Two-Dimensional Tomography from Noisy Projections Taken at Unknown Random Directions[*]

A. Singer[†] and H.-T. Wu[‡]

[†]Department of Mathematics and PACM, Princeton University, Princeton, NJ 08544-1000 (amits@math.princeton.edu).

[‡]Department of Mathematics, Princeton University, Princeton, NJ 08544-1000 (hauwu@math.princeton.edu).

Abstract

Computerized tomography 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. Another difficulty arises from the fact that the projections are often contaminated by noise, practically limiting all current methods, including the recently proposed diffusion map approach. In this paper, we introduce two denoising steps that allow reconstructions at much lower signal-to-noise ratios (SNRs) when combined with the diffusion map framework. In the first denoising step we use principal component analysis (PCA) together with classical Wiener filtering to derive an asymptotically optimal linear filter. In the second step, we denoise the graph of similarities between the filtered projections using a network analysis measure such as the Jaccard index. Using this combination of PCA, Wiener filtering, graph denoising, and diffusion maps, we are able to reconstruct the two-dimensional (2-D) Shepp–Logan phantom from simulative noisy projections at SNRs well below their currently reported threshold values. We also report the results of a numerical experiment corresponding to an abdominal CT. Although the focus of this paper is the 2-D CT reconstruction problem, we believe that the combination of PCA, Wiener filtering, graph denoising, and diffusion maps is potentially useful in other signal processing and image analysis applications.

1. Introduction

Transmission computerized tomography (CT) currently is a standard method for obtaining internal structures nondestructively, which is routinely used in medical imaging [13, 24, 31, 32]. The classical two-dimensional (2-D) 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

---

$$R_\theta f(s) = \int_{x \cdot \theta = s} f(x)\, dx = \int_{-\infty}^{\infty} f\left(s\theta + r\theta^\perp\right) dr,$$

where $\theta \in S^1$ is perpendicular to the beaming direction $\theta^\perp \in S^1$ (S^1 is the unit circle), and $s \in \mathbb{R}$. The reconstruction of f from its Radon transform $R_\theta f$ is made possible due to the Fourier projection slice-theorem that relates the one-dimensional (1-D) Fourier transform $\widehat{R_\theta f}$ of the Radon transform to the 2-D Fourier transform $\hat{f}$ of the function [13, 24, 31, 32]:

$$\widehat{R_\theta f}(\xi) = \hat{f}(\xi\theta) \quad \text{for all} \quad \xi \in \mathbb{R}. \quad (1.1)$$

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 cases, one is given samples of the Radon transform $R_{\theta_i}(\cdot)$ 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 s is usually known and is dictated by the physical setting of the acquisition process; for example, if the detectors are equally spaced, then the values of s correspond to the location of the detectors along the line of detectors, while the origin may be set at the center of mass. An alternative method for estimating the shifts will be discussed in section 8.

In this paper we address the reconstruction problem for the 2-D parallel-beam model with unknown acquisition directions. Formally, we consider the following problem: Given n projection vectors $(R_{\theta_i}f(s_1), R_{\theta_i}f(s_2), \ldots, R_{\theta_i}f(s_p))$ taken at unknown directions $\{\theta_i\}_{i=1}^n$ that were randomly drawn from the uniform distribution over S^1 and given that $s_1, s_2, \ldots, s_p$ are fixed p equally spaced pixels in s, find the underlying density function f of the object. The observed n projection vectors are often contaminated by noise; in such cases, the problem is to find the beaming directions of the noisy projections.

This 2-D reconstruction problem from unknown directions was previously considered by Basu and Bresler in [2, 3]. In particular, in [3] they derive conditions for the existence of unique reconstruction from unknown directions and shifts. The recovery problem is formulated as a nonlinear system using the Helgason–Ludwig consistency conditions, which are used to derive uniqueness conditions. Stability conditions for the angle recovery problem under deterministic and stochastic perturbation models are derived in [2], 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 [2], and it consists of three steps: (1) initial direction estimation; (2) direction ordering; and (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 [12], where the ordering was obtained by a proper application of the diffusion map framework [11, 27]. Specifically, the method of [12] 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 [10].

In this paper, we combine the diffusion map approach of [12] with two other denoising techniques that together allow reconstructions at much lower SNRs. The first denoising step consists of using principal component analysis (PCA) together with the classical Wiener filtering approach for minimizing the mean squared error. The advantage of the basis found by PCA over the wavelet basis that was used in [12] is in its adaptivity to the data. Empirically, we have observed that for many images, a few principal components capture most of the variability of their Radon transform, and projecting the 1-D projections onto this basis diminishes the noise while capturing most of the signal features. We derive an asymptotically optimal linear filter in the limit $n, p \rightarrow \infty$ and with $p/n = \gamma$ fixed by employing the Wiener filtering approach together with recent results concerning PCA in high dimensions (see, e.g., [22]). Our filter requires us to estimate the noise variance, the number of components, and their corresponding eigenvalues. To that end, we use the method of Kritchman and Nadler [25, 26].

We then compute the Euclidean distances between all pairs of filtered projections and use these distances to construct a matrix of their pairwise similarities. Our second denoising step consists of further denoising the similarity matrix using a network analysis measure such as the Jaccard index (see, e.g., [18], where the Jaccard index was used to denoise protein interaction maps). When two projections share a similar beaming direction, then it is expected not only that the similarity between the two of them would be significant, but also that their similarity to 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.

We performed numerical experiments for testing this combination of PCA, Wiener filtering, graph denoising, and diffusion maps, and were able to reconstruct the 2-D Shepp–Logan phantom from simulative noisy projections at SNRs well below their currently reported threshold values [12]. We also report the results of a numerical experiment corresponding to an abdominal CT. Although the focus of this paper is the 2-D CT reconstruction problem, we believe that such a combination of PCA, Wiener filtering, graph denoising, and diffusion maps has the potential to become a useful tool for other signal processing and image analysis applications. In particular, we expect the asymptotically optimal linear filter that we derive here (see (4.30)) to be useful in many other applications, regardless of the other steps of our proposed algorithm.

The paper is organized in the following way. In section 2 we discuss the underlying geometry of the Radon projections as a 1-D closed curve in a high-dimensional ambient space. In section 3 we give a brief introduction to the diffusion map method and review its application to the 2-D reconstruction problem as proposed in [12]. In section 4 we develop the PCA-Wiener filter for the optimal weighted projection. In section 5 we explain the usage of the Jaccard index for denoising the graph of similarities. In section 6 we summarize the

algorithm for solving the 2-D reconstruction problem. The algorithm has only two free parameters, and we describe a method for choosing them automatically. In section 7 we detail the results of our numerical experiments. Finally, section 8 is a summary and discussion.

2. Underlying geometry

The following proposition is given as an exercise in Epstein's book on the mathematics of medical imaging [15, Exercise 6.6.1, p. 215].

Proposition 2.1

Suppose that $f \in L^2\left(\mathbb{R}^2\right)$ *and that f vanishes outside the unit disk. Then,* $R_\theta f$ *is in* $L^2(\mathbb{R})$ *for all θ, and* $\left\| R_{\theta_2} f - R_{\theta_1} f \right\|_{L^2(\ )}$ *tends to zero as* θ_2 *approaches* θ_1*. In other words, the map* $\theta \mapsto R_\theta f$ *is a continuous map from* S^1 *into* $L^2(\mathbb{R})$*.*

Proof. See Appendix A.

From Proposition 2.1 it follows that the image of the Radon transform is a compact and connected continuous curve $\mathscr{C}$ in $L^2(\mathbb{R})$ parameterized by θ, whenever f is in $L^2\left(\mathbb{R}^2\right)$ and has compact support. Note that the function *f* is not required to be continuous. For example, the function *f* corresponding to the Shepp–Logan phantom is discontinuous, yet Proposition 2.1 guarantees that its Radon transform is a continuous function of θ, a fact that can also be verified by a direct calculation.

We note that it is possible for the closed curve $\mathscr{C}$ to intersect with itself. We refer to closed curves without self-intersections as simple curves. For example, if *f* has some axis of symmetry, then the closed curve $\mathscr{C}$ intersects with itself and is therefore nonsimple. Symmetry, however, is not the only case for which $\mathscr{C}$ is nonsimple. There are other nonsymmetric functions that give rise to nonsimple curves. From the slice-theorem it follows that $\mathscr{C}$ is self-intersecting iff there are $\theta_1 = \theta_2$ such that $\hat{f}(\xi\theta_1) = \hat{f}(\xi\theta_2)$ for all $\xi \in \mathbb{R}$ (notice that for $f \in L^2\left(\mathbb{R}^2\right)$ with compact support, *f* is also in $L^1\left(\mathbb{R}^2\right)$, and as a result $\hat{f}$ is continuous). In what follows we assume that $f \in L^2\left(\mathbb{R}^2\right)$ has compact support and that its curve $\mathscr{C}$ is simple.

The measurements are assumed to be discrete samples of the Radon transform of *f*. Every projection vector $(R_\theta f(s_1), R_\theta f(s_2), \ldots, R_\theta f(s_p))$ can be viewed as a point in $\mathbb{R}^p$. When varying the beaming direction $\theta^\perp$ over S^1, the projection vectors sample a closed curve, denoted $\mathscr{C}^p$ in $\mathbb{R}^p$. The discretization operator can be viewed as a projection operator from $L^2(\mathbb{R})$ to its finite-dimensional subspace $\mathbb{R}^p$. The continuity of the projection operator and the fact that $\mathscr{C}$ is a compact, continuous, closed curve in $L^2(\mathbb{R})$ imply that $\mathscr{C}^p$ is also a compact, continuous, closed curve in $\mathbb{R}^p$. In the limit of an infinitely large number of discretization points $p \to \infty$, the projections sample a simple closed curve $\mathscr{C}$ in $L^2(\mathbb{R})$. The collection of *n* projection vectors $(R_{\theta_i} f(s_1), \ldots, R_{\theta_i} f(s_p))$ ($i = 1, \ldots, n$) are therefore *n* sampling points of a closed curve in $\mathbb{R}^p$ that approximates a simple closed curve in $L^2(\mathbb{R})$. In order to have a unique solution for the finite-dimensional problem, we need to assume that the closed curve $\mathscr{C}^p$ is also simple. Note that noise contamination perturbs $\mathscr{C}^p$, and its effect is examined later.

3. Diffusion maps

Diffusion mapping is a nonlinear dimensionality reduction technique [11, 27], whose application to the reconstruction problem at hand was studied in [12]. As discussed above, although the sampled projections are points in a high-dimensional Euclidean space, they are restricted to a 1-D closed curve. This curve may have a complicated nonlinear structure that may not be captured by projecting it linearly onto a low-dimensional subspace. Unlike linear methods such as PCA, the diffusion map technique successfully finds the correct parametrization of the nonlinear curve. In this section, we give a brief description of the diffusion map technique and discuss some of its properties and limitations. Readers who are familiar with [12] are advised to proceed to the next section.

We now outline the steps of the diffusion map algorithm. Suppose $x_1, \ldots, x_n \in \mathbb{R}^p$ 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 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 $\varepsilon > 0$ in the following way:

$$W_{ij} = K\left(\frac{\|x_i - x_j\|_p}{\sqrt{2\epsilon}}\right) \quad \text{for} \quad i, j = 1, \ldots, n. \quad (3.1)$$

Clearly, the matrix W is symmetric. The second step is to normalize W into a probability transition matrix A of a random walk on the data points by letting

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

where D is a diagonal matrix whose entries are given by

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

The matrix A is similar to the symmetric matrix $D^{-1/2}WD^{-1/2}$ through

$$A = D^{-1/2}\left(D^{-1/2}WD^{-1/2}\right)D^{1/2}.$$

Therefore, A has a complete set of eigenvectors $\varphi_0, \varphi_1, \ldots, \varphi_{n-1}$ with corresponding eigenvalues $1 = \lambda_0 \quad \lambda_1 \quad \cdots \quad \lambda_{n-1} \quad -1$, where $\varphi_0 = (1, 1 \ldots, 1)^T$. The eigenvectors $\varphi_0, \ldots, \varphi_{n-1}$ are vectors in $\mathbb{R}^n$, and we denote their ith element by $\varphi_0(i), \varphi_1(i), \ldots, \varphi_{n-1}(i)$. Moreover, for positive definite kernel functions in the sense of Bochner (see, e.g., [35, pp. 329–332]), the matrix W is positive definite, and as a consequence all the eigenvalues of A are positive. For example, by Bochner's theorem, the Gaussian kernel $K(u) = \exp\{-u^2/2\}$ is positive definite, because its Fourier transform is positive. In the last step, the data points are embedded in $\mathbb{R}^m$ ($m \quad n-1$) via

$$\Phi_t^m : x_i \mapsto \left(\lambda_1^t \phi_1(i), \ldots, \lambda_m^t \phi_m(i)\right) \quad \text{for} \quad i = 1, \ldots, n, \quad (3.2)$$

where $t > 0$ is a parameter. The map Φ_t^m is known as the diffusion map.

Whenever the data points are sampled from a *d*-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 $\varepsilon \to 0$, provided that $n\varepsilon^{d/2+1}$. This convergence can be stated in terms of the normalized graph Laplacian *L*, which is defined as

$$L = I - A,$$

where *I* is the $n \times n$ identity matrix. In the case where the data points $\{x_i\}_{i=1}^n$ are independent samples from a probability density function $p(x)$ whose support is a *d*-dimensional manifold $\mathcal{M}^d$, the graph Laplacian converges pointwise to the Fokker–Planck operator, as we have the following proposition [4, 20, 27, 36]: if $f : \mathcal{M}^d \to \mathbb{R}$ is a smooth function (e.g., $f \in C^3(\mathcal{M})$), then with high probability

$$\frac{1}{\epsilon}\sum_{j=1}^n L_{ij} f(x_j) = \frac{1}{2}\Delta_M f(x_i) + \nabla U(x_i) \cdot \nabla f(x_i) + O\left(\epsilon + \frac{1}{n^{1/2}\epsilon^{1/2+d/4}}\right), \quad (3.3)$$

where Δ_M is the Laplace–Beltrami operator on $\mathcal{M}^d$ and the potential term *U* is given by $U(x) = -2\log p(x)$. In the special case of a uniform density (*p* and *U* are constants, and ∇U vanishes) the limiting operator is merely the Laplace–Beltrami operator. The error consists of two terms: a bias term $O(\varepsilon)$ and a variance term that decreases as $1/\sqrt{n}$, but also depends on ε. Balancing the two terms can lead to an optimal choice of the parameter ε as a function of the number of points *n*. In the case of uniform sampling, Belkin and Niyogi [5] have proved spectral convergence, that is, they showed that the eigenvectors of the normalized graph Laplacian converge almost surely in the L^2 sense to the eigenfunctions of the Laplace–Beltrami operator on the manifold, which is stronger than the pointwise convergence stated in (3.3). We refer the reader to Theorem 2.1 in [5] for the precise conditions and statement of their theorem.

It is possible to recover the Laplace–Beltrami operator also for nonuniform sampling processes using a different normalization of the similarity matrix [11]. Indeed, if one defines the matrix $\tilde{W}$ as $\tilde{W} = D^{-1}WD^{-1}$, the matrix $\tilde{D}$ as a diagonal matrix with $\tilde{D}_{ii} = \sum_{j=1}^n \tilde{W}_{ij}$, and the matrix $\tilde{A}$ as $\tilde{A} = \tilde{D}^{-1}\tilde{W}$, then $\tilde{L} = I - \tilde{A}$ converges pointwise to the Laplace–Beltrami operator even if the sampling process is nonuniform. Belkin and Niyogi [5, last paragraph of section 1 and first paragraph of section 2] observed that the arguments in their paper are likely to allow one to show the spectral convergence of *L* to the Laplace–Beltrami operator, although to the best of our knowledge no such proof currently exists in the literature. The proof of the spectral convergence in the nonuniform case is beyond the scope of this paper.

In our case, the data points are the projections which are restricted to the closed curve $\mathscr{C}^p$. Although the beaming directions are assumed to be uniformly distributed over S^1, the projections are not necessarily uniformly distributed over $\mathscr{C}^p$, due to the nontrivial Jacobian of the transformation from S^1 to $L^2(\mathbb{R})$ (and to $\mathbb{R}^p$) that takes θ to $R_\theta f$ (and its discretization in $\mathbb{R}^p$). Assuming that the spectral convergence result by Belkin and Niyogi also holds in the case of nonuniform sampling, the eigenvectors of *L* computed by the diffusion map will therefore be discrete approximations of the eigenfunctions of the Fokker–Planck operator over $\mathscr{C}^p$. If instead we apply the normalization that leads to the Laplace–Beltrami operator, then (again, assuming that the result of Belkin and Niyogi holds also for this particular normalization) the computed eigenvectors will be discrete approximations of the eigenfunctions of the Laplace–Beltrami operator over $\mathscr{C}^p$ which are nothing but the trigonometric functions of the normalized arclength *l* given by 1, $\sin(2\pi jl)$, $\cos(2\pi jl)$, $j = 1$,

2, . . . , where the arclength l is normalized such that the total arclength of $\mathscr{C}^p$ is 1. In particular, a diffusion map with $m = 2$ in (3.2) that uses the first two nontrivial eigenfunctions $\sin(2\pi l)$ and $\cos(2\pi l)$ embeds $\mathscr{C}^p$ onto the unit circle S^1 in $\mathbb{R}^2$ (note that the eigenvalues associated with these eigenfunctions are equal). In practice, the eigenvectors and their corresponding eigenvalues are only an approximation of their continuous counterparts, and so the embedding may not coincide exactly with the unit circle but can be more "wiggly."

Suppose we compute the first and second nontrivial eigenvectors of $\tilde{L}$ and denote them by $\tilde{\phi}_1$ and $\tilde{\phi}_2$. The eigenvectors $\tilde{\phi}_1$ and $\tilde{\phi}_2$ are vectors of length n, and from the convergence theorem stated above it follows that their ith components $\tilde{\phi}_1(i)$ and $\tilde{\phi}_2(i)$ approximate the corresponding values of the eigenfunctions of the Laplace–Beltrami operator over $\mathscr{C}^p$ at the point x_i. Since the first and second eigenfunctions of the Laplace–Beltrami operator are known to be $\sin(2\pi l)$ and $\cos(2\pi l)$, it follows that $\tilde{\phi}_1(i) \approx \sin(2\pi l(x_i))$ and $\tilde{\phi}_2(i) \approx \cos(2\pi l(x_i))$, where $l(x_i)$ is a particular choice of the normalized arclength function. We can therefore estimate the ordering of the (perpendicular) beaming directions $\theta_1, \ldots, \theta_n \in S^1$ by the ordering of the phases

$$\hat{\theta}_i = \frac{\left(\tilde{\phi}_1(i), \tilde{\phi}_2(i)\right)}{\sqrt{\tilde{\phi}_1^2(i) + \tilde{\phi}_2^2(i)}} \quad \text{for} \quad i = 1, 2, \ldots, n.$$

In other words, the embedding $x_i \mapsto \left(\tilde{\phi}_1(i), \tilde{\phi}_2(i)\right)$ practically solves the problem, up to some nonlinear ("warp") transformation which is due to the arclength function $l(x)$. Still, the monotonicity of the arclength function ensures that the ordering of the beaming directions is estimated correctly. Once the ordering is revealed, the beaming directions are estimated by equally spacing them over the unit circle. This estimator is consistent due to the underlying assumption about the uniform distribution of the beaming directions.

We remark that it is also possible to use the eigenvectors φ_1 and φ_2 of L in order to estimate the beaming directions, despite the fact that L approximates the Fokker–Planck operator and that the eigenfunctions of the Fokker–Planck operator are no longer the sine and cosine functions. As discussed in [12], the Fokker–Planck operator over a simple closed curve is a Sturm–Liouville operator with periodic boundary conditions and positive coefficients, and from the classical Sturm–Liouville theory of [9] it follows that the embedding of $\mathscr{C}^p$ into $\mathbb{R}^2$ given by $x \mapsto (\phi_1, \phi_2)$ also circles the origin exactly once in such a manner that the angle is monotonic (here, φ_1 and φ_2 are eigenfunctions rather than eigenvectors). In other words, upon writing the embedding in polar coordinates

$$(\phi_1(l), \phi_2(l)) = r(l) e^{i\varphi(l)}, \quad l \in [0, 1],$$

the argument $\varphi(l)$ is a monotonic function of l, with $\varphi(0) = 0$, $\varphi(1) = 2\pi$. Despite the fact that the explicit form of the eigenfunctions is no longer available, the graph Laplacian embedding reveals the ordering of the projections through the angle ϕ_i attached to x_i. Once the order of the beaming directions is revealed, they are estimated by spreading them evenly over S^1 (see also section III in [12]). This is a consistent estimator of the beaming directions if they are assumed to be uniformly distributed.

Note that the correct ordering of the beaming directions is revealed even if the beaming directions are not uniformly distributed. However, it is impossible to accurately estimate the beaming directions without prior knowledge of the underlying nonuniform density. This is perhaps the main difference between the 2-D reconstruction problem considered here and its three-dimensional (3-D) analogue that corresponds to the reconstruction of macromolecules from their 2-D tomographic images taken at unknown random directions as in cryoelectron microscopy images [17]. In the 3-D problem, the slice-theorem implies that any two central slices share a common line of intersection that can be used to find the unknown imaging directions even when they are not uniformly distributed. This implication is of great importance since the imaging directions are nonuniform whenever the biological object has a preferred orientation, as is often the case. However, in the 2-D problem, the implication of the slice-theorem is trivial, as it implies only that any two line projections intersect at a point (the origin in the Fourier domain). This trivial intersection cannot be used to improve the estimate of the beaming directions beyond their ordering.

We now give a detailed explanation for the impossibility of accurately estimating the beaming directions if their distribution is not known in advance and is not necessarily uniform. To that end, suppose that $f(x, y)$ is the image to be reconstructed and $\hat{f}(\omega_x, \omega_y)$ is its Fourier transform. We assume that f is compactly supported and is in L^2 so that both its Radon and Fourier transforms are continuous. In light of the projection slice-theorem, it is convenient to regard the frequency $\omega = (\omega_x, \omega_y)$ as a complex number and its representation in polar coordinates as $\omega = re^{i\theta}$, where r is the modulus of the frequency and θ is the phase. Now, suppose that $g : S^1 \rightarrow S^1$ is a continuous, 1-to-1, and onto mapping from the unit circle to itself, also satisfying $g(-\omega) = -g(\omega)$ for all $\omega \in S^1$. The last condition implies that g maps antipodal points to antipodal points. As a result, g can be regarded as a continuous, 1-to-1, and onto mapping from the projective space $\mathbb{R}P^1$ to itself. That is, g maps central lines to central lines. The rigid mapping that rotates the circle by a fixed angle is an example of such a mapping. However, there are many nonrigid (nonlinear) transformations that "warp" the circle. Note that only rigid transformations preserve the underlying distribution of the beaming directions. For example, the uniform distribution remains uniform under rotation. Nonrigid transformations change the distribution. For example, if g is differentiable, then the density is changed according to the derivative of g. See Figure 1 for the following example of such a nonrigid transformation g:

$$
g\left(e^{i\theta}\right) = \begin{cases} e^{\frac{i\pi(1-\cos(\theta))}{2}} & \text{for} \quad 0 \leq \theta < \pi, \\ -e^{\frac{i\pi(1-\cos(\theta-\pi))}{2}} & \text{for} \quad \pi \leq \theta < 2\pi. \end{cases}
$$

Consider now the function $h(x, y)$ whose Fourier transform $\hat{h}$ is given by $\hat{h}(\omega_x, \omega_y) = \hat{h}(re^{i\theta}) = \hat{f}(rg(e^{i\theta}))$. Clearly, $\hat{h}$ and $\hat{f}$ agree on central lines (though possibly at different angles). Combined with the slice-theorem, this means that the set of line projections of h is the same as the set of lines projections of f. Of course, this does not mean that the Radon transform of h equals the Radon transform of f. However, this establishes that f and h are indistinguishable given just samples of their Radon transforms at unknown angles, unless the distribution of the viewing directions is known in advance, such as uniform.

We comment that it is also possible to search for the underlying closed curve directly, an approach that boils down to solving the traveling salesman problem (TSP) in high dimensions. Although the TSP is NP hard, there are many algorithms (heuristics and approximation algorithms) for finding its solution. The diffusion map approach that we invoke here is an efficient way of finding a solution to this specific case of TSP (which is far from being the general TSP problem, as the underlying geometry corresponds to a closed

curve). Due to noise, however, the original Euclidean distances between the Radon projections are not so meaningful for low SNRs. Thus, TSP solvers face the same difficulties as the diffusion mapping approach faces, emphasizing the importance of denoising.

Indeed, noise is the main limitation of the diffusion mapping approach, since the measurement noise causes 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 simple closed curve. This is conceptually illustrated in Figure 2. It is reported in [12] that for noisy projections of the Shepp–Logan phantom, the diffusion map approach succeeds only for an SNR above 10.5dB (the SNR is later defined in (7.1)). It is further reported in [12] that applying classical wavelet noise filtering techniques prior to the diffusion mapping step allowed successful reconstructions for SNR above 2dB. This significant improvement encourages us to explore other denoising techniques, with our main goal being to further improve the robustness of the algorithm to noise. We proceed to show that it is possible to significantly increase the robustness to noise by applying PCA, Wiener filtering, and graph denoising prior to the diffusion mapping step.

4. PCA and Wiener filtering

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 *W*. A good denoising procedure will retain most characteristic features of the true signal while diminishing the contribution of noise. Thus, it is more beneficial to construct a similarity matrix *W* from the properly denoised projections. For example, in [12], 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 prechosen 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 PCA. The main contribution of this section is the derivation of the linear filtering procedure (4.29), which we prove to be asymptotically optimal in the mean squared error sense in the limit $n, p \to \infty$ and with $p/n = \gamma$ fixed.

4.1. PCA: The basics

PCA is a linear dimensionality reduction method dating back to 1901 [34] and is one of the most useful techniques in data analysis. Indeed, usually the first step in the analysis of most types of high-dimensional data is performing PCA, and the situation here is no different. PCA finds an orthogonal transformation which maps the data to a new coordinate system such that the greatest variance by any projection of the data comes to lie on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on. Given a *p*-dimensional random variable *x* with mean $\mathbb{E}x = \mu \, (\in \mathbb{R}^p)$ and a $p \times p$ covariance matrix $\mathbb{E}(x - \mu)(x - \mu)^T = \Sigma$, the solution to the maximization problem

$$\max_{\|u\|=1} \quad \mathrm{Var}\left(u^T x\right) = \max_{\|u\|=1} \mathbb{E}\left[u^T (x - \mu)(x - \mu)^T u\right] = \max_{\|u\|=1} u^T \Sigma u$$

is given by the top eigenvector u_1 of Σ satisfying $\Sigma u_1 = \lambda_1 u_1$. The eigenvalues of Σ denoted $\lambda_1 \quad \lambda_2 \quad \cdots \quad \lambda_p \quad 0$ are also known as the population eigenvalues. Similarly, the first *d* eigenvectors $u_1, \ldots, u_d$ corresponding to $\lambda_1, \ldots, \lambda_d$ are the solution to

$$\max_{u_1,\dots,u_d : u_i^T u_j = \delta_{ij}} \sum_{i=1}^{d} u_i^T \Sigma u_i.$$

The *d*-dimensional subspace spanned by the first *d* eigenvectors is therefore optimal in the sense that it captures the most variability of the data. From an alternative point of view, the first principal component u_1 is the axis which minimizes the sum of squared distances from the mean shifted data points to their orthogonal projections on that axis:

$$\min_{\|u\|=1} \mathbb{E}\| \left(I - uu^T \right) (x - \mu) \|^2.$$

In general, the first *d* principal components span a *d*-dimensional linear subspace that minimizes the mean square of the approximation error of the data by its orthogonal projection (after mean shift).

Many real world data sets, though possibly complex and nonlinear, are well approximated by a low-dimensional subspace, hence the usefulness of PCA. In our case, Figure 7(a) shows the top 20 eigenvalues of the sample covariance matrix for projections with very high SNR of the Shepp–Logan phantom. The rapid decay of the eigenvalues is evident. As a result, projecting the data onto the subspace spanned by the first 15 or so principal components results in a very small approximation error. This means that although the curve $\mathscr{C}^p$ is highly nonlinear, its deviations from this 15-dimensional linear subspace in $\mathbb{R}^p$ are small. We illustrate the first 8 principal components in Figure 8. Since white noise is evenly distributed over all components, projecting the data onto that 15-dimensional subspace would decrease the ℓ_2 energy of the noise by a factor of $15/p$ while almost completely preserving the true signal features. This has, of course, a most desirable denoising effect.

4.2. Linear Wiener filtering

We see in the example of the Shepp–Logan phantom that PCA can be used to denoise the 1-D projections. But what about other phantoms, and how many components should be used in general? And, would it be perhaps more beneficial to use a weighted projection? To answer these questions, we revisit the classical linear Wiener filtering approach [38] for finding the optimal weighted projection.

Consider the additive noise model, where the noisy observation *y* is given by

$$y = x + \xi, \quad (4.1)$$

where *x* is the underlying clean signal and ξ is an additive white noise.[1] We assume that $\mathbb{E}x = \mu$ and $\mathbb{E}(x - \mu)(x - \mu)^T = \Sigma$, while $\mathbb{E}\xi = 0$ and $\mathbb{E}\xi\xi^T = \sigma^2 I_{p \times p}$. Given the noisy observation *y*, we want to derive an estimator $\hat{x}$ for its unknown underlying clean signal *x*. A possible optimality criterion for deriving the estimator $\hat{x}$ is the minimum conditional mean squared error

---

[1]In some situations errors can be correlated; in such cases, a standard procedure consists of prewhitening the noise prior to the PCA step. Notice that the noise is not assumed to be Gaussian, just white. Later on, in subsections 4.3 and 4.4 we also require the noise to have finite fourth moments.

$$\min_{\widehat{x}} \mathbb{E}\left[\left\|x - \widehat{x}\right\|^2 | y\right]. \quad (4.2)$$

The well-known minimizer of (4.2) is the conditional expectation

$$\widehat{x} = \mathbb{E}\left[x|y\right]. \quad (4.3)$$

However, computation of the conditional expectation (4.3) requires the knowledge of the probability distribution of x, which is not available to us: Only the second order statistics of x is assumed to be known.[2] An alternative approach is therefore required. One of the standard alternative approaches consists of two modifications: first, replacing the conditional mean squared error in (4.2) by the mean squared error (i.e., removing the conditioning on y), and second, restricting the minimizer of (4.2) to a smaller class of linear estimators instead of all possible estimators. That is, the estimator x îs restricted to be of the form

$$\widehat{x} = \mu + H\left(y - \mu\right), \quad (4.4)$$

where H is a $p \times p$ matrix which is the solution to the minimization problem

$$\min_{H \in \quad p \times p} \mathbb{E}\left\|x - \left(\mu + H\left(y - \mu\right)\right)\right\|^2. \quad (4.5)$$

The solution to (4.5) is given by (see, e.g., [28, Chap. 46, eqs. (46.9)–(46.12), pp. 550–551])

$$H = \Sigma\left(\Sigma + \sigma^2 I\right)^{-1}. \quad (4.6)$$

This solution is obtained in two steps. First, rewrite the mean squared error using the trace

$$
\begin{aligned}
\mathbb{E}\left\|x - \left(\mu + H\left(y - \mu\right)\right)\right\|^2 &= \mathbb{E}\left\|x - \mu - H\left(x - \mu + \xi\right)\right\|^2 \\
&= \mathbb{E}\,\mathrm{Tr}\left[\left(I - H\right)\left(x - \mu\right)\left(x - \mu\right)^T\left(I - H\right)^T + H\xi\xi^T H^T\right] \quad (4.7) \\
&= \mathrm{Tr}\left[\left(I - H\right)\Sigma\left(I - H\right)^T + \sigma^2 H H^T\right],
\end{aligned}
$$

where we used the independence of noise and signal as well as their first and second order moments. Second, differentiate (4.7) with respect to H (using the fact that $\frac{d}{dH}\mathrm{Tr}\left[HB^T\right]$) to obtain

$$-2\left(I - H\right)\Sigma + 2\sigma^2 H = 0,$$

whose solution is given in (4.6). Plugging (4.6) into (4.4) provides the "optimal" linear filter

$$\widehat{x} = \mu + \Sigma\left(\Sigma + \sigma^2 I\right)^{-1}\left(y - \mu\right). \quad (4.8)$$

The estimator x ĥas a simple form when written in the basis of eigenvectors of Σ:

---

[2]Finding the probability density of x given noisy samples $y_1, \ldots, y_n$ is a deconvolution problem (assuming knowledge of the distribution of noise, e.g., Gaussian noise). This high-dimensional deconvolution problem is ill-conditioned, especially for Gaussian noise; see [16]. We remark that in cases where $n \gg p$ it may be possible to estimate and use higher (than second) order moments of x, but we do not pursue this possibility here.

$$\widehat{x} = \mu + \sum_{k=1}^{p} \frac{1}{1 + \frac{1}{\mathrm{SNR}_k}} \langle y - \mu, u_k \rangle u_k, \quad (4.9)$$

where u_k is the kth eigenvector of Σ, that is,

$$\Sigma = \sum_{k=1}^{p} \lambda_k u_k u_k^T, \quad (4.10)$$

and

$$\mathrm{SNR}_k = \frac{\lambda_k}{\sigma^2} \quad (4.11)$$

can be considered as the SNR of the kth component.

4.3. PCA in high dimensions

The filtering formula (4.9) requires knowledge of the first and second order moments μ, Σ, and σ^2. In practice, however, these are unknown and need to be estimated from a finite collection of noisy observations. Once estimated, the naive approach to filtering would be to replace μ, Σ, and σ^2 in (4.9) by their estimated counterparts. However, as we show below, the optimal filter turns out to be different from this naive procedure. But before constructing the optimal filter, it is important to quickly review a few recent results regarding PCA in high dimensions. The reader is referred to [22] for an extensive review of this topic.

Let $y_1, \ldots, y_n \in \mathbb{R}^p$ be n noisy observations from the additive noise model (4.1). The sample mean estimator $\widehat{\mu}_n$ is defined as

$$\widehat{\mu}_n = \frac{1}{n} \sum_{i=1}^{n} y_i,$$

and by the law of large numbers $\widehat{\mu}_n \to \mu$ (almost surely) as $n \to \infty$.

The sample covariance matrix S_n is defined as the following $p \times p$ matrix:

$$S_n = \frac{1}{n} \sum_{i=1}^{n} (y_i - \widehat{\mu}_n)(y_i - \widehat{\mu}_n)^T. \quad (4.12)$$

We denote by $l_1 \quad l_2 \quad \cdots \quad l_p$ the eigenvalues of the sample covariance matrix S_n and by $\widehat{u}_1, \widehat{u}_2, \ldots, \widehat{u}_p$ the corresponding computed eigenvectors, that is,

$$S_n = \sum_{k=1}^{p} l_k \widehat{u}_k \widehat{u}_k^T. \quad (4.13)$$

The sample covariance matrix converges to $\Sigma + \sigma^2 I_{p \times p}$ as $n \to \infty$ (while p is fixed). Assuming that Σ is rank deficient (i.e., its smallest eigenvalue is 0), the noise variance σ^2 and the covariance matrix Σ can be estimated from the sample covariance matrix in a straightforward manner. Our assumptions about the function f imply that the covariance matrix is indeed rank deficient. Specifically, recall that f is assumed to be compactly supported in a disk. As a result, in the 1-D projections, the pixel values at the boundaries correspond to noise without any signal contribution. It follows that σ^2 can be accurately and

simply estimated using the second order statistics of the values of the boundary pixels, provided that $n \gg 1$ regardless of the value of p. If, in addition, $n \gg p$, then we can also estimate Σ from the sample covariance matrix and the estimate for σ^2.

More thought is needed whenever the number of samples n is not exceedingly large (compared to p), and indeed much attention has been given in recent years to the analysis of PCA in the regime $p, n \to \infty$ with $p/n = \gamma$ fixed ($0 < \gamma < \infty$) [22]. This is also the interesting regime of parameters for the 2-D tomography problem at hand, since typical values for p and n range between several hundreds to several thousands. In such cases, the largest eigenvalue due to noise can be significantly larger than σ^2. It is therefore possible for the smaller eigenvalues of Σ to be "buried" inside the limiting Marcenko–Pastur (MP) distribution for the eigenvalues of the noise covariance matrix. As a result, the principal components that correspond to such small eigenvalues cannot be identified. The identifiability of the principal components from the eigenvalues of the sample covariance matrix was studied in [1, 33]. The key result is the presence of a phase transition phenomenon: In the joint limit $p, n \to \infty$, $p/n \to \gamma$, only components of Σ whose eigenvalues are larger than the critical value

$$\lambda_{crit} = \sigma^2 \sqrt{\gamma} \quad (4.14)$$

can be identified (almost surely) in the sense that their corresponding sample covariance eigenvalues "pop" outside the MP distribution. Formally (see, e.g., [30, Theorem 2.3]), if ξ is white noise and ξ and x have finite fourth moments, then in the joint limit $p, n \to \infty$, $p/n = \gamma$, the kth eigenvalue (k is fixed, e.g., $k = 2$) of the sample covariance matrix converges with probability one to

$$l_k \to \begin{cases} \sigma^2 \left(1 + \sqrt{\gamma}\right)^2 & \text{if} \quad \lambda_k < \sigma^2 \sqrt{\gamma}, \\ \left(\lambda_k + \sigma^2\right) \left(1 + \gamma \frac{\sigma^2}{\lambda_k}\right) & \text{if} \quad \lambda_k \geq \sigma^2 \sqrt{\gamma}. \end{cases} \quad (4.15)$$

Moreover, the dot product between the population eigenvector u_k and the eigenvector $\hat{u}_k$ computed by PCA also undergoes a phase transition almost surely:

$$|\langle u_k, \hat{u}_k \rangle|^2 \to c_k^2 = \begin{cases} 0 & \text{if} \quad \lambda_k < \sigma^2 \sqrt{\gamma}, \\ \dfrac{1 - \frac{\gamma \sigma^4}{\lambda_k^2}}{1 + \frac{\gamma \sigma^2}{\lambda_k}} & \text{if} \quad \lambda_k \geq \sigma^2 \sqrt{\gamma}. \end{cases} \quad (4.16)$$

This behavior is illustrated in Figure 7: As the level of noise σ increases, fewer components can be identified (here we varied σ and fixed p and n, but similarly, the theory can also be tested by varying any of the three parameters). Figure 8 shows principal components computed from clean projections and principal components that are computed from noisy projections. The resemblance between the two sets of components is evident, with the correlation between the "noisy" components and their "clean" counterparts being greatest for the first component and monotonically decreasing until it becomes completely random due to the phase transition, as predicted by the theory.

4.4. Linear Wiener filter using PCA in high dimensions

Motivated by the form of the optimal linear filter (4.9), we propose the following linear filter:

$$\widehat{x} = \widehat{\mu}_n + \sum_{k=1}^{p} h_k \langle y - \widehat{\mu}_n, \widehat{u}_k \rangle \widehat{u}_k, \quad (4.17)$$

where the scalar coefficients $h_1, \dots, h_p$, which we refer to as the filter coefficients, are to be determined from the minimum mean squared error criterion similar to (4.5). In other words, the filter coefficients are the solution to the minimization problem

$$\min_{h_1, \dots, h_p} \mathbb{E} \|x - \widehat{x}\|^2 = \min_{h_1, \dots, h_p} \mathbb{E} \left\| x - \widehat{\mu}_n - \sum_{k=1}^{p} h_k \langle y - \widehat{\mu}_n, \widehat{u}_k \rangle \widehat{u}_k \right\|^2. \quad (4.18)$$

Using (4.1) and the fact that $\widehat{u}_1, \widehat{u}_2, \dots, \widehat{u}_p$ form an orthonormal basis to $\mathbb{R}^p$, we have

$$
\begin{aligned}
\|x - \widehat{x}\|^2 &= \left\| x - \widehat{\mu}_n - \sum_{k=1}^{p} h_k \langle x - \widehat{\mu}_n + \xi, \widehat{u}_k \rangle \widehat{u}_k \right\|^2 \\
&= \left\| \sum_{k=1}^{p} (1 - h_k) \langle x - \widehat{\mu}_n, \widehat{u}_k \rangle \widehat{u}_k - \sum_{k=1}^{p} h_k \langle \xi, \widehat{u}_k \rangle \widehat{u}_k \right\|^2 \quad (4.19) \\
&= \sum_{k=1}^{p} [(1 - h_k) \langle x - \widehat{\mu}_n, \widehat{u}_k \rangle - h_k \langle \xi, \widehat{u}_k \rangle]^2.
\end{aligned}
$$

Since ξ is white noise with variance σ^2 independent of x, and $\mathbb{E}(x - \mu)(x - \mu)^T = \Sigma$, we get

$$\mathbb{E}[(1 - h_k) \langle x - \widehat{\mu}_n, \widehat{u}_k \rangle - h_k \langle \xi, \widehat{u}_k \rangle]^2 = (1 - h_k)^2 \left[\widehat{u}_k^T \Sigma \widehat{u}_k + |\langle \mu - \widehat{\mu}_n, \widehat{u}_k \rangle|^2 \right] + h_k^2 \sigma^2. \quad (4.20)$$

It follows that

$$\mathbb{E} \|x - \widehat{x}\|^2 = \sum_{k=1}^{p} (1 - h_k)^2 \left[\widehat{u}_k^T \Sigma \widehat{u}_k + |\langle \mu - \widehat{\mu}_n, \widehat{u}_k \rangle|^2 \right] + h_k^2 \sigma^2. \quad (4.21)$$

The optimal filter is therefore given by

$$h_k = \frac{1}{1 + \frac{\sigma^2}{\widehat{u}_k^T \Sigma \widehat{u}_k + |\langle \mu - \widehat{\mu}_n, \widehat{u}_k \rangle|^2}} \quad \text{for} \quad k = 1, \dots, p. \quad (4.22)$$

In practice, however, this filter cannot be used, since σ^2, Σ, and μ are unknown. Instead, we determine the filter coefficients in the limit $n, p \to \infty$ and with $p/n = \gamma$ fixed, and refer to the resulting filter as the asymptotically optimal filter. First, $\widehat{\mu}_n$ converges (almost surely) to μ in the limit, so the term $|\langle \mu - \widehat{\mu}_n, \widehat{u}_k \rangle|^2$ converges to zero for all k. Second, from (4.10) it immediately follows that

$$\widehat{u}_k^T \Sigma \widehat{u}_k = \sum_{l=1}^{p} \lambda_l |\langle \widehat{u}_k, u_l \rangle|^2 = \lambda_k |\langle \widehat{u}_k, u_k \rangle|^2 + \sum_{l=1, l \neq k} \lambda_l |\langle \widehat{u}_k, u_1 \rangle|^2. \quad (4.23)$$

From (4.16) it follows that the first term on the right-hand side of (4.23) converges almost surely to $\lambda_k c_k^2$. We proceed to show that the second term (involving the summation) tends to 0 under appropriate rapid decay assumptions about the population eigenvalues. For example, under the "spike" model [22], where Σ is assumed to have only a finite number K of nonzero eigenvalues (i.e., $\lambda_1 \quad \lambda_2 \quad \cdots \lambda_K > 0$ and $\lambda_{K+1} = \cdots = \lambda_p = 0$), we have that

$$\sum_{l=1,l\neq k}^{p} \lambda_l |\langle \widehat{u}_k, u_l \rangle|^2 \sum_{l=1,l\neq k}^{p} \lambda_l |\rangle \widehat{u}_k, u_l\rangle|^2.$$

Let $c_k \tilde{u}_k$ be the orthogonal projection of $\hat{u}_k$ onto u_k, that is, $\hat{u}_k = c_k\tilde{u}_k + r_k$, where r_k is perpendicular to u_k, and $\|r_k\|^2 = 1 - \tilde{c}_k^2$ (notice that $\tilde{c}_k^2 \to c_k^2$ almost surely as $p \to \infty$). Since u_k is perpendicular to u_l for $l \neq k$, we have that $|\langle \hat{u}_k, u_l \rangle|^2 = |\langle r_k, u_l \rangle|^2$. The vector r_k is a random vector uniformly distributed over the $(p-1)$-dimensional sphere of radius $\sqrt{1-\tilde{c}_k^2}$ [6, 29]. As a result, the expected value of the squared correlation is $\mathbb{E}|\langle r_k, u_l \rangle|^2 = \frac{1-\tilde{c}_k^2}{p-1}$ (recall that the expected value for the squared correlation of any unit vector with a random unit vector in $\mathbb{R}^{p-1}$ is $\frac{1}{p-1}$). Moreover, with a probability that goes to 1 as $p \to \infty$, the squared correlation is bounded by $C\frac{\log p}{p}$ for some $C > 0$. Altogether, we get that for $l \neq k$, $|\langle \hat{u}_k, u_l \rangle|^2 = |\langle r_k, u_l \rangle|^2 \to 0$ (almost surely) as $p \to \infty$. Therefore, in the spike model we obtain that

$$\widehat{u}_k^T \Sigma \widehat{u}_k \to \lambda_k c_k^2 \quad (4.24)$$

almost surely for all k. The spike model may be considered to be too restrictive for the tomography problem at hand. We would like to relax the assumption about the finite number of nonzero eigenvalues. A more realistic assumption would be that the eigenvalues decay sufficiently quickly. For example, suppose that there exists $a < 1$ such that

$$\text{Tr}\left(\Sigma^2\right) = \sum_{l=1}^{p} \lambda_l^2 = O\left(p^\alpha\right). \quad (4.25)$$

The assumption (4.25) does not hold for all images. For example, it does not hold for a 2-D image consisting of white noise windowed to a disk, as in this case $\text{Tr}(\Sigma^2) = O(p)$. However, (4.25) is expected to hold for a large class of images that arise in CT. We conjecture, without providing a formal proof, that assumption (4.25) implies (4.24). We are motivated to make this conjecture by the following heuristic arguments. The Cauchy–Schwarz inequality and (4.25) imply

$$\sum_{l=1,l\neq k}^{p} \lambda_l |\langle \widehat{u}_k, u_l \rangle|^2 \leq \sqrt{\sum_{l=1,l\neq k}^{p} \lambda_l^2} \sqrt{\sum_{l=1,l\neq k}^{p} |\langle \widehat{u}_k, u_l \rangle|^4} \leq C\sqrt{p^\alpha} \sqrt{\sum_{l=1,l\neq k}^{p} |\langle \widehat{u}_k, u_l \rangle|^4},$$

where $C > 0$ is some constant. We foresee a large deviation bound for the term

$$\sqrt{\sum_{l=1,l\neq k}^{p} |\langle \widehat{u}_k, u_l \rangle|^4}$$

since each of its $p-1$ individual summands is expected to concentrate at $O(1/p^2)$. A Bernsteinlike inequality would then imply that this term is $O\left(\sqrt{\frac{\log p}{p}}\right)$ with high probability that goes to 1 as $p \to \infty$. Since $\sqrt{p^\alpha \frac{\log p}{p}} \to 0$ for $a < 1$, we expect the conjecture to hold true.

Under the spike model or, alternatively, assumption (4.25) (provided that the above conjecture holds true), the optimal filter coefficients (4.22) converge almost surely to

$$h_k = \frac{1}{1+\frac{\sigma^2}{\lambda_k c_k^2}} = \frac{1}{1+\frac{1}{c_k^2 \mathrm{SNR}_k}}, \quad (4.26)$$

where SNR_k is given in (4.11). Using (4.16), we rewrite the filter coefficients as

$$h_k = \frac{1}{1+\frac{1}{\mathrm{SNR}_{\gamma,k}}}, \quad (4.27)$$

where

$$\mathrm{SNR}_{\gamma,k} = c_k^2 \mathrm{SNR}_k = \begin{cases} 0 & \text{if } \mathrm{SNR}_k < \sqrt{\gamma}, \\ \frac{\mathrm{SNR}_k^2 - \gamma}{\mathrm{SNR}_k + \gamma} & \text{if } \mathrm{SNR}_k \geq \sqrt{\gamma}. \end{cases} \quad (4.28)$$

We conclude that in the limit $n, p \to \infty$ and with $p/n = \gamma$ fixed, the asymptotically optimal linear Wiener filter is given by

$$\widehat{x} = \mu + \sum_{k=1}^{\infty} \frac{1}{1+\frac{1}{\mathrm{SNR}_{\gamma,k}}} \langle y - \mu, \widehat{u}_k \rangle \widehat{u}_k. \quad (4.29)$$

Comparing (4.29) and (4.9), we find that filtering using the computed eigenvectors is more aggressive compared to filtering using the population eigenvectors, in the sense that the decay of the filter coefficients is faster due to the extra c_k^2 terms. We explain the excess aggressiveness of the filter (4.29) in its need to compensate for the fact that the computed principal components are noisier for smaller population eigenvalues.

Although we derived the optimal filter (4.29) in the limit $n, p \to \infty$, we suggest using it in the practical case of a finite number of sample points. That is, the filter we propose using is

$$\widehat{x} = \widehat{\mu}_n + \sum_{k=1}^{\widehat{K}} \frac{1}{1+\frac{1}{\mathrm{SNR}_{\gamma,k}}} \langle y - \widehat{\mu}_n, \widehat{u}_k \rangle \widehat{u}_k, \quad (4.30)$$

where $\widehat{K}$ is the number of components estimated to satisfy $\lambda_k > \sigma^2 \sqrt{\gamma}$.

In order to use the filter (4.30) in practice, we need to know the noise variance σ^2 and the eigenvalues $\lambda_1, \ldots, \lambda_p$. These are, however, unknown and need to be estimated from the data. We mentioned earlier that σ^2 can be estimated using the boundary pixels that correspond merely to noise. Here we adapt the estimation method suggested recently by Kritchman and Nadler [25, 26] and conveniently use their MATLAB code.[3] The exact details of their procedures are beyond the scope of this paper, but the interested reader is urged to check their papers for details, analysis, and the history of the problem. Their method provides an estimator $\widehat{\sigma}^2$ for the noise variance and an estimator $\widehat{K}$ for the number of components satisfying (4.14) under the (more restrictive) spike model assumption. The top $\widehat{K}$ population eigenvalues $\lambda_1, \ldots, \lambda_{\widehat{K}}$ are then estimated as the positive solutions to the decoupled quadratic equations

---

[3]The MATLAB code is freely available from Nadler's website at http://www.wisdom.weizmann.ac.il/~nadler/Rank_Estimation/rank_estimation.html.

$$l_k = \left(\widehat{\lambda}_k + \widehat{\sigma}^2\right)\left(1 + \gamma\frac{\widehat{\sigma}^2}{\widehat{\lambda}_k}\right), \quad k=1,\ldots,\widehat{K},$$

as implied by (4.15). Figure 7 illustrates the estimators for a particular example.

Figures 11 and 12 show several denoised projections using (4.30) for different viewing directions and different levels of noise. The root mean squared error (RMSE) using the combined PCA-Wiener filter method is significantly smaller than the RMSE of the wavelet spin-cycling method. We emphasize that the PCA-Wiener filtering method does not require any external parameters: It is completely adaptive to the data and is free of any tuning parameters.

We remark that while optimality of the filter is guaranteed in the limit $n, p \rightarrow \infty$, we cannot claim optimality in the finite sample case. The deviation of (4.26) from (4.22) due to finite sample effects is not explored in this paper and is left for future research. It is plausible that suitable finite sample corrections would lead to an improved filtering procedure in the practical finite sample case.

Finally we remark on the possibility of using "sparse PCA" [7, 23] for an improved filtering scheme. The clean projections (shown in Figure 11) and the clean principal components shown in Figure 8 are piecewise smooth functions and are therefore expected to have a sparse representation in a suitable wavelet basis. We emphasize that while this property seems to hold for the Shepp–Logan phantom, it is not expected to hold in general for all possible 2-D images. In cases where the components are piecewise smooth, one may benefit from applying sparse PCA techniques [7, 23] in order to produce more accurate, and, under certain conditions, even consistent, estimators of the principal components and their eigenvalues. Our empirical experience with sparse PCA for the 2-D tomography problem is positive, but we postpone the derivation of the optimal filter for sparse PCA to future investigation.

4.5. Parity of components and their minimal required number

We conclude with a discussion of two PCA related issues that are more specific to the 2-D tomography problem. The first issue concerns the parity of the principal components. The perceptive reader has probably noticed that all components shown in Figure 8 are either even or odd functions. This is not a mere coincidence: The principal components are either even or odd functions, regardless of the underlying image, whether it is the Shepp–Logan phantom or another image. To see this, note that the projection taken at direction $-\theta$ is related to the projection at direction θ through

$$R_{-\theta}f(s) = R_{\theta}f(-s). \quad (4.31)$$

This motivates us to artificially double the number of projections from n to $2n$ by including all mirrored projections that correspond to the antipodal directions. The resulting sample covariance matrix commutes with the reflection matrix; therefore its eigenvectors are either even or odd, as inspected. However, the reflected projections are clearly dependent on the original ones; in particular, the realizations of noise are no longer independent, a necessary assumption for the method of [25, 26]. Thus, we cannot simply employ [25, 26] using the parameters $2n$ for the number of samples and p for the dimension. Fortunately, there is a simple remedy to this problem. Instead of doubling the number of projections, we first project the n projections onto the two orthogonal linear subspaces of even and odd functions, each of which is of dimension $p/2$ (for simplicity, we assume p is even). The even and odd

projectors, restricted to the positive axis $s > 0$, are denoted P_E and P_O, respectively, and are given by

$$P_E f(s) = \frac{f(s) + f(-s)}{2} \quad (4.32)$$

and

$$P_O f(s) = \frac{f(s) - f(-s)}{2}. \quad (4.33)$$

We restrict the projected projections to the positive axis ($s > 0$), hence representing each projection with just $p/2$ pixels, form two sample covariance matrices of size $(p/2) \times (p/2)$, compute their eigenvectors and eigenvalues, and reflect the computed eigenvectors to the negative axis ($s < 0$) based on the parity (even or odd). This procedure results in exactly the same eigenvectors and eigenvalues of the $p \times p$ sample covariance formed from the n original projections and their n reflections. Also, the realization of noise in the two sample covariance matrices remains independent. Thus, we apply the method of [25, 26] for each of the two $(p/2) \times (p/2)$ sample covariance matrices separately, using parameters $p/2$ for the effective dimension and n for the number of samples.

The second issue concerns the minimum number of principal components required to solve the 2-D tomography problem. Clearly, it is impossible to determine the viewing directions of the projections by using just the top principal component; at least two principal components are necessary to preserve the topology of a closed curve. But is it also a sufficient number of components? Figure 9 shows the embedding of the closed curve $\mathscr{C}^p$ of projections of the Shepp–Logan phantom onto the subspace spanned by the top two principal components. Clearly, the projected curve has a nontrivial self-intersection, rendering the impossibility of unique viewing direction determination.[4] The self-intersection is a byproduct of the parity of the top two principal components. Note that in the case of the Shepp–Logan phantom, the first principal component u_1 is an even function, while the second component u_2 is an odd function. Consider the expansion of the projection $R_\theta f$ in terms of the mean projection μ and the principal components $u_1, u_2, \ldots$:

$$R_\theta f(s) = \mu(s) + a_1(\theta) u_1(s) + a_2(\theta) u_2(s) + \cdots.$$

Suppose θ^* is a viewing direction for which $a_2(\theta^*) = 0$, that is, $R_{\theta^*} f - \mu$ is perpendicular to u_2: Such a direction must exist due to the continuity of the Radon transform (Proposition 2.1) and due to the fact that

$$\int \langle R_\theta f - \mu, u_2 \rangle \, d\theta = \langle \int (R_\theta f - \mu) \, d\theta, u_2 \rangle = \langle 0, u_2 \rangle = 0.$$

The reflection property (4.31), the fact that the mean μ and the first component u_1 are even functions, and $a_2(\theta^*) = 0$ together imply that the projections onto the top two principal components of $R_{-\theta^*} f$ and $R_{\theta^*} f$ coincide:

$$R_{\theta^*} f(s) = \mu(s) + a_1(\theta^*) u_1(s) + 0 \cdot u_2(s) + \cdots,$$

---

[4]Since there is a single intersection, it is possible to "traverse" the curve in two different ways, giving rise to two different viewing direction orderings. The user may still be able to correctly choose between the two possibilities, either automatically by using multiway clustering methods [8], or manually by examining the two resulting reconstructions.

$$
\begin{aligned}
R_{-\theta^*} f(s) &= R_{\theta^*} f(-s) \\
&= \mu(-s) + a_1(\theta^*) u_1(-s) + a_2(\theta^*) u_2(-s) + \dots \\
&= \mu(s) + a_1(\theta^*) u_1(s) + 0 \cdot u_2(s) + \cdots.
\end{aligned}
$$

This explains the nontrivial self-intersection of the 2-D PCA map in the case of the Shepp–Logan phantom. While the above discussion focused on the Shepp–Logan example, it can be easily generalized to any image, and it allows us to conclude that, for a general image, at least two odd principal components are needed in order to avoid nontrivial self-intersections. This is a necessary condition, although it may not be sufficient. Returning to the Shepp–Logan example, we observe that the 3-D PCA mapping of $\mathscr{C}^p$ also exhibits a nontrivial self-intersection, since u_3 happens to be an even function. The fourth component u_4 turns out to be an odd function, and Figure 9 shows that the PCA mappings in dimension four (and therefore also in higher dimensions) successfully preserve the topology of $\mathscr{C}^p$. This is also demonstrated by the 2-D diffusion map embeddings shown in Figure 10. From this discussion we also conclude a theoretical limitation of any PCA-based method for solving the 2-D tomography problem in the case of a general underlying image. Indeed, recall that the identifiable components are those whose eigenvalues are greater than the critical value (4.14) (where the effective dimension is $p/2$ instead of p as discussed above). It follows that there exists a theoretical limitation for any PCA-based method such as ours: A necessary condition is that the number of samples n is sufficiently high as well as that the noise variance σ^2 is sufficiently low so that at least two odd components can be identified; i.e., their corresponding eigenvalues must be greater than $\sigma^2 \sqrt{\frac{p}{2n}}$.

5. Graph denoising

As mentioned above, the diffusion map method is limited in the presence of noise, as the latter may change the topology of the underlying manifold, even when using the combined PCA-Wiener filtering scheme. In our case, noise can "shortcut" the curve (see, e.g., Figure 2). It is therefore desirable to detect such shortcut edges in advance and remove them from the similarity matrix W.

After their introduction by Watts and Strogatz [37], small-world graphs were extensively used to describe many natural phenomena [21]. 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 q each ring edge is rewired to a random vertex, and with probability $1 - q$ 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 i and j with a "shortcut" edge $e = (i, j)$ between them is expected to be much smaller than the number of common neighbors of two nearby vertices [37]. 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

$$
J(i, j) = \frac{|N_i \cap N_j|}{|N_i \cup N_j|}, \quad (5.1)
$$

where N_i is the set of vertices connecting to vertex i. 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 [18] 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. While the underlying geometry of the graph should be that of a simple closed curve, such confusion due to noise is realized by shortcut edges that may change its topology. This change in topology can affect the long time behavior of the random walk on the graph. Indeed, it was observed in [37] 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 estimating the beaming directions using the diffusion map technique. We use the Jaccard coefficient 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. The threshold value is chosen by the number of edges we wish to keep.

The Jaccard index can be computed efficiently as follows. Suppose W is the adjacency matrix of the graph (that is, the entries of W are either 0 or 1). The graph may be either directed or undirected, where in the latter case the matrix W is symmetric. The number of common neighbors to i and j is

$$|N_i \cap N_j| = \sum_k w_{ik} w_{jk} = \left(WW^T\right)_{ij},$$

and the number of neighbors of i is

$$|N_i| = \sum_k w_{ik} = (W1)_i,$$

where $\mathbf{1} = (1, 1, \ldots, 1)^T$ is the all-ones vector. The inclusion-exclusion principle implies that

$$|N_i \cup N_j| = |N_i| + |N_j| - |N_i \cap N_j| = (W1)_i + (W1)_j - \left(WW^T\right)_{ij}.$$

The Jaccard index $J(i, j)$ can therefore be written as

$$J(i, j) = \frac{\left(WW^T\right)_{ij}}{(W1)_i + (W1)_j - (WW^T)_{ij}} \quad (5.2)$$

or, equivalently, in matrix notation

$$J = WW^T ./ \left[W11^T + 11^T W^T - WW^T \right], \quad (5.3)$$

where ./ denotes elementwise division (as in MATLAB). Note that J is a symmetric matrix; that is, $J = J^T$ even if W is nonsymmetric as in the directed graph case.

The benefit of using (5.1) together with diffusion maps can be summarized by the viewpoint of different time scales. On the one hand, the beaming directions are estimated from the top first and second nontrivial eigenvectors of the random walk matrix A. These eigenvectors

correspond to the long time behavior of the random walk over the data points, since they correspond to the largest (nontrivial) eigenvalues of A. On the other hand, the computation of (5.1) involves only the common neighbors, which is related to the diffusion process at a short time scale, corresponding to at most two steps of the random walk. While the purpose of the Jaccard index is to remove the "bad" edges, the purpose of the diffusion mapping using the top two eigenvectors is to reveal the global ordering of the beaming directions.

6. Algorithm

In this section we summarize the steps of our reconstruction algorithm. The input to the algorithm consists of n noisy projections $y_1, \ldots, y_n$, each of which is a vector in $\mathbb{R}^p$ corresponding to the discretization of the p equally spaced detectors. The algorithm depends on only two parameters, denoted α and β, that are explained in Steps 2 and 3 below. These parameters either can be prechosen by the user or the algorithm can automatically search for their optimal values.

Step 1: PCA and linear filtering. Project $y_1, \ldots, y_n$ onto the $\lceil \frac{p}{2} \rceil$−dimensional subspace of even functions and onto the $\lfloor \frac{p}{2} \rfloor$−dimensional subspace of odd functions; see (4.32)–(4.33). Perform PCA twice, once for each subspace, and extend the computed eigenvectors to vectors of length p based on the parity. Each PCA can be computed by either forming the sample covariance matrix or by using the singular value decomposition (SVD), which is the preferred method due to computational considerations.[5] The computed eigenvalues are fed into the method of [25] to estimate the number of components K, the noise variance[6] σ^2, and the signal-to-noise ratios $\text{SNR}_{\gamma,k}$. Denoise all projections using the filter (4.30), and denote the filtered projections $\widehat{x}_1, \widehat{x}_2, \ldots, \widehat{x}_n \in \mathbb{R}^p$. Compress each filtered projection $\widehat{x}_i \in \mathbb{R}^p$ to $\mathbb{R}^{\widehat{K}}$ using its first $\widehat{K}$ expansion coefficients $\langle x_i, \widehat{u}_k \rangle$ $(k = 1, \ldots, \widehat{K})$, and denote

$$\widehat{x}_i^{\widehat{K}} = \left(\langle \widehat{x}_i, \widehat{u}_1 \rangle, \ldots, \langle \widehat{x}_i, \widehat{u}_{\widehat{K}} \rangle \right), \quad i = 1, \ldots, n. \quad (6.1)$$

Step 2: Similarity matrix W. For each of the n compressed projections in $\mathbb{R}^{\widehat{K}}$ search for its N nearest neighbors with respect to the Euclidean distance $\| \cdot \|_{\widehat{K}}$. Assuming that changing the beaming direction by α degrees has a small effect on the Radon projection, choose $N = \lfloor n \frac{2\alpha}{360°} \rfloor$. If α is not prechosen by the user, then the algorithm is repeated with different values for α, and in Step 6 of the algorithm it automatically chooses the optimal reconstruction based on a criterion described in Step 6. From the results of the nearest neighbors search, construct a directed graph with n vertices corresponding to the projections, and put a directed edge from i to j iff projection j is one of the N nearest neighbors of projection i (that is, iff $j \in N_i$). Construct an $n \times n$ similarity matrix W^α whose entries W_{ij}^α are defined by

$$W_{ij}^\alpha = \begin{cases} 1 & \text{if } j \in N_i, \\ 0 & \text{if } j \notin N_i. \end{cases} \quad (6.2)$$

Note that W^α is not necessarily symmetric. That is, we may have a pair of nodes i and j satisfying $j \in N_i$, but $i \notin N_j$.

---

[5] For large values of n and p, even the SVD may be computationally infeasible. In such cases we recommend using the recently proposed randomized algorithms for computing the SVD; see [19] for a comprehensive review of such methods.
[6] Each of the two PCAs may give a slightly different estimate for the noise variance. We therefore estimate the noise variance by averaging the two estimators. Alternatively, the noise variance can be estimated using the boundary pixels.

Step 3: Graph denoising. Calculate the Jaccard index $J^\alpha(i, j)$ for all edges. The second parameter of our algorithm is the threshold value β. Remove all edges whose Jaccard index is less than β. Moreover, keep only edges for which both $i \in N_j$ and $j \in N_i$. We denote the thresholded matrix by $W^{\alpha,\beta}$, that is,

$$W_{ij}^{\alpha,\beta} = \begin{cases} 1 & \text{if} \quad W_{ij}^\alpha W_{ji}^\alpha > 0 \quad \text{and} \quad J^\alpha(i,j) \geq \beta, \\ 0 & \text{if} \quad W_{ij}^\alpha W_{ji}^\alpha = 0 \quad \text{or} \quad J^\alpha(i,j) < \beta. \end{cases} \quad (6.3)$$

Note that $W^{\alpha,\beta}$ is a symmetric matrix that can be viewed as the adjacency matrix of an undirected graph. If β is not prechosen by the user, then the algorithm tries several different values of β and executes the following steps for each of them separately. That is, the remaining steps of the algorithm are performed on several $W^{\alpha,\beta}$s with different values for β, until Step 6, where we automatically detect the optimal threshold value β, based on the criterion described in Step 6. We remark that for large values of β some nodes may become isolated, that is, all their edges have been removed. We remove such nodes from the graph and, as a result, estimate only the beaming directions of the projections that correspond to the remaining nodes.

Step 4: Diffusion map embedding. Form an $n \times n$ diagonal matrix $D^{\alpha,\beta}$ with $D_{ii}^{\alpha,\beta} = \sum_{j=1}^n W_{ij}^{\alpha,\beta}$, the normalized weighted matrix $W^{\tilde{\alpha},\beta} := (D^{\alpha,\beta})^{-1} W^{\alpha,\beta}(D^{\alpha,\beta})^{-1}$, and an $n \times n$ diagonal matrix $D^{\tilde{\alpha},\beta}$ so that $\tilde{D}_{ii}^{\alpha,\beta} = \sum_{j=1}^n \tilde{W}_{ij}^{\alpha,\beta}$. Then we compute the top two nontrivial eigenvectors φ_1 and φ_2 of $A^{\alpha,\beta}$ whose entries are given by $A_{ij}^{\alpha,\beta} = \frac{\tilde{W}_{ij}^{\alpha,\beta}}{\tilde{D}_{ii}^{\alpha,\beta}}$. The embedding $y_i \mapsto (\phi_1(i), \phi_2(i))$ reveals the ordering of the beaming directions. We estimate the beaming directions as equally spaced points on S^1 according to their ordering.

Step 5: 2-D reconstruction. Invert the Radon transform to reconstruct the 2-D image $f^{\alpha,\beta}$ from the noisy projections and their estimated beaming directions.

Step 6: Automatic estimation of parameters (optional). At this stage we get several different 2-D image reconstructions $f^{\alpha,\beta}$ corresponding to the different choices of the parameters α and β used in Steps 2 and 3. Out of all available reconstructions we need to automatically choose the best one using some optimality criterion. One possible criterion is to use the ℓ_1 norm of the expansion of the reconstructed image in a wavelet basis. Specifically, for each reconstruction, we compute the 2-D wavelet decomposition of the ℓ_2−normalized image $f^{\alpha,\beta}/\|f^{\alpha,\beta}\|_2$. Then, we choose the reconstruction whose wavelet coefficient vector has the smallest ℓ_1 norm. For the wavelet transform we use the Daubechies db2 mother wavelet with 4-level decomposition. The motivation for this step is that the underlying clean image is expected to be sparse in the wavelet domain. While this ℓ_1 heuristic seems to perform well for the Shepp–Logan phantom (see section 7), we do not find it to perform well for many other natural images, including the image that is used in subsection 7.1. For such images other quality measures are expected to perform better, depending on the application domain. We do not explore other quality measures in this paper, and we do not claim that the ℓ_1 heuristic is optimal in any sense. We emphasize that Step 6 is optional, and in practice the user may want to use parameters α and β that are predetermined using some sort of manual training.

7. Numerical results

We performed several numerical simulations in order to test the performance of our algorithm. In the first set of experiments the underlying 2-D object was the Shepp–Logan

phantom, while in the other set of experiments the underlying 2-D image was a more realistic abdominal CT. For the simulations with the Shepp–Logan phantom, the number of projections was $n = 1024$, and the number of discretization points was $p = 512$. 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

$$\text{SNR} \quad [\text{dB}] = 10 \quad \log_{10} \left(\frac{\text{Var} \quad S}{\sigma^2} \right), \quad (7.1)$$

where S is the array of the noiseless projections. As a reference to later reconstructions, Figure 3(a) shows the original Shepp–Logan phantom, while Figures 3(b)–3(e) show reconstructions of the Shepp–Logan phantom from noisy projections with known beaming directions at different levels of noise.

The results of applying the algorithm described in section 6 to noisy projections with unknown beaming directions are illustrated in Figure 4. These results are obtained by fixing the parameter values to $\alpha = 6$ and $\beta = 0.5$. That is, the final optional step of the algorithm was not applied. Obviously, for same level of noise, the reconstructions in Figure 3 have better quality compared to the reconstructions in Figure 4, which are missing the extra knowledge of the beaming directions. Still, our algorithm succeeds in providing similar reconstructions (up to the unavoidable degrees of freedom of rotation and reflection) even when the beaming directions are unknown for SNR = –3dB and above. The main features of the original Shepp–Logan phantom are visible in our reconstructions even at such low SNRs. Figures 4(i)–4(p) demonstrate that the beaming directions are estimated successfully and mostly follow their true ordering for SNR = –3dB and above.

Figure 4 indicates that the algorithm fails to produce a reasonable estimation of the angles for SNR = –4dB and below when using the fixed parameter values $\alpha = 6$ and $\beta = 0.5$. We therefore turned to explore the behavior of the algorithm when the final step is also applied. We searched the parameter space by letting $\alpha \in \{3, 4, 5, 6, 7, 8, 9, 10\}$ and $\beta \in \{0.35, 0.36, \dots, 0.74, 0.75\}$. The reconstructions and the beaming angle estimation are shown in Figure 5. The result for SNR = –4dB is much more satisfactory (compared to Figure 4). Even for SNR = –5dB the ordering of the estimated beaming directions is not completely random, and some features of the phantom can still be observed, although the image is quite fuzzy. The optimal values for α and β that were found in Step 6 of the algorithm are summarized in Table 1.

Figure 6 shows reconstructions obtained by applying the method described in [12] to the same sets of noisy projections. That method uses wavelet spin-cycling denoising instead of PCA and diffusion maps without the graph denoising step, and provides successful reconstructions only for SNR = 2dB and above.

In the following we describe the numerical results that are specific to the different steps of the algorithm. We start with Step 1 for PCA and Wiener filtering in order to denoise the projections. Bar plots of the 20 largest eigenvalues of the sample covariance matrix (including both odd and even functions) corresponding to different levels of noise are shown in Figure 7. The caption of Figure 7 details the number of identifiable components as predicted by [25]. Figure 8 shows the principal components obtained from clean projections as well as the principal components obtained from noisy projections at SNR = –5dB.

Figure 9 shows the 2-D embedding of clean projections obtained by linearly projecting them onto the subspace spanned by the top two principal components. The embedded curve exhibits a nontrivial self-intersection rendering the impossibility of unique determination of the beaming directions when only two components are used. This phenomenon is explained

in section 4, where we emphasize that at least two odd principal components are necessary to avoid nontrivial self-intersections. Examination of the parity of the principal components shown in Figure 8 reveals that at least the top four components are needed in order to avoid nontrivial self-intersections. This finding is confirmed by the 2-D diffusion mappings of clean projections after linearly projecting them to subspaces spanned by the top K principal components ($K = 2, 3, 4, 5$), as shown in Figure 10.

A comparison between denoising the noisy projections using our combined PCA-Wiener filtering approach and denoising using wavelets is illustrated in Figures 11, 12, and 13. The wavelet denoising procedure consists of using the full spin-cycle algorithm [10] with hard thresholding of the Daubechies db2 wavelet coefficients of each projection image, as described in [12]. The comparison shows that both denoising methods do relatively well for SNR = 2dB (Figure 11), but the combination of PCA with the Wiener filter is clearly better for SNR = –1dB (Figure 12) and SNR = –5dB (Figure 13). We attribute the success of the combined PCA-Wiener filter approach at relatively low SNRs to the adaptivity of the principal components and to the optimality criterion of the Wiener filter.

The effect of graph denoising in Step 3 of the algorithm is demonstrated in Figure 14 corresponding to SNR = –4dB. The vertices are arranged on a circle according to the beaming directions of the projections they represent, while edges are represented by chords. The left panel shows the edges of the nearest neighbors graph that is formed in Step 2 with α = 6, while the right panel shows the edges after Step 3 with a thresholding level β = 0.5. A large portion of the "shortcut" edges were successfully removed. We attribute the seemingly nonrandom behavior of the shortcut edges that are left in the denoised graph shown in Figure 14(b) to the particular shape of the Shepp–Logan phantom, which gives rise to somewhat similar projections that are taken at particular different beaming directions.

Finally, we conducted a large scale experiment with 100 different realizations of noise for each value of the SNR. For that experiment we also incorporated the final step of the algorithm and searched for the optimal parameter values, with $\alpha \in \{3, 4, 5, 6, 7, 8, 9, 10\}$ and $\beta \in \{0.35, 0.36, \ldots, 0.74, 0.75\}$. The results are summarized in Table 2. Note that the standard deviation of the ℓ_1 norm is considerably smaller for SNR = –5dB compared to other SNR values. We explain this by the constant failure of our algorithm to produce satisfactory reconstructions at such a low SNR (the poor reconstruction is indicated by the large ℓ_1 norm associated with this SNR).

7.1. Numerical experiment for abdominal CT

In this section we demonstrate the applicability of our algorithm to a real abdominal CT image (the image is a CT of the second author's father and is used with his permission). The CT image (Figure 15) was obtained by a Toshiba Aquilion 64 CFX CT scanner and is of size 380×380 pixels. We randomly picked $n = 1024$ angles from $[0, \pi]$ and generated the projections related to these angles. The number of discretization points of each projection is $p = 541$. The clean projections were contaminated by Gaussian white noise at different noise levels SNR [dB] = 30, 10, 8, 5, 4, 3. The first 50 eigenvalues of the covariance matrix of the clean projections are shown in Figure 15; the first eight eigenvectors of the covariance matrix of the clean projections and the covariance matrix of the noisy projections (SNR = 8dB) are shown in Figure 16. There are only a few dominant principal components although their number is larger compared to the case of the Shepp–Logan phantom. We found that the ℓ_1 criterion proposed in Step 6 did not perform well in this case. Instead, we fix the parameters $\alpha = 5$ and $\beta = 0.6$ in all experiments. While these parameters were found to be optimal for the Shepp–Logan phantom, they are not necessarily optimal for the real CT image. Still, by eyeballing the reconstructions that our algorithm produced, we confirmed

that these parameters gave satisfactory results. At the moment we do not have a better automatic way of choosing the parameters for images of this kind.

The PCA-based denoising results are demonstrated in Figure 17 when the noise level is 8dB. Figure 18 shows the reconstruction results obtained by applying the entire algorithm (Steps 1–5, excluding Step 6). When the noise level is 4dB, the estimation of the projection angles is accurate, and the large structures are distinguishable; for example, the spinal cord and the liver are visible, although the other parts are blurred. The algorithm fails when the noise level is 3dB or below, as shown in Figure 19.

8. 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 a combination of PCA and classical Wiener filtering, while the second denoising step consists of denoising the graph of similarities between denoised projections using the Jaccard index from network analysis. The additional denoising steps significantly improve the noise tolerance of the reconstruction method for the Shepp–Logan phantom from a benchmark of SNR = 2dB reported in [12] using diffusion maps and wavelet denoising, to SNR = –3dB obtained here.

We expect the combination of PCA, Wiener filtering, graph denoising, and diffusion maps to be useful in many other applications that require the organization of high-dimensional data with an underlying nonlinear low-dimensional structure. While the diffusion map framework is well adjusted to studying and analyzing 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 PCA 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. We combined recent results for PCA in high dimensions with the classical Wiener filtering approach in order to derive an asymptotically optimal filter. We believe that this result may be valuable in many other applications. Our asymptotically optimal filter requires the estimation of the noise variance, the number of identifiable principal components, and the population eigenvalues. These are estimated in practice using the method of [25]. Our numerical experiments show that denoising by PCA outperforms denoising by a prechosen basis such as a wavelet basis, and we attribute this success to the data adaptivity of the PCA basis.

The second denoising step in our procedure consists of denoising the graph using the Jaccard index. The objective in this denoising step is to restore the correct topology of the data by removing "bad" edges that shortcut the underlying manifold.

Throughout this paper we assumed that the projections are centered, but in practice the projections may be shifted with unknown, though usually small, shifts. The problem of finding both the beaming directions and the shifts was considered in [2, 3]. We remark that we can still use the diffusion map framework by associating similarity weights w_{ij} with the translational-invariant distances d_{ij} that are given by

$$d_{ij} = \min_{\tau \in} \lVert R_i - T_\tau R_j \rVert_{\ell_2},$$

where T_τ is the translation operator over $\mathbb{R}$ satisfying $T_\tau f(x) = f(x + \tau)$, and R_i denotes the noisy version of $R_{\theta_i} f$. These distances factor out the single degree of freedom associated with translations so that diffusion maps should recover the correct parameterizations of the closed curve as before. Moreover, if the shift of projection R_i is $\tau_i \in \mathbb{R}$, then the relative shift of projections R_i and R_j is $\tau_i - \tau_j$, which can be estimated from computing

$$\tau_{ij} = \underset{\tau \in}{\operatorname{argmin}} \| R_i - T_\tau R_j \|_{\ell_2}.$$

After the diffusion map step, but prior to reconstruction, the shifts τ_is can be estimated by solving an overdetermined least squares problem of the form $\tau_i - \tau_j = \tau_{ij}$ for all (i, j) with similar beaming directions θ_i and θ_j (as estimated by the diffusion map).

Finally, we note that better reconstructions can be achieved by regularizing the inverse Radon transform (Step 5). The reconstruction problem can be viewed as finding the original function f from its projections $R_\theta f$. So once the beaming directions are estimated, this can be cast as an overdetermined linear system of the form $Ax = b$, where the operator A is the Radon projection operator, x relates to the function f, and b are the Radon samples. One way to reconstruct would be to minimize $\| Ax - b \|_2^2$ (least squares method), and this can be done efficiently using conjugate gradient iterations, where the matrices A (projection) and A^T (backprojection) can be applied quickly using the slice-theorem and the nonuniform FFT algorithm of Dutt and Rokhlin [14]. It is also possible to add a regularization term of the form $\mu \| \mathcal{W} x \|_1$, where $\mathcal{W}$ is the wavelet transform and $\mu > 0$ is the regularization parameter. This will encourage the reconstruction to have a sparse representation in the wavelet basis.

Acknowledgments

Appendix A

Proof of Proposition 2.1

Suppose $h \in L^2\left(\mathbb{R}^2\right)$ is a function like f that vanishes outside the unit disk. Clearly, $R_\theta h$ vanishes outside the interval $[-1, 1]$ since h vanishes outside the unit disk. The fact that $h \in L^2\left(\mathbb{R}^2\right)$ implies that $R_\theta h$ is in $L^2([-1, 1])$, since Cauchy–Schwarz inequality gives the following estimate:

$$\int_{-1}^1 |R_\theta h(s)|^2 ds = \int_{-1}^1 \left| \int_{-1}^1 h\left(s\theta + r\theta^\perp\right) dr \right|^2 ds \leq 2 \int_{-1}^1 \int_{-1}^1 |h\left(s\theta + r\theta^\perp\right)|^2 dr\, ds = 2\|h\|_{L^2(\ ^2)}^2.$$

That is,

$$\|R_\theta h\|_{L^2(\ )} \le \sqrt{2}\|h\|_{L^2(\ ^2)}. \quad \text{(A.1)}$$

In fact, it can be similarly shown that

$$\int_{-1}^{1} \frac{1}{\sqrt{1-s^2}} |R_\theta h(s)|^2 ds \le 2\|h\|_{L^2(\ ^2)}^2,$$

that is, $R_\theta h \in L^2\left([-1,1], \frac{1}{\sqrt{1-s^2}}\right)$ (see, e.g., [15, Proposition 6.6.2, p. 214]), but we do not need this finer estimate here.

Continuous functions that vanish outside the unit disk are dense in $L^2(B_1)$, where B_1 denotes the unit disk. Therefore, for any $\varepsilon > 0$, we can find a continuous function g that vanishes outside B_1 such that $\|f - g\|_{L^2(\ ^2)} < \varepsilon$. From the linearity of the Radon transform and the estimate (A.1) we get

$$\|R_\theta f - R_\theta g\|_{L^2(\ )} = \|R_\theta(f-g)\|_{L^2(\ )} \le \sqrt{2}\|f-g\|_{L^2(\ ^2)} < \sqrt{2}\varepsilon. \quad \text{(A.2)}$$

Since g is a continuous function and the unit disk is compact, g is uniformly continuous and there exists $\delta > 0$ such that $|g(x) - g(y)| < \varepsilon$ for all $x, y \in B_1$ satisfying $\|x - y\| < \delta$.

Suppose that $\theta_2, \theta_1 \in S^1$ satisfy $\|\theta_2 - \theta_1\| < \delta$. Then, for $s^2 + r^2 \quad 1$, we have

$$\left\|\left(s\theta_2 + r\theta_2^\perp\right) - \left(s\theta_1 + r\theta_1^\perp\right)\right\| = \|s(\theta_2 - \theta_1) + r(\theta_2^\perp - \theta_1^\perp)\| = \sqrt{s^2+r^2}\|\theta_2 - \theta_1\| \le \|\theta_2 - \theta_1\| < \delta.$$

Therefore,

$$|g(s\theta_2 + r\theta_2^\perp) - g(s\theta_1 + r\theta_1^\perp)| < \varepsilon.$$

The triangle inequality thus gives

$$|R_{\theta_2}g(s) - R_{\theta_1}g(s)| = \left|\int_{-1}^{1}\left(g(s\theta_2 + r\theta_2^\perp) - g(s\theta_1 + r\theta_1^\perp)\right)dr\right| \le 2\varepsilon$$

for all $s \in [-1, 1]$. Hence,

$$\|R_{\theta_2}g - R_{\theta_1}g\|_{L^2(\ )} = \left(\int_{-1}^{1}|R_{\theta_2}g(s) - R_{\theta_1}(s)|^2 ds\right)^{1/2} \le \sqrt{8}\varepsilon. \quad \text{(A.3)}$$

From (A.2), (A.3), and the triangle inequality, we get that

$$\|R_{\theta_2}f - R_{\theta_1}f\|_{L^2(\ )} \le \|R_{\theta_2}f - R_{\theta_2}g\|_{L^2(\ )} + \|R_{\theta_2}g - R_{\theta_1}g\|_{L^2(\ )} + \|R_{\theta_1}g - R_{\theta_1}f\|_{L^2(\ )} < \sqrt{2}\varepsilon + \sqrt{8}\varepsilon + \sqrt{2}\varepsilon = \sqrt{32}\varepsilon$$

whenever $\|\theta_2 - \theta_1\| < \delta$. This shows that

$$\lim_{\theta_2 \to \theta_1} \| R_{\theta_2} f - R_{\theta_1} f \|_{L^2(\ )} = 0. \quad \text{(A.4)}$$

REFERENCES

1. Baik J, Silverstein JW. Eigenvalues of large sample covariance matrices of spiked population models. J. Multivariate Anal. 2006; 97:1382–1408.

2. Basu S, Bresler Y. Feasibility of tomography with unknown view angles. IEEE Trans. Image Process. 2000; 9:1107–1122. [PubMed: 18255480]

3. Basu S, Bresler Y. Uniqueness of tomography with unknown view angles. IEEE Trans. Image Process. 2000; 9:1094–1106. [PubMed: 18255479]

4. Belkin, M.; Niyogi, P. Proceedings of the 18th Conference on Learning Theory (COLT), Lecture Notes in Comput. Sci. 3559. Springer-Verlag; Berlin: 2005. Towards a theoretical foundation for Laplacian-based manifold methods; p. 486-500.

5. Belkin, M.; Niyogi, P. Convergence of Laplacian eigenmaps, in Advances in Neural Information Processing Systems 19. The MIT Press; Cambridge, MA: 2007. p. 129-136.

6. Benaych-Georges F, Nadakuditi RR. The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. Adv. Math. 2011; 227:494–521.

7. Bickel PJ, Levina E. Covariance regularization by thresholding. Ann. Statist. 2008; 36:2577–2604.

8. Chen G, Lerman G. Spectral curvature clustering (SCC). Int. J. Comput. Vision. 2009; 81:317–330.

9. Coddington, EA.; Levinson, N. Theory of Ordinary Differential Equations. Krieger; Malabar, FL: 1984.

10. Coifman, RR.; Donoho, D. Wavelets and Statistics, Lecture Notes in Statist. Vol. 103. Springer-Verlag; New York: 1995. Translation-invariant de-noising; p. 125-150.

11. Coifman RR, Lafon S. Diffusion maps. Appl. Comput. Harmon. Anal. 2006; 21:5–30.

12. Coifman RR, Shkolnisky Y, Sigworth FJ, Singer A. Graph Laplacian tomography from unknown random projections. IEEE Trans. Image Process. 2008; 17:1891–1899. [PubMed: 18784036]

13. Deans, SR. The Radon Transform and Some of Its Applications. Krieger; Malabar, FL: 1993.

14. Dutt A, Rokhlin V. Fast Fourier transforms for nonequispaced data. SIAM J. Sci. Comput. 1993; 14:1368–1393.

15. Epstein, CL. Introduction to the Mathematics of Medical Imaging. Pearson Education; Upper Saddle River, NJ: 2003.

16. Fan J. On the optimal rates of convergence for nonparametric deconvolution problems. Ann. Statist. 1991; 19:1257–1272.

17. Frank, J. Three-Dimensional Electron Microscopy of Macromolecular Assemblies: Visualization of Biological Molecules in Their Native State. 2nd ed.. Oxford University Press; New York: 2006.

18. Goldberg DS, Roth FP. Assessing experimentally derived interactions in a small world. Proc. Natl. Acad. Sci. USA. 2003; 100:4372–4376. [PubMed: 12676999]

19. Halko N, Martinsson PG, Tropp JA. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. SIAM Rev. 2011; 53:217–288.

20. Hein, M.; Audibert, J.; von Luxburg, U. Proceedings of the 18th Conference on Learning Theory (COLT), Lecture Notes in Comput. Sci. 3559. Springer-Verlag; Berlin, Heidelberg: 2005. From graphs to manifolds—weak and strong pointwise consistency of graph Laplacians; p. 470-485.

21. Humphries MD, Gurney K. Network 'small-world-ness': A quantitative method for determining canonical network equivalence. PLoS ONE. 2008; 3:e0002051. [PubMed: 18446219]

22. Johnstone, IM. Proceedings of the International Congress of Mathematicians, Madrid, Spain, 2006. European Mathematical Society Publishing House; Zurich: 2007. High dimensional statistical inference and random matrices; p. 307-333.

23. Johnstone IM, Lu AY. On consistency and sparsity for principal components analysis in high dimensions. J. Amer. Statist. Assoc. 2009; 104:682–693.

24. Kak, AC.; Slaney, M. Principles of Computerized Tomographic Imaging. SIAM; Philadelphia: 2001.

25. Kritchman S, Nadler B. Determining the number of components in a factor model from limited noisy data. Chem. Intell. Lab. Syst. 2008; 94:19–32.

26. Kritchman S, Nadler B. Non-parametric detection of the number of signals, hypothesis testing and random matrix theory. IEEE Trans. Signal Process. 2009; 57:3930–3941.

27. Lafon, S. Ph.D. thesis. Yale University; New Haven, CT: 2004. Diffusion Maps and Geometric Harmonics.

28. MacKay, DJC. Information Theory, Inference and Learning Algorithms. Cambridge University Press; Cambridge, UK: 2004.

29. Nadakuditi RR, Benaych-Georges F. The breakdown point of signal subspace estimation. 2010 IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM). 2010:177–180.

30. Nadler B. Finite sample approximation results for principal component analysis: A matrix perturbation approach. Ann. Statist. 2008; 36:2791–2817.

31. Natterer, F. The Mathematics of Computerized Tomography. SIAM; Philadelphia: 2001.

32. Natterer, F.; Wübbeling, F. Mathematical Methods in Image Reconstruction. SIAM; Philadelphia: 2001.

33. Paul D. Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. Statist. Sinica. 2007; 17:1617–1642.

34. Pearson K. On lines and planes of closest fit to systems of points in space. Philosophical Magazine. 1901; 2:559–572.

35. Reed, M.; Simon, B. Methods of Modern Mathematical Physics, Volume 1, Functional Analysis. Academic Press; San Diego, CA: 1981.

36. Singer A. From graph to manifold Laplacian: The convergence rate. Appl. Comput. Harmon. Anal. 2006; 21:128–134.

37. Watts DJ, Strogatz SH. Collective dynamics of 'small-world' networks. Nature. 1998; 393:440–442. [PubMed: 9623998]

38. Wiener, N. Extrapolation, Interpolation, and Smoothing of Stationary Time Series: With Engineering Applications. The MIT Press; Cambridge, MA: 1964.
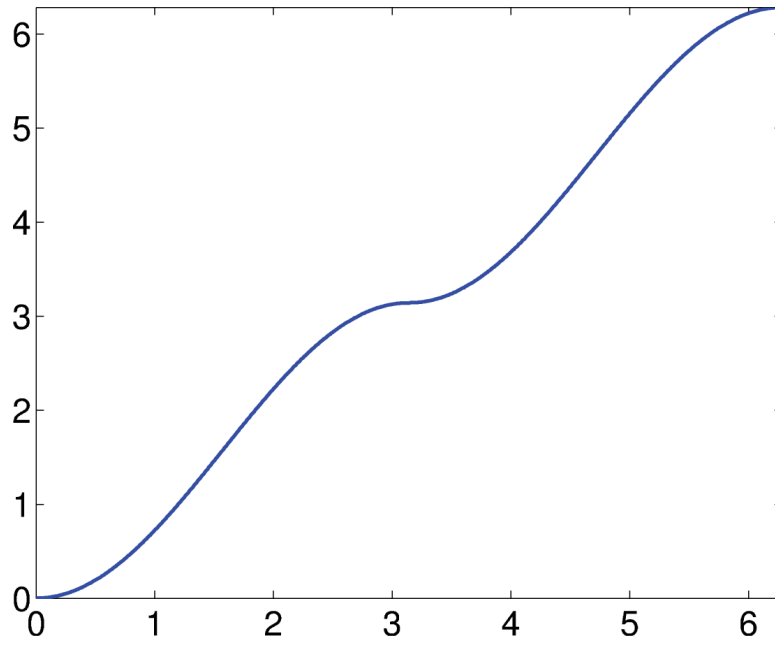
Figure 1.
An example of a nonrigid transformation g (both axes are from 0 *to* 2π).
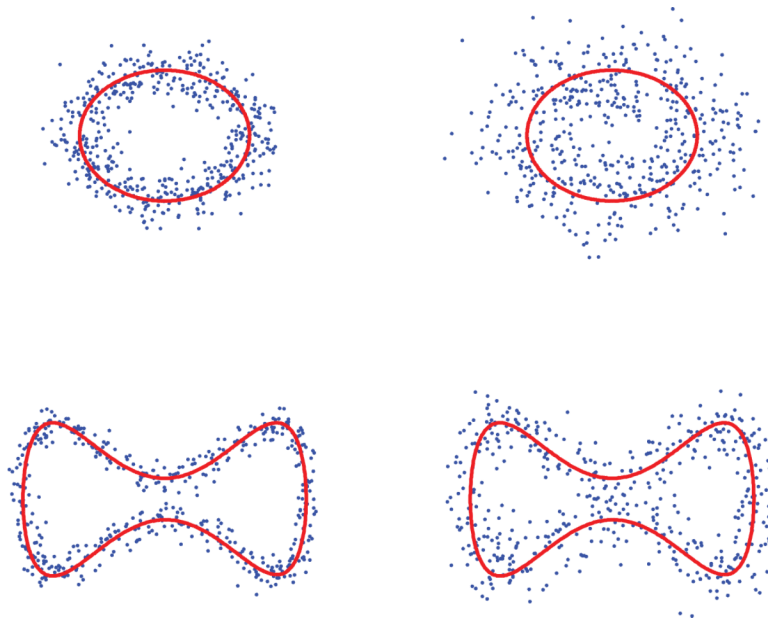
Figure 2.
Upper row: The data (blue points) sampled from a circle (red line) at different noise levels.
Lower row: The data (blue points) sampled from a closed curve (red line) at different noise levels.
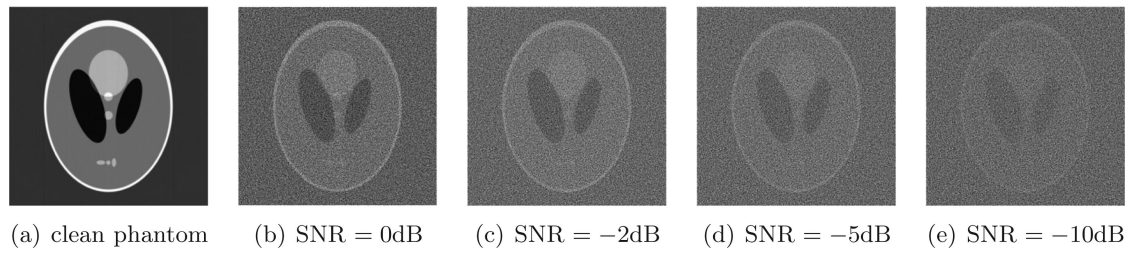
(a) clean phantom   (b) SNR = 0dB   (c) SNR = −2dB   (d) SNR = −5dB   (e) SNR = −10dB

Figure 3.
Reconstruction of the Shepp–Logan phantom image from projections with known beaming directions at different levels of noise.
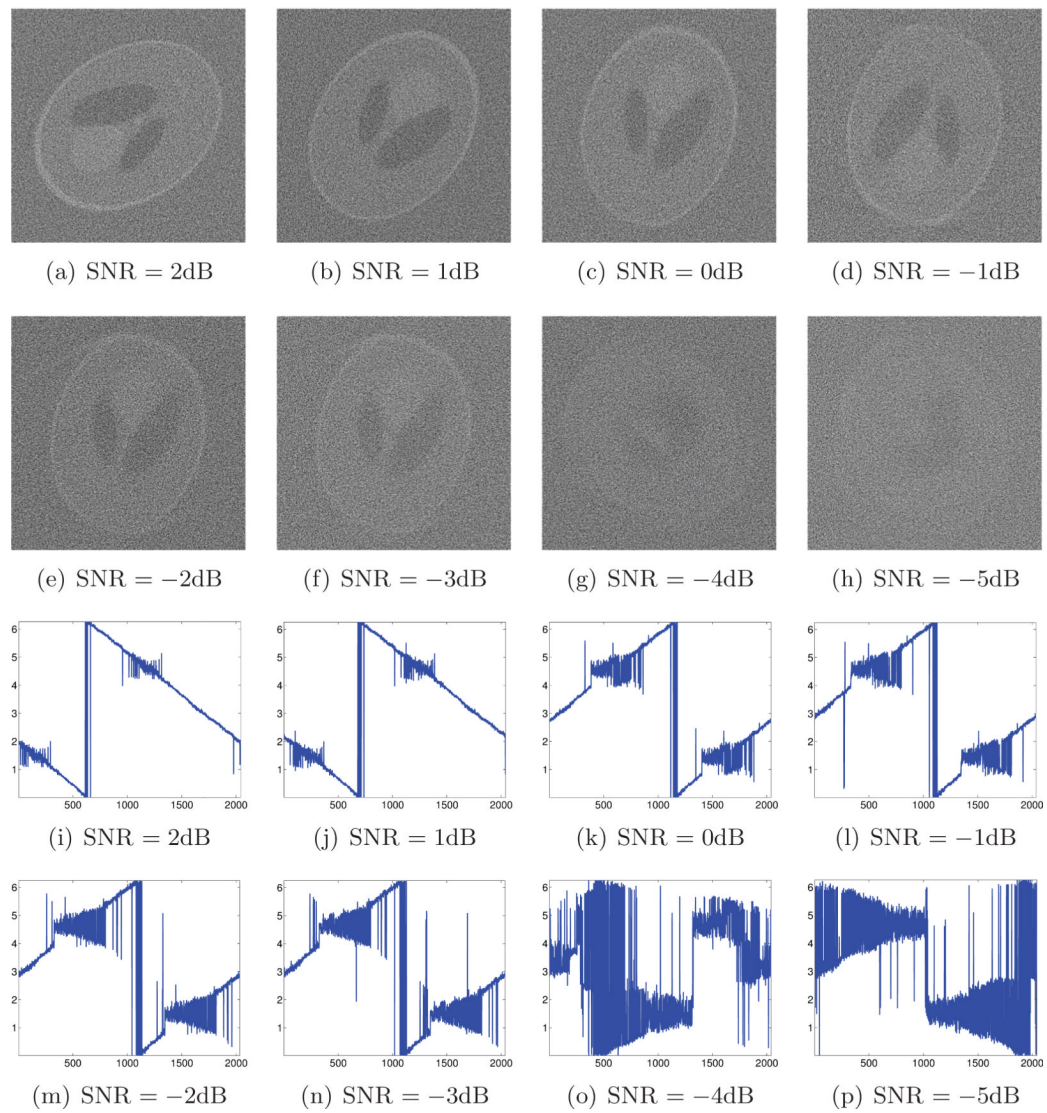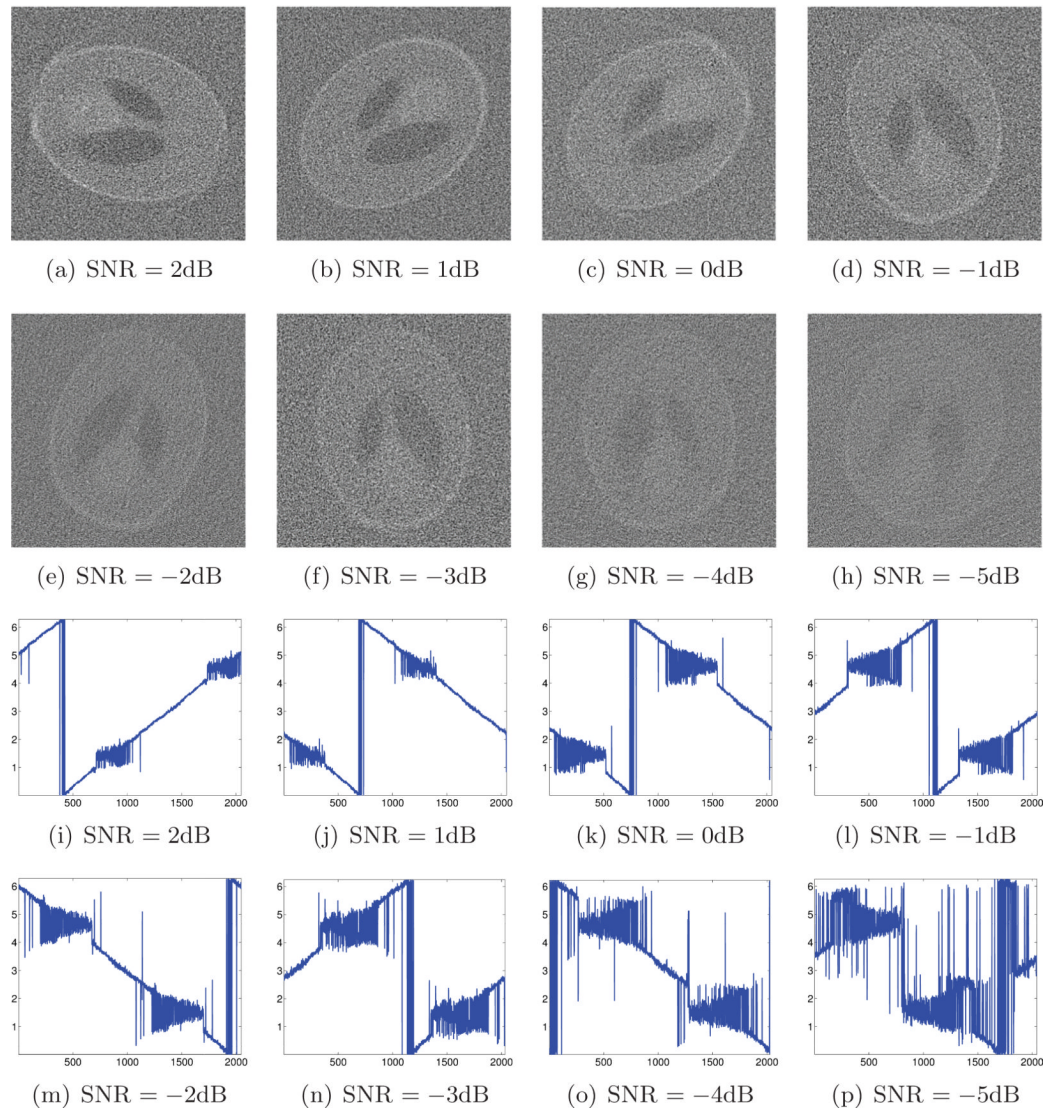
Figure 4.
Top: Reconstruction from noisy projections at unknown directions using the algorithm described in section 6 *(excluding Step 6) at different levels of noise with fixed α = 6 and β =* 0.5. Bottom: Estimated beaming directions (y-axis) against their correct ordering (x-axis).

Figure 5.
Top: Reconstruction from noisy projections at unknown directions using the algorithm described in section 6 *(including Step* 6) at different levels of noise. Bottom: Estimated beaming directions (y-axis) against their correct ordering (x-axis).

(a) SNR = 2dB   (b) SNR = 1dB   (c) SNR = 0dB   (d) SNR = −1dB

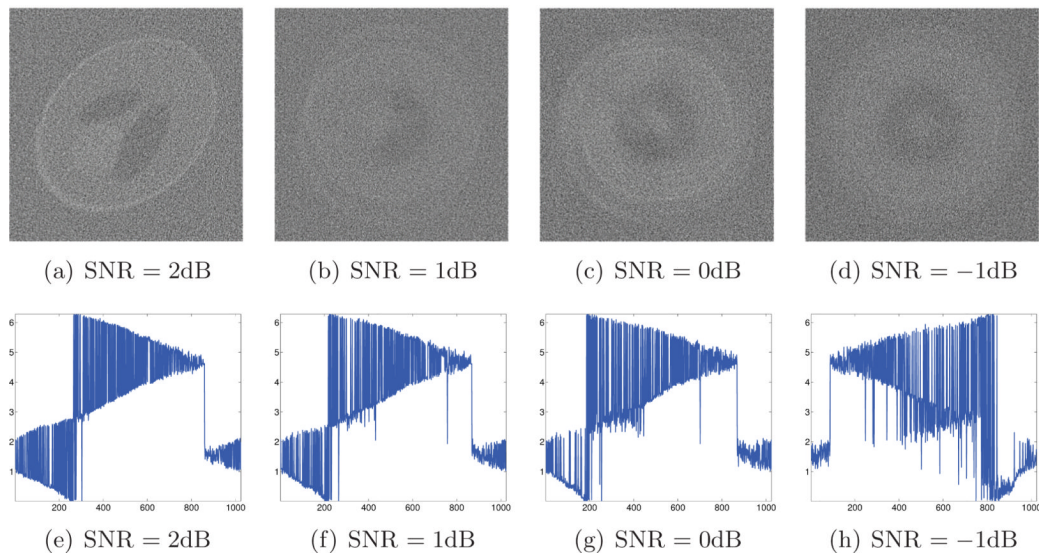(e) SNR = 2dB   (f) SNR = 1dB   (g) SNR = 0dB   (h) SNR = −1dB

Figure 6.
Top: Reconstructions from noisy projections at unknown directions using the algorithm described in [12] for different levels of noise. Bottom: Estimated beaming directions (y-axis) against their correct ordering (x-axis).
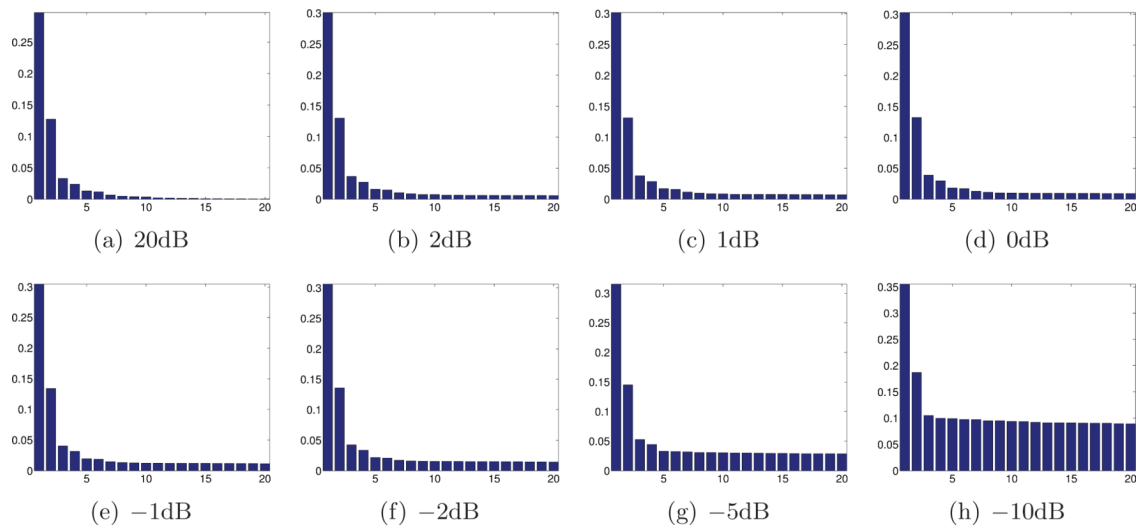
Figure 7.
Bar plots of the first 20 eigenvalues of the sample covariance matrix corresponding to noisy projections at different levels of noise. The numbers of significant principal components K̂ determined by [25] *are* 49, 12, 10, 10, 8, 7, 6, *and* 3.
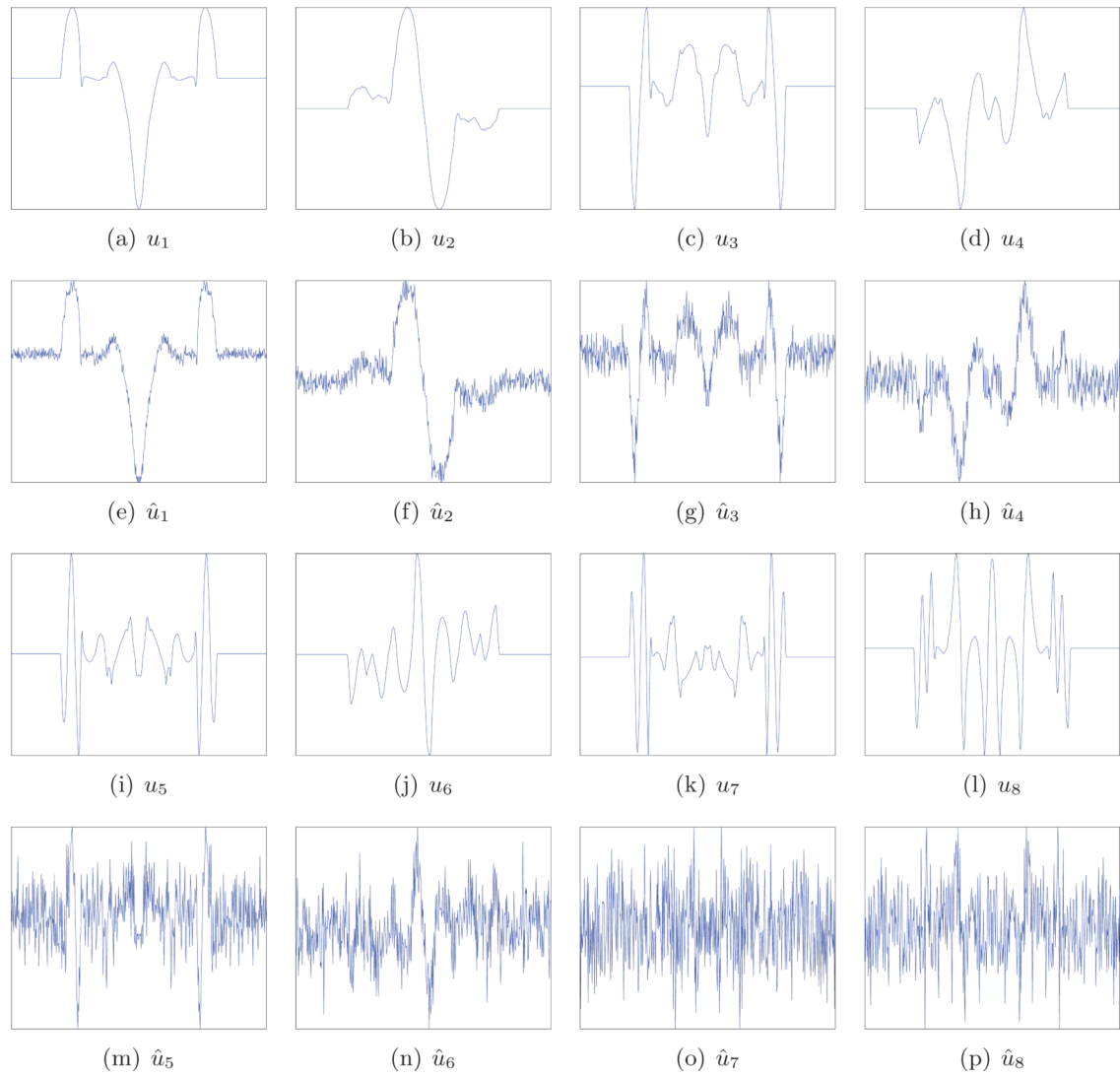
Figure 8.

The first eight principal components for clean projections and the first eight principal components for noisy projections with SNR = –5dB. Note that the principal components are determined up to an arbitrary sign, and we choose the signs so that corresponding pairs of components are positively correlated.
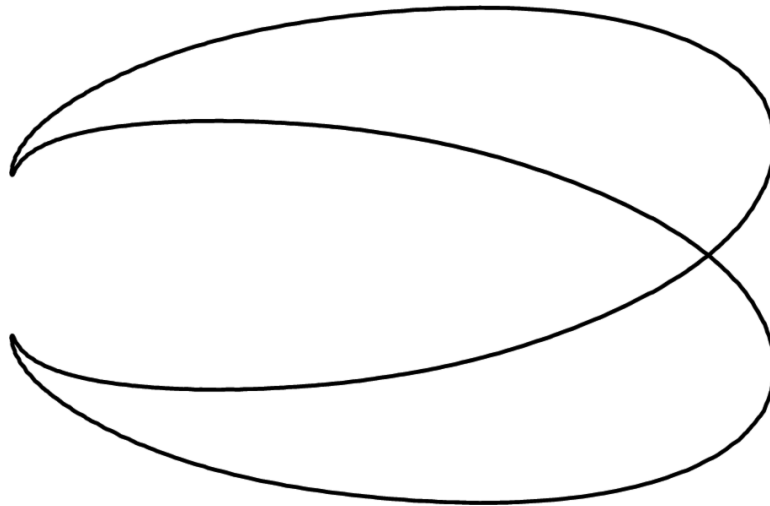
Figure 9.
Projecting the curve 𝒞 of clean projections of the Shepp–Logan phantom onto the linear subspace spanned by the top two principal components. The projected curve has a nontrivial self-intersection, implying the insufficiency of just two principal components.

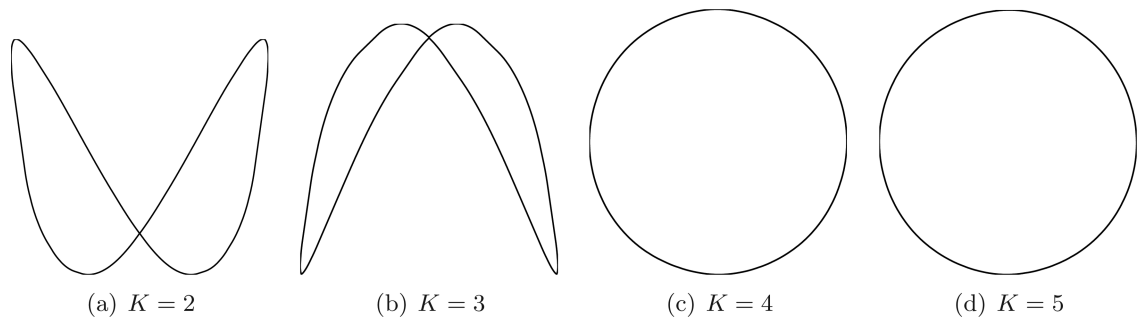(a) $K = 2$ (b) $K = 3$ (c) $K = 4$ (d) $K = 5$

Figure 10.
The diffusion map embedding of $\mathscr{C}$ after the latter was mean shifted and projected onto the linear subspace of the top K principal components (K = 2, 3, 4, 5). As expected from the parity sequence of the principal components, at least four components are required to avoid nontrivial self-intersections.
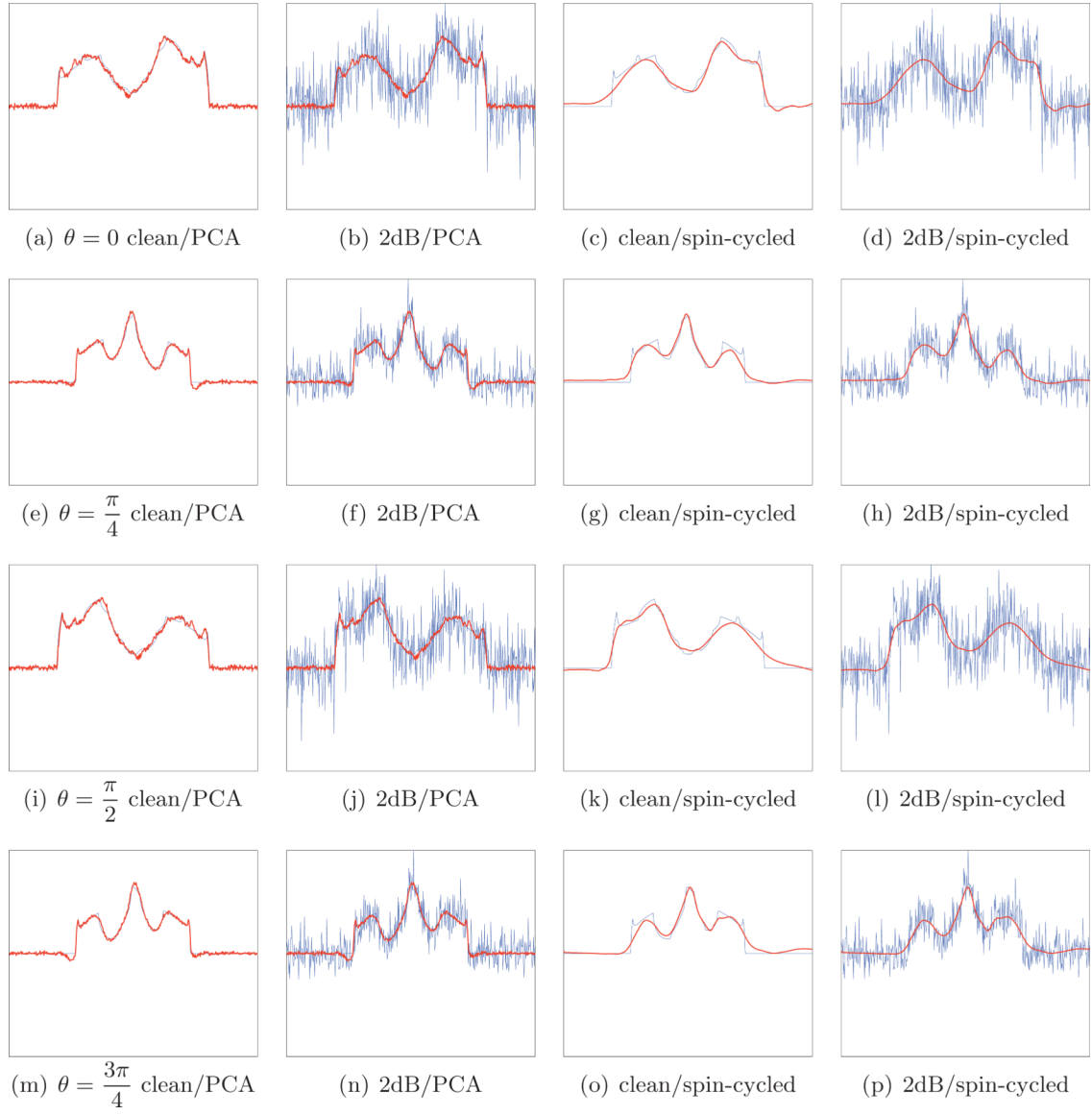
(a) $\theta = 0$ clean/PCA (b) 2dB/PCA (c) clean/spin-cycled (d) 2dB/spin-cycled

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

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

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

Figure 11.
Comparison between the combined PCA-Wiener filtering and wavelet denoising for four different noisy projections with SNR = *2dB taken at* θ=0, $\frac{\pi}{4}$, $\frac{\pi}{2}$, *and* $\frac{3\pi}{4}$. (a) Clean projection (blue) and the PCA denoising of the noisy projection (red); (b) noisy projection (blue) and its filtered version using PCA (red); (c) clean projection (blue) and the wavelet denoising of the noisy projection (red); (d) noisy projection (blue) and its filtered version using wavelets (red). The number of principal components used by the Wiener filter is 12. *RMSE:* 0.464 *for the PCA-Wiener scheme and* 0.612 *for wavelets.*
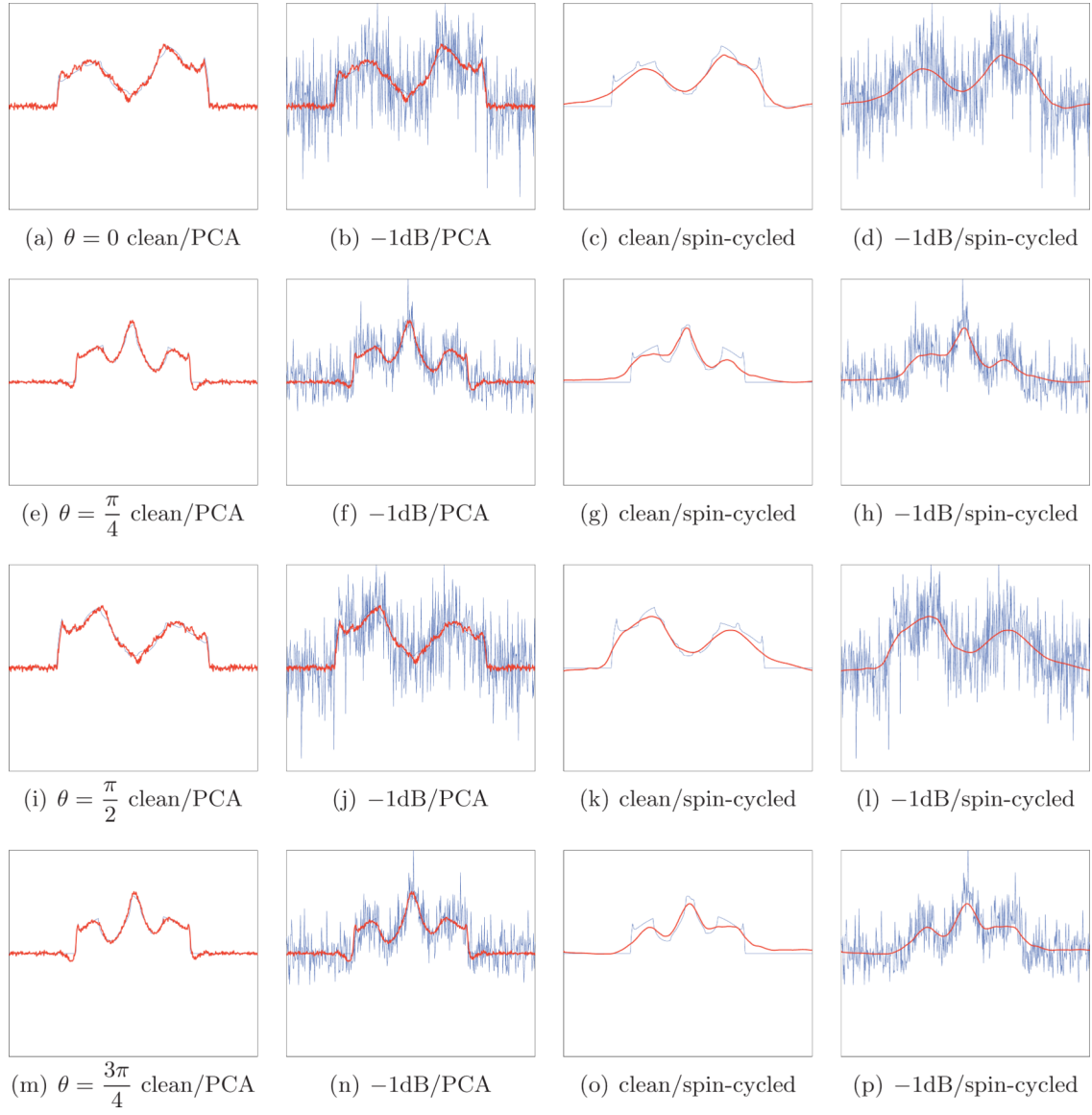
(a) $\theta = 0$ clean/PCA   (b) -1dB/PCA (c) clean/spin-cycled   (d) -1dB/spin-cycled

(e) $\theta = \dfrac{\pi}{4}$ clean/PCA   (f) -1dB/PCA (g) clean/spin-cycled   (h) -1dB/spin-cycled

(i) $\theta = \dfrac{\pi}{2}$ clean/PCA   (j) -1dB/PCA (k) clean/spin-cycled   (l) -1dB/spin-cycled

(m) $\theta = \dfrac{3\pi}{4}$ clean/PCA   (n) -1dB/PCA (o) clean/spin-cycled   (p) -1dB/spin-cycled

Figure 12.
Comparison between the combined PCA-Wiener filtering and wavelet denoising for four different noisy projections with SNR = $-1dB$ *taken at* $\theta{=}0$, $\frac{\pi}{4}$, $\frac{\pi}{2}$, *and* $\frac{3\pi}{4}$. (a) Clean projection (blue) and the PCA denoising of the noisy projection (red); (b) noisy projection (blue) and its filtered version using PCA (red); (c) clean projection (blue) and the wavelet denoising of the noisy projection (red); (d) noisy projection (blue) and its filtered version using wavelets (red). The number of principal components used by the Wiener filter is 8. *RMSE:* 0.514 *for the PCA-Wiener scheme and* 0.745 *for wavelets.*
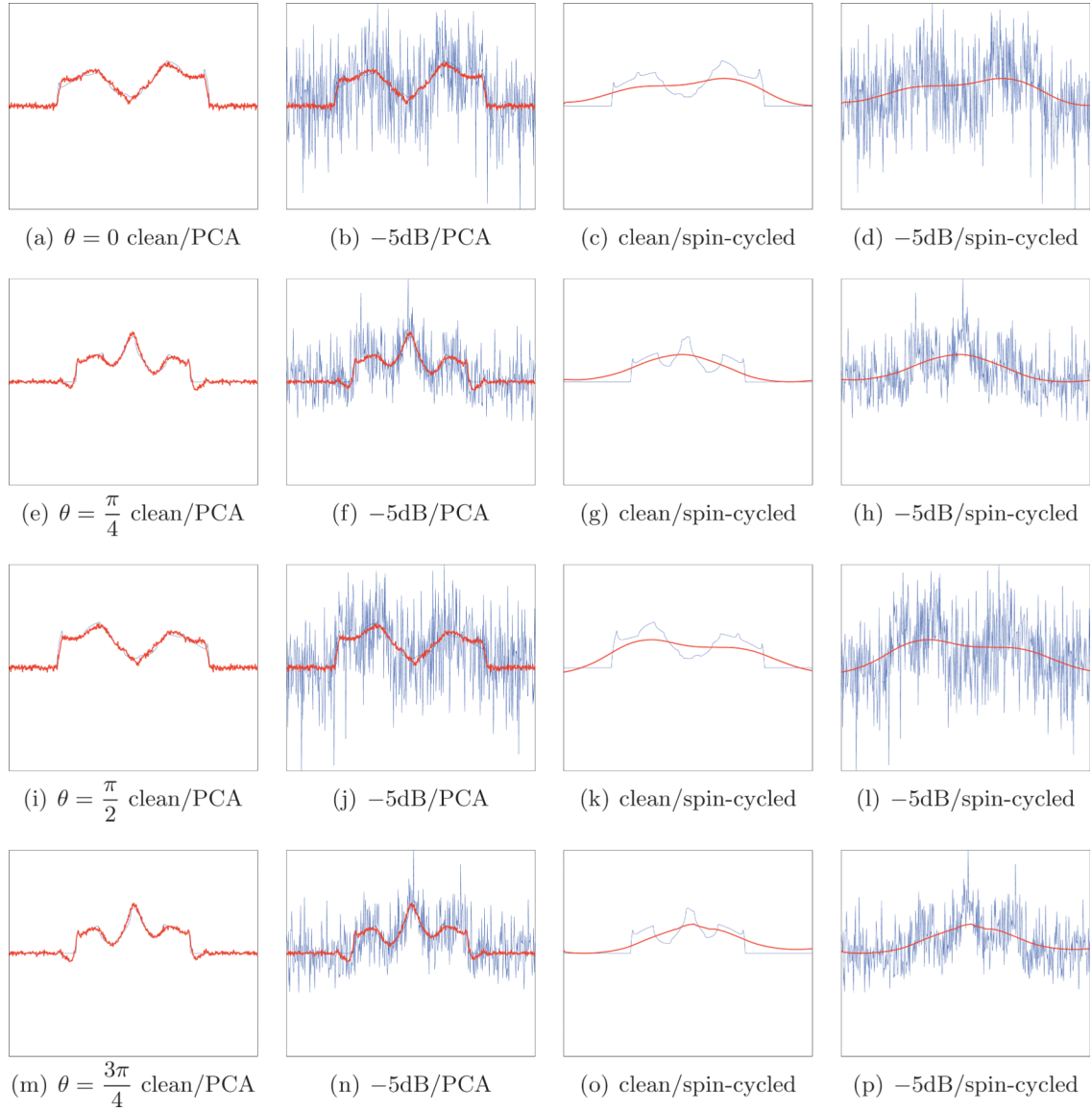
(a) $\theta = 0$ clean/PCA   (b) -5dB/PCA   (c) clean/spin-cycled   (d) -5dB/spin-cycled

(e) $\theta = \frac{\pi}{4}$ clean/PCA   (f) -5dB/PCA   (g) clean/spin-cycled   (h) -5dB/spin-cycled

(i) $\theta = \frac{\pi}{2}$ clean/PCA   (j) -5dB/PCA   (k) clean/spin-cycled   (l) -5dB/spin-cycled

(m) $\theta = \frac{3\pi}{4}$ clean/PCA   (n) -5dB/PCA   (o) clean/spin-cycled   (p) -5dB/spin-cycled

Figure 13.
Comparison between the combined PCA-Wiener filtering and wavelet denoising for four different noisy projections with SNR = $-5dB$ *taken at* θ=0, $\frac{\pi}{4}$, $\frac{\pi}{2}$, *and* $\frac{3\pi}{4}$. (a) Clean projection (blue) and the PCA denoising of the noisy projection (red); (b) noisy projection (blue) and its filtered version using PCA (red); (c) clean projection (blue) and the wavelet denoising of the noisy projection (red); (d) noisy projection (blue) and its filtered version using wavelets (red). The number of principal components used by the Wiener filter is 6. *RMSE:* 0.587 *for the PCA-Wiener scheme and* 0.919 *for wavelets.*
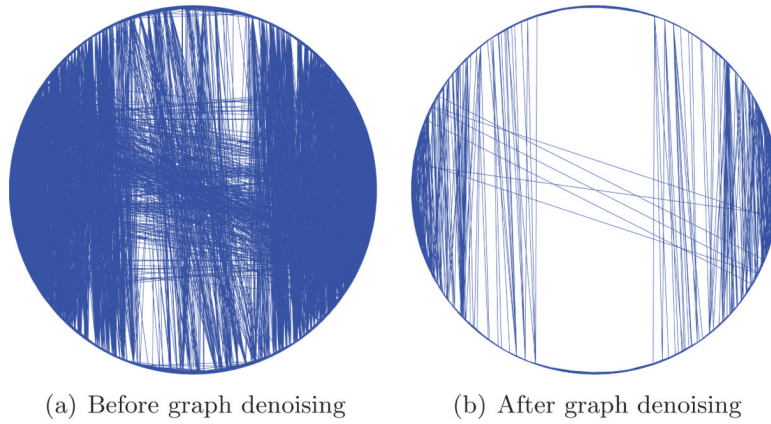
(a) Before graph denoising   (b) After graph denoising

Figure 14.
The effect of graph denoising (Step 3) for SNR = –4dB and α = 6 by thresholding edges whose Jaccard index is below β = 0.5. (a) *The graph after Step 2 with edges given by W^α* (6.2), *and* (b) *the graph after Step 3 with edges given by $W^{\alpha,\beta}$* (6.3).
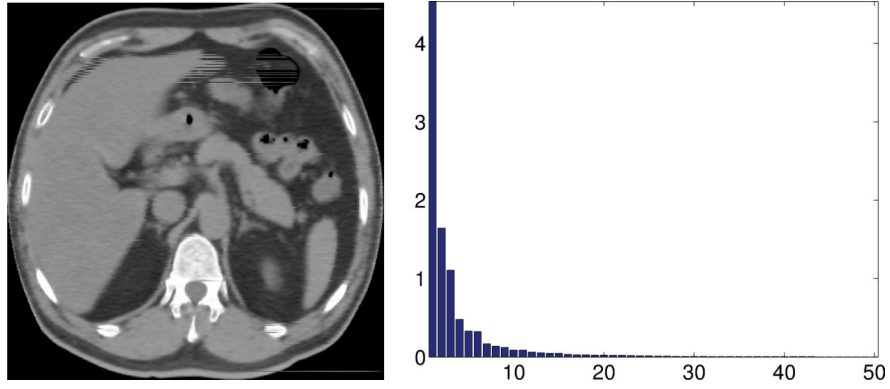
Figure 15.
Left: The abdominal CT image. Right: The first 50 eigenvalues of the covariance matrix of its clean projections.

(a) u_1 (b) u_2 (c) u_3 (d) u_4

(e) $\hat{u}_1$ (f) $\hat{u}_2$ (g) $\hat{u}_3$ (h) $\hat{u}_4$

(i) u_5 (j) u_6 (k) u_7 (l) u_8

(m) $\hat{u}_5$ (n) $\hat{u}_6$ (o) $\hat{u}_7$ (p) $\hat{u}_8$
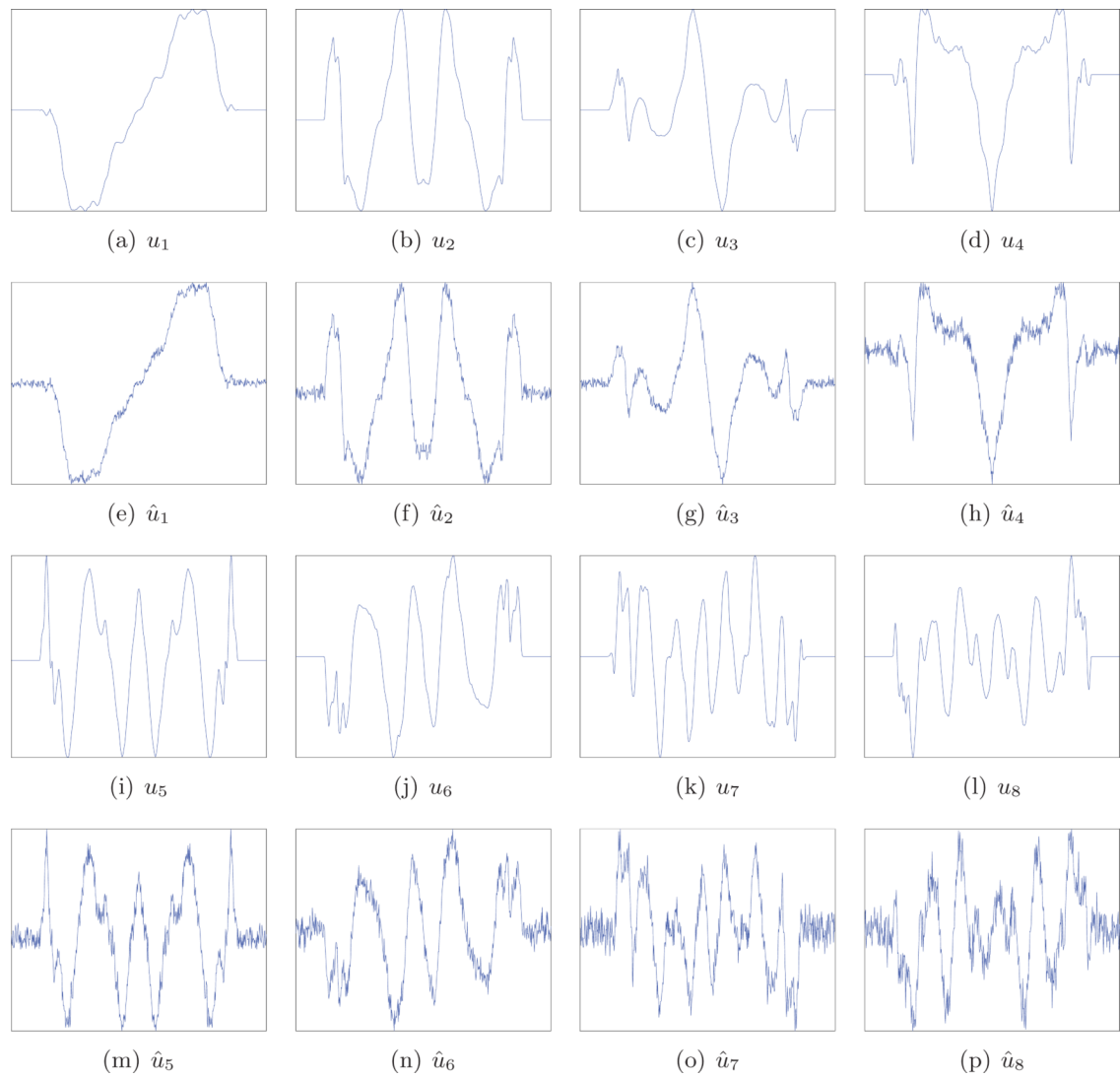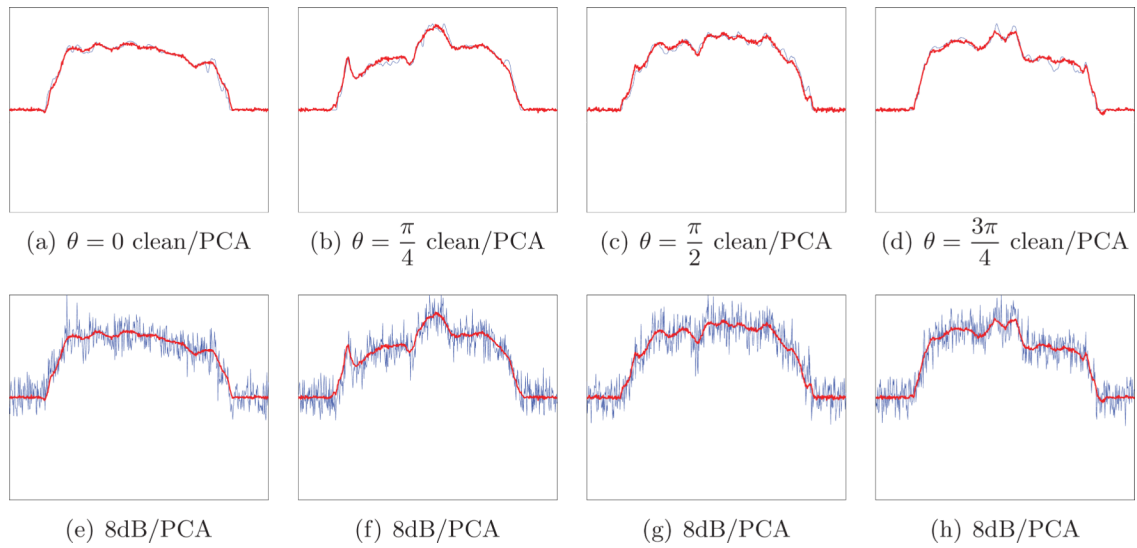
Figure 16.
The first eight principal components for clean projections of the abdominal CT image and the first eight principal components for noisy projections with SNR = 8dB. Notice that the principal components are determined up to an arbitrary sign, and we choose the signs so that corresponding pairs of components are positively correlated.

(a) $\theta = 0$ clean/PCA   (b) $\theta = \dfrac{\pi}{4}$ clean/PCA   (c) $\theta = \dfrac{\pi}{2}$ clean/PCA   (d) $\theta = \dfrac{3\pi}{4}$ clean/PCA

(e) 8dB/PCA   (f) 8dB/PCA   (g) 8dB/PCA   (h) 8dB/PCA

Figure 17.
Results of the combined PCA-Wiener filtering for four different noisy projections with SNR = 8*dB taken at* $\theta{=}0, \frac{\pi}{4}, \frac{\pi}{2}$, *and* $\frac{3\pi}{4}$. Top: Clean projection (blue) and the PCA denoising of the noisy projection (red). Bottom: Noisy projection (blue) its filtered version using PCA (red).
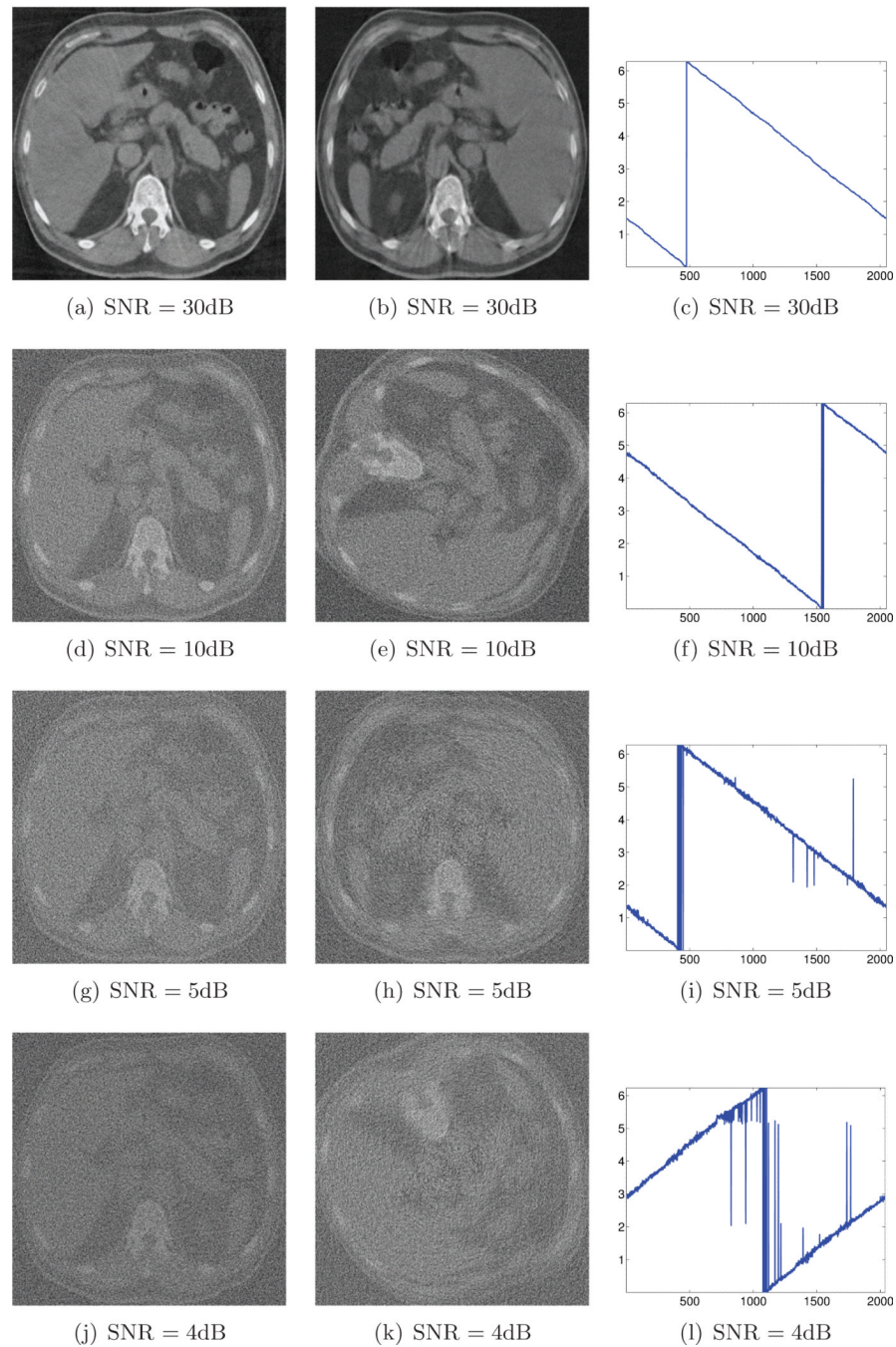
Figure 18.
Left column: Reconstructions from noisy projections at known directions for different levels of noise. Middle column: Reconstructions from noisy projections at unknown directions using the proposed algorithm for different levels of noise. Right column: Estimated beaming directions (y-axis) against their correct ordering (x-axis). The number of principal components used by the Wiener filter is 95, 21, 15, *and* 13 *for the noise level* 30*dB*, 10*dB*, 5*dB, and* 4*dB*.

(a) SNR = 3dB   (b) SNR = 3dB   (c) SNR = 3dB
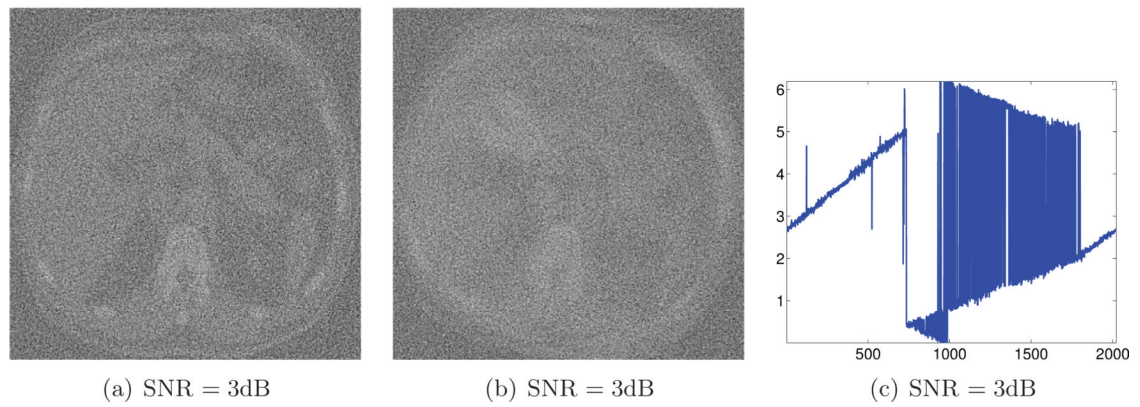
Figure 19.
The algorithm fails when SNR = 3dB. Left: Reconstruction from noisy projections at known directions. Middle: Reconstruction from noisy projections at unknown directions. Right: Estimated beaming directions (y-axis) against their correct ordering (x-axis). The number of principal components used by the Wiener filter is 10.

Table 1

The optimal parameter values for α and β as a function of the SNR.

SNR [dB]	α	β
3	5	0.40
2	5	0.41
1	5	0.42
0	5	0.40
-1	5	0.40
-2	7	0.45
-3	4	0.43
-4	4	0.46
-5	4	0.41

Table 2

Performance analysis of our algorithm. For each level of noise we performed 100 independent runs of the algorithm, corresponding to different independent realizations of noise and beaming directions. The table reports the mean and standard deviation (over 100 runs) of the RMSE for denoising using PCA with the asymptotically optimal linear filter, the optimal parameter values α and β, and the ℓ_1 norm of the wavelet expansion of the reconstructed image.

SNR [dB]	RMSE	a	β	ℓ_1
10	0.350 ± 0.001	7.1 ± 2.5	0.41 ± 0.02	228.9 ± 1.3
5	0.417 ± 0.001	5.2 ± 1.1	0.41 ± 0.02	256.3 ± 1.6
3	0.448 ± 0.001	5.1 ± 0.9	0.44 ± 0.05	266.7 ± 8.4
2	0.464 ± 0.002	5.3 ± 1.0	0.45 ± 0.05	270.4 ± 9.4
1	0.478 ± 0.002	5.3 ± 0.9	0.44 ± 0.05	271.1 ± 8.8
0	0.497 ± 0.002	5.6 ± 1.0	0.44 ± 0.04	271.4 ± 7.7
−1	0.514 ± 0.002	6.0 ± 1.1	0.45 ± 0.05	273.6 ± 8.4
−2	0.532 ± 0.002	6.4 ± 1.4	0.45 ± 0.05	273.1 ± 6.1
−3	0.550 ± 0.002	6.4 ± 1.6	0.45 ± 0.05	275.3 ± 7.1
−4	0.568 ± 0.002	6.3 ± 1.8	0.47 ± 0.06	284.3 ± 8.8
−5	0.588 ± 0.002	7.5 ± 2.1	0.53 ± 0.06	292.0 ± 0.2