example, in [6], denoising the projections using wavelet spin-cycling significantly improved the noise tolerance of the diffusion map algorithm. A possible limitation of the wavelet denoising approach is that the pre-chosen wavelet basis is not adaptive to the data, and it is reasonable to believe that an adaptive basis will lead to improved denoising. One way of constructing such an adaptive basis is using the SVD.

Suppose we have a data set $\{x_1, \ldots, x_N\} \subset \mathbb{R}^n$ stored in an $n \times N$ data matrix X. The SVD factorization of X is given by $X = \Psi \Sigma \Phi^T$ where Ψ is an $n \times n$ orthogonal matrix, Σ is an $n \times N$ diagonal matrix with decreasing singular values $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_{\min(n,N)}$ on its diagonal, and Φ is an $N \times N$ orthogonal matrix (see, e.g., [9]). By the Eckart-Young theorem, the best rank-k approximation X_k^* of X that minimizes the Frobenius norm, that is, $X_k^* = \operatorname{argmin}_{\operatorname{rank}(X_k) \leq k} \|X - X_k\|_F$ is given in terms of the SVD by $X_k^* = U \Sigma_k V^T$, where Σ_k is a diagonal matrix with the top k dominant singular values $\sigma_1, \ldots, \sigma_k$ on its diagonal while the other diagonal elements are set to zero.

In other words, for every fixed dimension k, SVD finds the linear subspace of that dimension that retains most of the energy the data upon projecting it onto that subspace. The left-singular vectors $\psi_1, \psi_2, \ldots, \psi_n \in \mathbb{R}^n$ (the columns of Ψ) define an orthogonal basis of $\mathbb{R}^n$. Note that the left singular vectors are also the eigenvectors of the $n \times n$ semi-definite positive matrix $XX^T = \Psi \Sigma^2 \Psi^T$ whose eigenvalues $\lambda_1, \ldots, \lambda_n$ are related to the singular values through $\lambda_i = \sigma_i^2$, for $i = 1, \ldots, n$. Denoising by SVD is obtained by projecting the data onto the subspace spanned by the first few singular vectors. More precisely, we project the dataset onto the k-dimensional subspace spanned by $\psi_1, \ldots, \psi_k$, for some choice of $k < n$. We later describe how to choose k based on the empirical distribution of the singular values $\sigma_1, \ldots, \sigma_n$ and its connection to the quarter circle law for the distribution of the eigenvalues of random matrices that are drawn from the Wishart distribution [12, 15].

2.4. Denoising by the Jaccard index. As mentioned above, the diffusion map method is limited by the presence of noise, as the latter may change the topology of the underlying manifold. In our case, noise can, for example, "shortcut" the curve (see, e.g., Figure 2.1). It is therefore desirable to detect such shortcut edges in advance and remove them from the similarity matrix $\boldsymbol{W}$.

After their introduction by Watts and Strogatz [23], small-world graphs were extensively used to describe many natural phenomena [11]. We briefly describe the small-world graph model. A d-regular ring graph is a graph whose vertices can be viewed as equally spaced points on the circle, and whose edges connect every point to its d nearest neighbors. The small-world network is constructed from the ring graph by randomly perturbing its edges: with probability p each ring edge is rewired to a random vertex, and with probability $1 - p$ it remains untouched. We refer to the rewired edges as "shortcuts".

The small-world graph obtained by rewiring the edges of the ring graph has the following useful property: the number of common neighbors for the vertices v and w with a "shortcut" edge $e = (v, w)$ between them is expected to be much smaller than the number of common neighbors of two nearby vertices. Thus, the number of common neighbors can be used as a measure for detecting shortcut edges from the edges of the original ring graph. One of the many possible measures for this detection

is the *Jaccard index*, defined by

$$J(v, w) = \frac{|N_v \cap N_w|}{|N_v \cup N_w|},$$

where N_v is the set of vertices connecting to the vertex v. It is therefore expected that the Jaccard index of shortcut edges will be smaller than that of the original ring edges.

Using the Jaccard index we can therefore detect the shortcut edges in the graph and further remove them in order to reveal the structure of the original graph. This observation was used in [8] to reveal the underlying structure of protein interaction maps. In our case, noise can fool us to believe that two projections of entirely different beaming directions correspond to two similar beaming directions. Such a confusion is realized by shortcut edges of our graph that change its topology and also the long time behavior of the random walk on the graph. Indeed, it was observed in [23] that the mixing time of the random walk on a small-world graph having a relatively small number of shortcut edges is significantly shorter compared to the mixing time of the random walk on the ring graph. It is therefore desirable to detect and remove the shortcut edges prior to applying the diffusion map technique. We use the Jaccard coefficient $J(v, w)$ in order to detect and remove such shortcut edges. Specifically, we set the similarity W_{ij} to zero for all edges (i, j) for which the Jaccard index $J(i, j)$ is below some threshold.

Put all the above together, the benefit we get can be summarized by the viewpoint of the Heisenberg Principle. Indeed, the top eigenvectors used by diffusion map to embed the data correspond to the long time behavior of the random walk over the data points. For example, the top trivial all-ones eigenvector corresponds to the steady state and the following top eigenvectors characterize the approach to steady state. The correspondence between the top eigenvectors and the long time behavior of the random walk can also be realized by considering their numerical computation using the iterative subspace power method, in which the transition matrix $\boldsymbol{A}$ is applied several times on a subset of initial vectors until convergence. Thus, the diffusion map is looking at high-orders powers of $\boldsymbol{A}$. On the other hand, the computation of the Jaccard index involves only the common neighbors, which is similar to looking at the diffusion process at a short time scale, corresponding to at most two steps of the random walk. Indeed, one way of computing the Jaccard index is by simply multiplying the adjacency matrix of the graph by itself. In this spirit, a weighted version J_w of the Jaccard index can also be defined as $J_w(i, j) = \left(\boldsymbol{A}^2\right)_{ij}$, which unravels the connection of the Jaccard index and the short time diffusion. It is also possible to define a new similarity matrix for the diffusion map framework based on this weighted Jaccard index, or even use a different Jaccard index based on other small powers, like the third or fourth power of $\boldsymbol{A}$. In any case, with the combination of the Jaccard index denoising and the diffusion map framework we benefit from both worlds of different time scales: the short time scale (Jaccard) and the long time scale (diffusion map).

3. Algorithm. In this section we detail the steps of our reconstruction algorithm from noisy projections at unknown directions. The input to the algorithm are N noisy projections $x_1, \ldots, x_N$, each of which is a vector in $\mathbb{R}^n$ corresponding to the discretization of the n uniformly spaced detectors.

Step 1: SVD. We apply SVD to the projections $x_1, \ldots, x_N$ to get the first k left-singular vectors $\psi_1, \ldots, \psi_k$ and their corresponding singular values $\sigma_1 \geq \sigma_2 \geq$

$\ldots \geq \sigma_k$. We further apply soft-thresholding wavelet denoising to the left singular vectors $\psi_1, \ldots, \psi_k$ in order to obtain a smoother version of them, denoted $\tilde{\psi}_1, \ldots, \tilde{\psi}_k$. The denoised projections $\tilde{x}_i$ are obtained by the linear operation

$$\tilde{x}_i = \sum_{j=1}^{k} \langle x_i, \tilde{\psi}_j \rangle \tilde{\psi}_j, \tag{3.1}$$

where $\langle , \rangle$ is the inner product of $\mathbb{R}^n$. Note that although the singular vectors $\{\psi_j\}_{j=1}^k$ are orthonormal, the denoised set $\{\tilde{\psi}_j\}_{j=1}^k$ need not be orthonormal, and the linear operation (3.1) is not necessarily a linear projection. To facilitate further computation, we compress the denoised projections $\tilde{x}_i \in \mathbb{R}^n$ and represent them as vectors in $\mathbb{R}^k$ given by $\hat{x}_i = (\langle x_i, \tilde{\psi}_1 \rangle, \ldots, \langle x_i, \tilde{\psi}_k \rangle)^T$ (for $i = 1, \ldots, N$).

Since the top singular vectors capture the main characteristic features of the data, we bring the spirit of soft-thresholding to enhance the importance of these vectors, by defining a weighted dot product and a corresponding weighted norm $\| \cdot \|_{\ell_2(\mathbb{R}^k, w)}$ in $\mathbb{R}^k$ using a vector $w = (w_1, w_2, \ldots, w_k)$ of positively decreasing weights (that is, $w_1 \geq w_2 \geq \ldots \geq w_k > 0$).

Step 2: Nearest Neighbors Search. For each of the N denoised projections we search for its kNN nearest neighbors with distances implied by the weighted norm $\| \cdot \|_{\ell_2(\mathbb{R}^k, w)}$.

Step 3: Jaccard index denoising. From the results of the nearest neighbors search we construct an undirected graph $G = (V, E)$ with N vertices corresponding to the projections, and where we put an edge between (i, j) iff either projection i is one of the kNN nearest neighbors of projection j or projection j is one of the kNN nearest neighbors of projection i. We compute the Jaccard index for every edge in the graph. We use several different threshold values and produce the corresponding denoised graphs, such that every edge whose Jaccard index is below the threshold is deleted from G. We also remove the isolated vertices of degree 0 and vertices of degree 1. The remaining steps of the algorithm are performed on each of the denoised graphs, until Step 6 where we automatically detect the threshold value that performed better than all others. We denote the denoised graph as $\tilde{G} = (\tilde{V}, \tilde{E})$, where $\tilde{V} \subset V$ and $\tilde{E} \subset E$.

Step 4: Diffusion map embedding. We construct an $|\tilde{V}| \times |\tilde{V}|$ similarity matrix $\boldsymbol{W}$ whose entries W_{ij} are defined by

$$W_{ij} = \begin{cases} \exp \left\{ -\dfrac{\|\hat{x}_i - \hat{x}_j\|_{\ell_2(w)}^2}{2\epsilon} \right\} & \text{for } (i, j) \in \tilde{E}, \\ 0 & \text{for } (i, j) \notin \tilde{E} \end{cases} . \tag{3.2}$$

We compute the top two non-trivial eigenvectors of $\boldsymbol{L} = \boldsymbol{I} - \boldsymbol{D}^{-1}\boldsymbol{W}$ denoted ϕ_1 and ϕ_2. The embedding $x_i \mapsto (\phi_1(i), \phi_2(i))$ reveals the ordering of the beaming directions. We estimate the beaming direction as equally spaced points on S^1 according to their ordering.

Step 5: 2-D Reconstruction. Inverse Radon transform is performed to reconstruct the 2-D image from the noisy projections and their estimated beaming directions.

Step 6: Automatic threshold detection using SVD. At this stage we get several different 2-D image reconstructions corresponding to the different choices of the threshold value of Step 3, and perhaps to different choices of other parameters such as ϵ in Step 4. Out of all available reconstructions we automatically choose the one of best quality, a measure we now define.

For each reconstructed image we compute its Radon transform and produce a series of projections at different beaming directions. We apply SVD to these projections and analyze the eigenvalues (squared singular values). Our empirical studies showed that the spectrum of eigenvalues of high quality reconstructions often has the following characteristics: the top first and second eigenvalues tend to be larger, while the small eigenvalues tend to get more concentrated near the origin, compared to all other reconstructions. The concentration of the small eigenvalues near the origin stems from the quarter circle law in random matrix theory (see, e.g., [12, 15]), which characterizes the non-uniform distribution of the eigenvalues of a random covariance matrix. The larger the noise level, the wider the distribution. We understand this phenomenon by viewing the data we have at hand as a superposition of the clean data with a random matrix due to noise, where in our case, the "noise" comes not only from the noise added to the projections but also from the wrong assignment of the angles. Based on these observations we apply an ad hoc quality measure that takes into account the values of the top first and second eigenvalues as well as the width of the quarter-circle.

4. Numerical results. We performed several numerical simulations in order to test our algorithm and its tolerance to noise. In all of our simulations the underlying 2-D object was the Shepp-Logan phantom, the number of projections was $N = 1024$, and the number of discretization points was $n = 512$. In each simulation, we added to the clean projections a Gaussian zero-mean white noise of a fixed variance σ^2. We define the SNR (measured in dB) by

$$\text{SNR [dB]} = 10 \log_{10} \left(\frac{\text{Var} \, S}{\sigma^2} \right),$$

where S is the array of the noiseless projections. As a reference to later reconstructions, Figure 4.1(a) shows the original Shepp-Logan phantom, while Figures 4.1(b) and 4.1(c) show reconstructions of the Shepp-Logan phantom from noisy projections with SNR $= -2$dB and SNR $= -3$dB, respectively, for which the beaming directions are fully known.

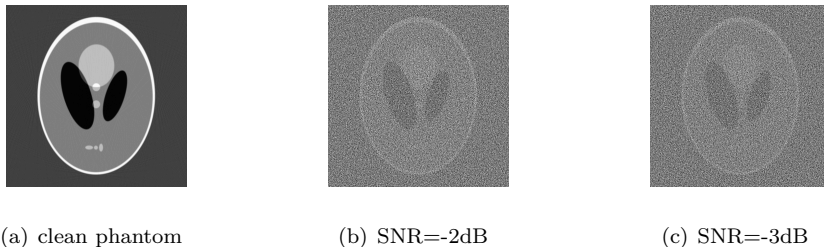(a) clean phantom   (b) SNR=-2dB   (c) SNR=-3dB

FIG. 4.1. *Reconstruction of the Shepp-Logan phantom image from projections with known beaming directions at different levels of noise: (a) no noise; (b) SNR $= -2dB$; (c) SNR $= -3dB$.*

The results of applying the algorithm described in Section 3 to noisy projections of the same level of noise (SNR $= -2$dB and SNR $= -3$dB) but without knowledge of

the beaming directions are illustrated in Figure 4.2. Obviously, the reconstructions in Figures 4.1(b)-(c) have better quality compared to the reconstructions in Figure 4.2(a) and Figure 4.2(d), that are missing the extra knowledge of the beaming directions. Still, our algorithm succeeds to provide similar reconstructions (up to the unavoidable degrees of freedom of rotation and reflection) even when the beaming directions are unknown. The main features of the original Shepp-Logan phantom are visible in our reconstructions even at such low SNRs. Figure 4.2(b) and Figure 4.2(e) demonstrate that the beaming directions are estimated successfully and mostly follow their true ordering. The different scaling of the x-axis between Figure 4.2(b) and Figure 4.2(e) is due to Step 3 of our algorithm in which vertices (projections) of degree 0 and 1 are removed without any further consideration.
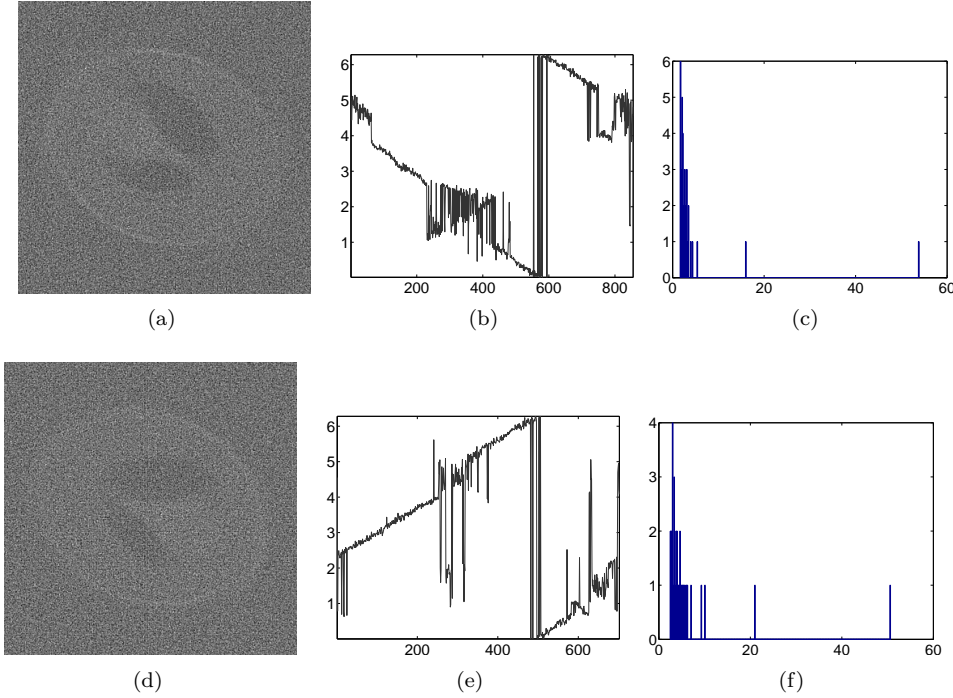

(a)   (b)   (c)

(d)   (e)   (f)

FIG. 4.2. *Reconstruction from noisy projections at unknown directions using the algorithm described in Section 3 for SNR = −2dB (top row) and SNR = −3dB (bottom row). (a) and (d) Reconstructed image; (b) and (e) Estimated beaming directions (y-axis) against their correct ordering (x-axis); (c) and (f) Histogram of the eigenvalues that are calculated in the final stage (Step 6) of our algorithm.*

Figure 4.3(a) and Figure 4.3(d) show reconstructions obtained by applying the method described in [6] to the same sets of noisy projections. Unlike the reconstructions provided by our algorithm, these reconstructions are blurry, and the estimated beaming directions do not follow their ground truth, as illustrated in Figure 4.3(b) and Figure 4.3(e). In the final stage of our algorithm (Step 6) we apply SVD to the projections obtained from the reconstructed image. Comparing the eigenvalue histograms, we see that the top eigenvalue in Figure 4.2(c) and Figure 4.2(f) is approximately 60, while the top eigenvalue in Figure 4.3(c) and Figure 4.3(f) is much smaller with a value near 10. The top eigenvalue thus provides yet another characterization for the quality of the reconstruction.
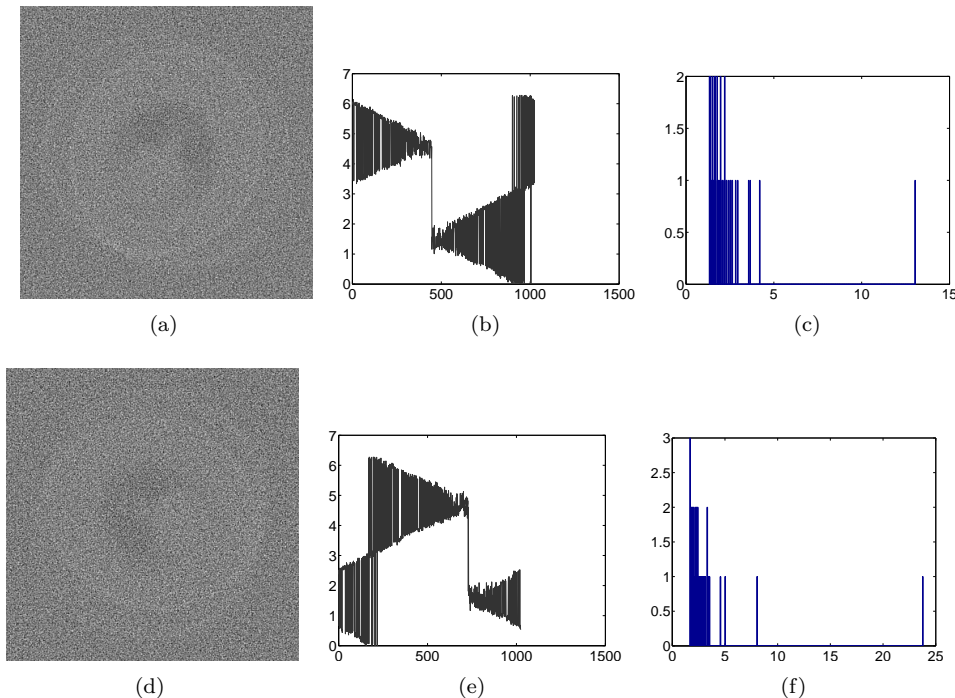
FIG. 4.3. *Reconstruction from noisy projections at unknown directions using the algorithm described in [6] for SNR = −2dB (top row) and SNR = −3dB (bottom row). (a) and (d) Reconstructed image; (b) and (e) Estimated beaming directions (y-axis) against their correct ordering (x-axis); (c) and (f) Histogram of the eigenvalues that are calculated in the final stage (Step 6) of our algorithm.*

In the following we describe the numerical results of the different steps of the algorithm and the specific choice of parameters in more detail. We start with Step 1 of the algorithm in which we compute the SVD of the 512×1024 data matrix X in order to denoise the projections. In all experiments the top left-singular vector was the constant vector $\psi_1 = \frac{1}{\sqrt{n}}(1, 1, \ldots, 1)^T$. From the Fourier slice theorem it follows that all clean projections share the same DC term, that is, $\int R_{\theta_1}(f)(r)\, dr = \int R_{\theta_2}(f)(r)\, dr$ for all $\theta_1, \theta_2 \in S^1$. Therefore, deviations in the dot product $\langle x_i, \psi_1 \rangle$ from one data projection to the other are the result of different realizations of noise, dominating the much smaller discretization errors. Since the trivial singular vector ψ_1 does not differentiate between clean projections at different directions, we set its weight to zero, that is, $w_1 = 0$. The other non-zero weights were chosen as $w_2 = w_3 = w_4 = 2$ and $w_5 = w_6 = w_7 = 1$. That is, the data is denoised by projecting it onto the subspace spanned by the top $k = 7$ singular vectors, with the exception of the trivial vector.

To justify this particular choice of k and w, we first applied SVD to different sets of noisy projections at different SNRs and examined the distribution of eigenvalues (squared singular values). The histograms and bar plots of the non-trivial eigenvalues are shown in Figure 4.4 (the eigenvalue λ_1 corresponding to ψ_1 is not shown). From the bar plot corresponding to the clean projections shown in Figure 4.4(f) it is possible to distinguish about 6 dominant eigenvalues, while the remaining eigenvalues are significantly smaller. Figure 4.4(g) and 4.4(h) illustrate that it is possible to distinguish the top 6 eigenvalues also at SNR = 2dB. These feature eigenvalues also

appear in the histogram plots in Figures 4.4(a)-(e), but they are hardly noticeable there due to the large number of small (near zero) eigenvalues. The small eigenvalues are attributed to noise, and as the noise increases the small eigenvalues get larger, as indicated in both histograms and bar plots. While the feature eigenvalues remain roughly the same (e.g., 300, 130, 35, 25), the eigenvalues corresponding to noise get larger as the level of noise increases. At SNR $= -2$dB and SNR $= -5$dB it is still possible to distinguish the top 4 feature eigenvalues, while at SNR $= -10$dB only the top 2 feature eigenvalue prevails. The small eigenvalues corresponding to noise follow the quarter circle law whose support increases as the level of noise increases until it swallows the feature eigenvalues as illustrated in the series of histograms 4.4(a)-(e). Our choice of $k = 7$ is thus simply motivated by the fact that we want to project the data onto the subspace of singular vectors that correspond to signal features rather than noise, and the eigenvalues attached to these feature singular vectors correspond to the discrete part of the spectrum rather than to the continuum (quarter circle) part of the distribution.
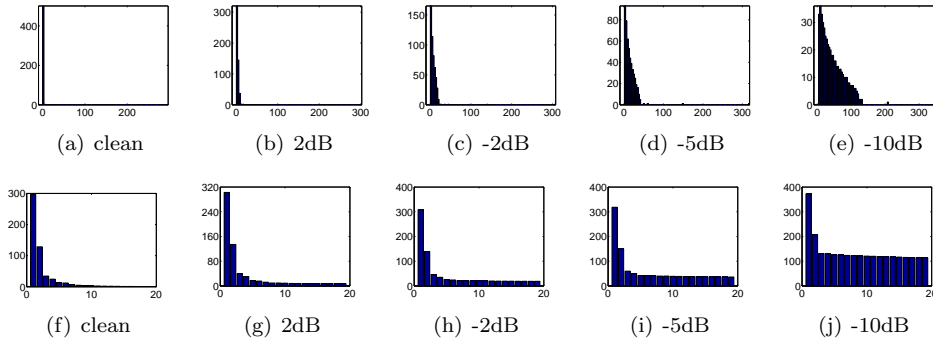

FIG. 4.4. *Distribution of the eigenvalues of XX^T (or equivalently, squared singular values of X) corresponding to noisy projections at different levels of noise. Top: Histograms; Bottom: Bar plots.*

Examining the top left singular vectors reveals that they somewhat resemble the shape of wavelets, though they are not really wavelet functions. Figures 4.5 and 4.6 show the first few left singular vectors obtained from clean projections and from noisy projections at SNR $= -2$dB, respectively. The resemblance between the two sets of singular vectors is striking, with the singular vectors in Figure 4.6 being the noisy versions of the clean singular vectors of Figure 4.5. The robustness of the top singular vectors to noise, as indicated by Figures 4.5 and 4.6 is explained by random matrix theory [12], that suggests that the singular vectors whose corresponding eigenvalues are above the support of the quarter circle distribution are highly correlated with the singular vectors of the clean projections data matrix.

Note that the clean projections are compactly supported due to the compactness of the Shepp-Logan phantom. Outside this support, the noisy projections are just made of noise. This leads to the wild oscillatory behavior of the noisy singular vectors outside the support (Figure 4.6). Loosely speaking, the SNR outside the support is $-\infty$ as the variance of the signal vanishes there. The local variations in the singular vectors may be explained by the local SNR, but we do not pursuit in this direction any further.

The purpose of the wavelet soft-thresholding denoising in Step 1 of the algorithm,

given symbolically by $\psi_j \to \tilde{\psi}_j$ $(j = 1, \ldots, k)$, is simply to smooth out the wild oscillations of the noisy singular vectors outside the compact support, as well as to smooth the singular vectors inside the support in a way that makes them resemble the (unknown) singular vectors of the clean data matrix even better.

We also performed experiments with PCA. However, while the top singular vectors of both the clean and noisy data matrices are well-behaved functions inside the support (Figures 4.5-4.6), the top principal components (eigenvectors of the sample covariance matrix where all projections are first centered by removing their average) have large oscillations both inside and outside the support. This lead us to believe that the SVD is more appropriate than PCA for denoising our data.
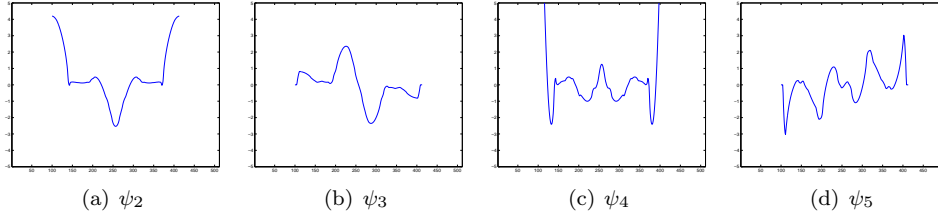

| (a) ψ_2 | (b) ψ_3 | (c) ψ_4 | (d) ψ_5 |

FIG. 4.5. *Singular vectors of the data matrix for clean projections.*
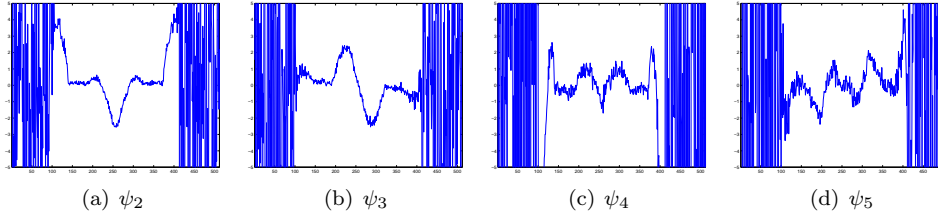

| (a) ψ_2 | (b) ψ_3 | (c) ψ_4 | (d) ψ_5 |

FIG. 4.6. *Singular vectors of the data matrix for noisy projections with $SNR = -2dB$.*

A comparison between denoising the projections using SVD and wavelet denoising is illustrated in Figure 4.7 and Figure 4.8. While denoising by SVD makes use of just the top 6 non-trivial singular vectors, the wavelet denoising procedure consists of using the full spin-cycle algorithm [4] with hard thresholding the Daubechies db2 wavelet coefficients, following the same denoising procedure used in [6]. The comparison shows that both denoising methods do relatively well for SNR = 2dB (Figure 4.7), but the SVD is doing a much better job for SNR = −2dB (Figure 4.8). One may conclude that denoising by SVD succeeds at relatively low SNRs due to the adaptivity of the singular vectors to the data.

In Step 2 of the algorithm, we computed the $kNN = 50$ nearest neighbors of each denoised projection and constructed the graph of nearest neighbors. This corresponds to linking each projection with a fraction of $\frac{50}{1024} \approx 5\%$ of all other projections. The edges of the resulting adjacency matrix were then denoised in Step 3 by Jaccard index thresholding, whose effectiveness is demonstrated in Figure 4.9. The vertices are arranged on a circle according to the beaming directions of the projections they represent, while edges are represented by chords. Figure 4.9(a) is a drawing of the graph prior to denoising by the Jaccard index, while Figure 4.9(b) is a drawing of the denoised graph for a particular threshold value. A large portion of the "shortcut" edges were successfully removed. We attribute the seemingly non-random behavior
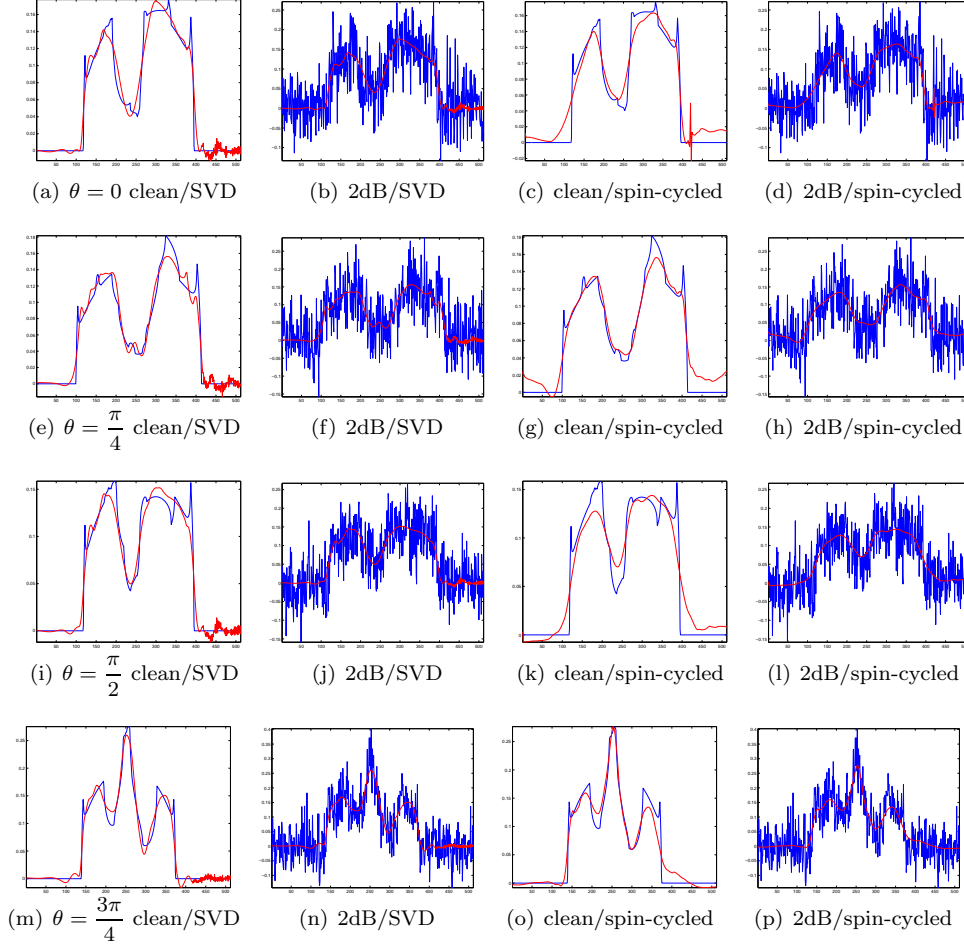
(a) $\theta = 0$ clean/SVD   (b) 2dB/SVD   (c) clean/spin-cycled   (d) 2dB/spin-cycled

(e) $\theta = \frac{\pi}{4}$ clean/SVD   (f) 2dB/SVD   (g) clean/spin-cycled   (h) 2dB/spin-cycled

(i) $\theta = \frac{\pi}{2}$ clean/SVD   (j) 2dB/SVD   (k) clean/spin-cycled   (l) 2dB/spin-cycled

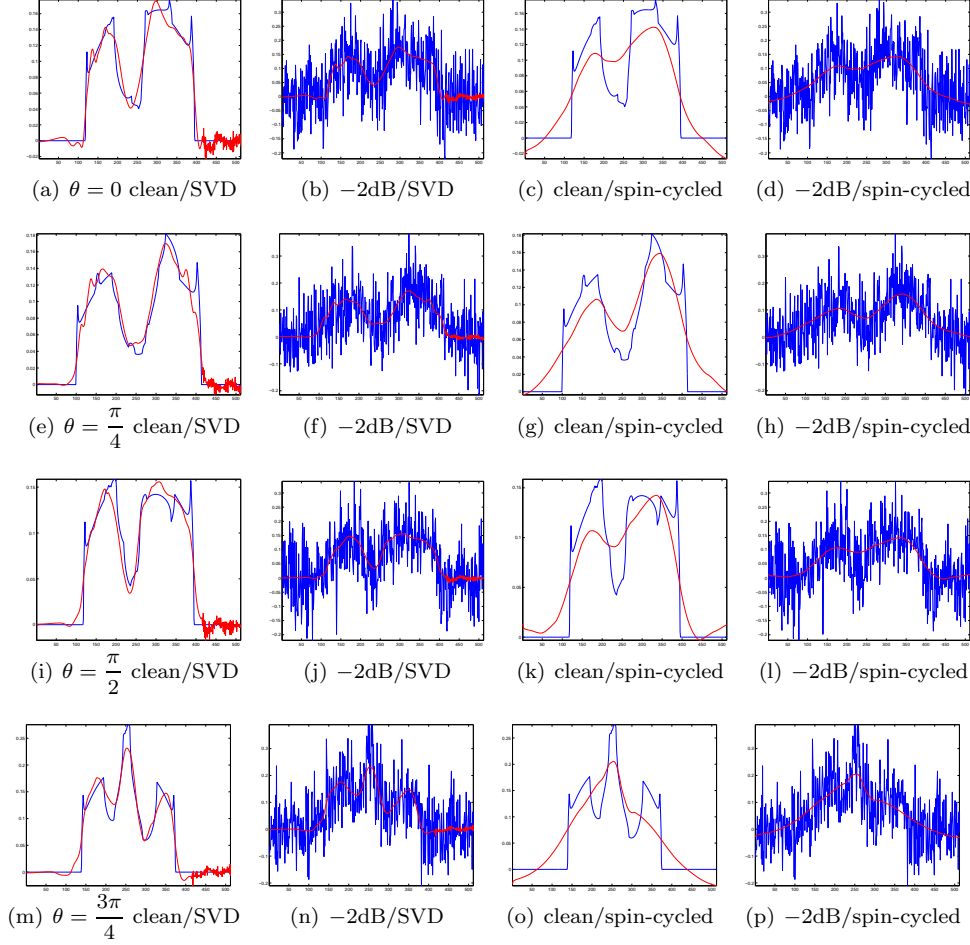(m) $\theta = \frac{3\pi}{4}$ clean/SVD   (n) 2dB/SVD   (o) clean/spin-cycled   (p) 2dB/spin-cycled

FIG. 4.7. *Comparison between SVD denoising and wavelet denoising for four different noisy projections with SNR = 2dB taken at $\theta = 0$, $\frac{\pi}{4}$, $\frac{\pi}{2}$, and $\frac{3\pi}{4}$. (a) clean projection (Blue) and the SVD denoising of the noisy projection (Red); (b) noisy projection (Blue) and the SVD denoising of the noisy projection (Red); (c) clean projection (Blue) and the wavelet denoising of the noisy projection (Red); (d) noisy projection (Blue) and the wavelet denoising of the noisy projection (Red).*

of the shortcut edges that are left in the denoised graph shown in Figure 4.9(b) to the particular shape of the Shepp-Logan phantom, that give rise to somewhat similar projections that are taken at particular different beaming directions.

In Step 4 we computed several diffusion map embeddings corresponding to different values of ϵ chosen in the following way. We first make a histogram of all pairwise distances between the denoised projections. We then look for the "cutoff" distances that separate the smallest 0.5%, 1%, 2% and 4% from all other larger distances. The values of ϵ are chosen as the squared cutoff distances corresponding to the different smallest percentiles. This choice is motivated by viewing ϵ as the squared radius of the ball that contains similar projections for which the similarities in (3.2) are non-negligible.

All in all, the results of our algorithm are summarized in Figures 4.2 showing successful reconstructions at SNR $= -2$dB and SNR $= -3$dB. For higher levels of

(a) $\theta = 0$ clean/SVD   (b) -2dB/SVD   (c) clean/spin-cycled   (d) -2dB/spin-cycled

(e) $\theta = \dfrac{\pi}{4}$ clean/SVD   (f) -2dB/SVD   (g) clean/spin-cycled   (h) -2dB/spin-cycled

(i) $\theta = \dfrac{\pi}{2}$ clean/SVD   (j) -2dB/SVD   (k) clean/spin-cycled   (l) -2dB/spin-cycled

(m) $\theta = \dfrac{3\pi}{4}$ clean/SVD   (n) -2dB/SVD   (o) clean/spin-cycled   (p) -2dB/spin-cycled

FIG. 4.8. *Comparison between SVD denoising and wavelet denoising for four different noisy projections with $SNR = -2dB$ taken at $\theta = 0, \frac{\pi}{4}, \frac{\pi}{2},$ and $\frac{3\pi}{4}$. (a) clean projection (Blue) and the SVD denoising of the noisy projection (Red); (b) noisy projection (Blue) and the SVD denoising of the noisy projection (Red); (c) clean projection (Blue) and the wavelet denoising of the noisy projection (Red); (d) noisy projection (Blue) and the wavelet denoising of the noisy projection (Red).*

noise with SNR below -3dB the algorithm sometimes breaks down and fails to recover the correct ordering of the beaming directions. Of course, the lower the SNR the more likely the algorithm would fail. In particular, the algorithm seems to fail consistently at SNR $= -4$dB.

5. Summary and Discussion. In this paper we introduced a reconstruction method of 2-D objects from noisy tomographic projections taken at unknown beaming directions. The method combines diffusion maps for finding the unknown beaming directions with two preliminary denoising steps. The first denoising step consists of applying the SVD to the data matrix of noisy projections and projecting them onto the subspace spanned by the top left singular vectors, while the second denoising step is based on the Jaccard index and is applied to denoise the edges of the graph of similarities between denoised projections. The additional two denoising steps significantly improve the noise tolerance of the reconstruction method from a

15
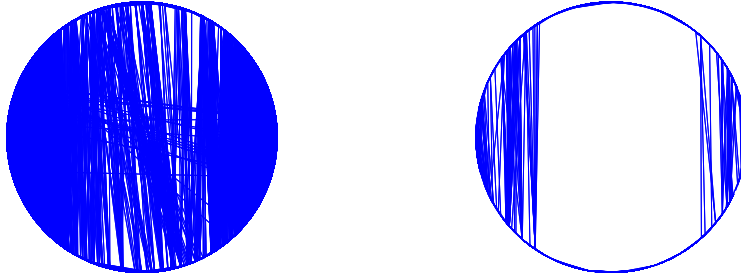
(a) Before Jaccard denoising　　　　　　(b) After Jaccard denoising

FIG. 4.9. *The effect of denoising by the Jaccard index at $SNR = -3dB$ and thresholding edges with Jaccard index below 0.598. (a) The graph $G = (V, E)$ prior to denoising, and (b) Graph $\tilde{G} = (\tilde{V}, \tilde{E})$ after denoising.*

benchmark of $SNR = 2dB$ reported in [6] using diffusion maps and wavelet denoising, to $SNR = -3dB$ obtained here.

We expect the combination of the three tools, namely, SVD, Jaccard index and diffusion maps, to be useful in many other applications that require the organization of high-dimensional data with an underlying non-linear low-dimensional structure. While the diffusion map framework is well adjusted to study and analyze complex data sets, it is somewhat limited by noise that may change both the dimensionality and the topology of the underlying data.

The role of the SVD in our procedure is to denoise the noisy projections by projecting them onto a low-dimensional subspace that captures most of the variability of the data and is adaptive to the data in that sense. By examining the distribution of the singular values, we use only the top singular vectors whose squared singular values reside outside the support of the numerically observed deformed quarter circle distribution. Our numerical experiments show that denoising by SVD outperforms denoising by a pre-chosen basis such as a wavelet basis and we attribute this success to the data adaptivity of the basis of left singular vectors as well as to choosing the correct number of singular vectors as suggested by the eigenvalue histogram. While SVD and PCA are often used as methods for dimensionality reduction, the usage here is merely as a denoising method. Although we use SVD to reduce the data from $n = 512$ dimensions to 6 dimensions, the underlying manifold of the data is one-dimensional, and this non-linear structure is revealed by the diffusion map, which is a non-linear dimensionality reduction method. We comment that in the case of sufficiently large sample size it may also be possible to perform many local SVDs instead of one global SVD and to detect different features for different beaming directions. The combination of SVD with the diffusion map can be recognized as independent component analysis (ICA) [20], while the combination of many local SVDs with the diffusion map can be regarded as non-linear ICA [21].

The second denoising step in our procedure consists of denoising the similarity graph by removing all edges whose Jaccard index is below a certain threshold. The objective in this denoising step is to restore the correct topology of the data by removing falsely edges that shortcut the underlying manifold. The Jaccard index examines whether the short-time diffusion neighborhoods of two vertices are alike, and we remove edges whose endpoints have dissimilar neighborhoods. In that sense,

16

this denoising step is complimentary to the diffusion map that embeds the data points based on their long time diffusion properties.

Finally, we remark that in this paper we assumed that the projections are centered. In [2, 1], the shift problem was considered to handle the uncertain nature of the imaging system, that can be formulated as follows. During the imaging process, suppose the image was shifted by h in the direction θ, that is, the imaged object is $g(x) = f(x + h\theta)$, then the projection we get is $R_\theta(g)(r) = R_\theta(f)(r + h)$, which is also equivalent to a shift by $-h$ of the camera. We can still use the diffusion map framework by computing the translational-invariant distances that are given by

$$d_{ij} = \min_{h \in \mathbb{R}} \|R_i - T_h R_j\|_{\ell_2},$$

where T_h is the translation operator over $\mathbb{R}$ satisfying $T_h f(x) = f(x + h)$, and R_i denotes $R_{\theta_i} f$. These distances factor out the one degree of freedom of translation, so that diffusion map should recover the correct parameterizations of the closed curve as before. However, due to noise, the translational alignment of two projections taken at dissimilar beaming directions can have a small translational-invariant distance. In order to detect such outliers, we can apply a similar technique to the angular synchronization technique that we recently suggested for the class averaging problem in cryo-EM [22]. Specifically, if the shift of projection R_i is $h_i \in \mathbb{R}$, then the relative shift of projections R_i and R_j is $h_i - h_j$ and we can estimate this relative shift by computing

$$h_{ij} = \operatorname*{argmin}_{h \in \mathbb{R}} \|R_i - T_h R_j\|_{\ell_2}.$$

Note that the h_{ij}'s can take values over the non-compact group $\mathbb{R}$, which we therefore compactify by the many-to-one mapping from $\mathbb{R} \to S^1$ given by $x \mapsto e^{\iota x/\delta}$, where δ is a parameter to be carefully chosen. As explained in [22], from the top eigenvectors of the Hermitian matrix $\boldsymbol{H}$ whose entries are given by $H_{ij} = e^{-d_{ij}^2/2\epsilon} e^{ih_{ij}/\delta}$ we can reveal the shifts of all projections as well as the parametrization of the closed curve.

REFERENCES

[1] S. Basu and Y. Bresler. Feasibility of tomography with unknown view angles. *IEEE Transactions on Image Processing*, 9(6):1107–1122, June 2000.

[2] S. Basu and Y. Bresler. Uniqueness of tomography with unknown view angles. *IEEE Transactions on Image Processing*, 9(6):1094–1106, June 2000.

[3] M. Belkin and P. Niyogi. Towards a theoretical foundation for laplacian-based manifold methods. In *Proceedings of the 18th Conference on Learning Theory (COLT)*, pages 486–500, 2005.

[4] R. R. Coifman and D. Donoho. Translation-invariant de-noising. In *Wavelets and Statistics*, pages 125–150. Springer-Verlag, New York, 1985.

[5] R. R. Coifman and S. Lafon. Diffusion maps. *Applied and Computational Harmonic Analysis*, 21(1):5 – 30, 2006.

[6] R. R. Coifman, Y. Shkolnisky, F. J. Sigworth, and A. Singer. Graph laplacian tomography from unknown random projections. *IEEE Trans. Image Proc.*, 17(10):1891–1899, Oct. 2008.

[7] S. R. Deans. *The Radon Transform and Some of its Applications*. Krieger, 1993.

[8] D. S. Goldberg and F. P. Roth. Assessing experimentally derived interactions in a small world. *Proceedings of the National Academy of Sciences of the United States of America*, 100(8):4372–4376, 2003.

[9] G. H. Golub and C. F. Van Loan. *Matrix computations*. Johns Hopkins University Press, Baltimore, 3rd edition, 1996.

[10] M. Hein, J. Audibert, and U. von Luxburg. From graphs to manifolds - weak and strong pointwise consistency of graph laplacians. In *Proceedings of the 18th Conference on Learning Theory (COLT)*, pages 470–485, 2005.

[11] M. D. Humphries and K. Gurney. Network small-world-ness: A quantitative method for determining canonical network equivalence. *PLoS ONE*, 3(4):e0002051, 2008.

[12] I. M. Johnstone. High dimensional statistical inference and random matrices. *arXiv:math/0611589v1*, 2006.

[13] A. C. Kak and M. Slaney. *Principles of Computerized Tomographic Imaging*. SIAM, 2001.

[14] S. Lafon. *Diffusion maps and geometric harmonics*. PhD thesis, Yale University, 2004.

[15] V. A. Marcenko and L. A. Pastur. Distribution of eigenvalues for some sets of random matrices. *Math USSR Sb*, 1(4):457–483, 1967.

[16] B. Nadler, S. Lafon, R. R. Coifman, and I. G. Kevrekidis. Diffusion maps, spectral clustering and reaction coordinates of dynamical systems. *Applied and Computational Harmonic Analysis*, 21(1):113–127, 2006.

[17] F. Natter and F. Wubbeling. *Mathematical Methods in Image Reconstruction*. SIAM, 2001.

[18] F. Natterer. *The Mathematics of Computerized Tomography*. SIAM, 2001.

[19] A. Singer. From graph to manifold Laplacian: The convergence rate. *Applied and Computational Harmonic Analysis*, 21(1):128–134, 2006.

[20] A. Singer. Spectral independent component analysis. *Applied and Computational Harmonic Analysis*, 21(1):135–144, 2006.

[21] A. Singer and R. R. Coifman. Non linear independent component analysis with diffusion maps. *Applied and Computational Harmonic Analysis*, 25(2):226–239, 2008.

[22] A. Singer and Y. Shkolnisky. Angular synchronization by eigenvectors and semidefinite programming: Analysis and application to class averaging in cryo-electron microscopy. *submitted*.

[23] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(1):440–442, 1998.