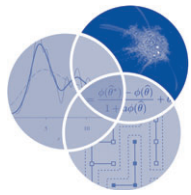
Advanced Review



# Robust dimension reduction

Shojaeddin Chenouri, Jiaxi Liang and Christopher G. Small*

Information in the data often has far fewer degrees of freedom than the number of variables encoding the data. Dimensionality reduction attempts to reduce the number of variables used to describe the data. In this article, we shall survey some dimension reduction techniques that are robust. We consider linear dimension reduction first and describe robust principal component analysis (PCA) using three approaches. The first approach uses a singular value decomposition of a robust covariance matrix. The second approach employs robust measures of dispersion to realize PCA as a robust projection pursuit. The third approach uses a low-rank plus sparse decomposition of the data matrix.

We also survey robust approaches to nonlinear dimension reduction under a unifying framework of kernel PCA. By using a kernel trick, the robust methods available for PCA can be extended to nonlinear cases. 

**How to cite this article:**
*WIREs Comput Stat* 2015, 7:63–69. doi: 10.1002/wics.1331

**Keywords:** principal component analysis; outlier; robust statistics; kernel; manifold; dimension reduction

## INTRODUCTION

Many datasets involving images, DNA microarrays, documents, etc. are high-dimensional in the sense that they have a large number of variables. For this reason, they can be represented as points in some high-dimensional space $\mathbb{R}^D$. For example, an image in raw format might be encoded as a $256 \times 256$ matrix. Upon vectorization, this image can be defined as a vector or point in dimension $D = 256^2$. A dataset of $n$ such images can be encoded as a set of $n$ points in $\mathbb{R}^D$.

However, the information in the data often has far fewer degrees of freedom than the number of variables used to describe the data. This implies that the data really lie close to some restricted lower-dimensional subset of $\mathbb{R}^D$. This subset may be assumed to be an unknown Riemannian submanifold $\mathbb{M} \subset \mathbb{R}^D$, with the data $\mathbf{x}_1, \ldots, \mathbf{x}_n$ lying on or near $\mathbb{M}$. Typically $\mathbb{M}$ is of dimension $d$ where $d \ll D$. In this setting, dimensionality reduction attempts to represent $\mathbf{x}_1, \ldots, \mathbf{x}_n$ by a corresponding set $\mathbf{y}_1, \ldots, \mathbf{y}_n \in \mathbb{R}^d$ such that the interpoint Euclidean distance between $\mathbf{y}_i$ and $\mathbf{y}_j$ approximates the interpoint geodesic distance between $\mathbf{x}_i$ and $\mathbf{x}_j$ on $\mathbb{M}$ for all $i$ and $j$. It is assumed that each point $\mathbf{x}_i \in \mathbb{R}^D$ is related to a latent point $\mathbf{y}_i \in \mathbb{R}^d$ through an equation $\mathbf{x}_i = g(\mathbf{y}_i) + \epsilon_i$, where $g(\mathbf{y}_i) \in \mathbb{M}$, and $\epsilon_i$ is some small error vector in $\mathbb{R}^D$. In many, but not all cases, $\epsilon_i$ is set to zero, such that the data lie exactly on the manifold.

When $\mathbb{M}$ is a linear $d$ -flat of $\mathbb{R}^D$, then the dimension reduction problem is also said to be linear. Perhaps the best known tool for linear dimension reduction is principal component analysis (PCA). Nonlinear dimension reduction covers all other cases where the manifold $\mathbb{M}$ is not a $d$-flat of $\mathbb{R}^D$ and is curved. In this case, the curvature is extrinsic to $\mathbb{M}$. That is, it is inherited from the embedding of $\mathbb{M}$ in $\mathbb{R}^D$, and is not intrinsic to the Riemannian geometry of $\mathbb{M}$. In those cases that are ideal for dimension reduction, $\mathbb{M}$ is assumed to be Riemannian isometric to some subset of $\mathbb{R}^d$. The task of a nonlinear dimension reduction algorithm is to 'remove' the extrinsic curvature of the data $\mathbf{x}_i \in \mathbb{M}$ by approximately recovering the latent points $\mathbf{y}_i \in \mathbb{R}^d$. Over the last two decades, many nonlinear methods have been introduced in the literature. These include the

*Correspondence to: cgsmall@uwaterloo.ca

Department of Statistics & Actuarial Science, University of Waterloo, Waterloo, ON, Canada

Conflict of interest: The authors have declared no conflicts of interest for this article.

principal curves and surfaces,[1] Kernel PCA,[2] local linear embedding (LLE),[3] isometric feature mapping (ISOMAP),[4] Laplacian Eigenmap,[5] Hessian Eigenmaps,[6] local tangent space alignment (LTSA),[7] local MDS,[8] stochastic neighbor embedding (SNE),[9] *t*-SNE,[10] diffusion maps,[11] manifold charting,[12] and multilayer neural networks,[13] among others.

All existing methods require assumptions to be successful in achieving their stated goals. For example, the ISOMAP algorithm ideally requires that $\mathbb{M}$ be geodesically convex. Although geodesic convexity is not a requirement for the success of LLE, examples with holes can have distortions from the ideal embedding function (see Ref 6, p. 5594). In Ref 14 it is also noted that the LLE algorithm has problems with nonuniform distributions on the manifold and is sensitive to noise. A variety of modifications have been proposed to fix these problems leading to algorithms such as Hessian local linear embedding (HLLE) by Donoho and Grimes,[6] robust local linear embedding (RLLE) by Chang and Yeung,[15] modified local linear embedding (MLLE) by Zhang and Wang,[16] local linear transformation embedding (LLTE) by Hou et al.[14] Another approach that is similar in some of its steps to LLE is the Laplacian Eigenmap algorithm by Belkin and Niyogi.[5] At this stage, there is no one method that clearly stands out as the complete solution to the dimension reduction problem.

Even when a method is successful under its stated assumptions, we must also study its performance in the presence of small departures from those assumptions. A method that performs well under small departures from standard assumptions is said to be *robust* in the statistics literature or *stable* in the engineering literature. Most of the literature on robust dimension reduction focuses on the robustness of methods against outliers. Therefore, this survey will concentrate on this. To motivate the problem of nonrobust dimension reduction, consider the PCA displayed in Figure 1. On the left-hand side, the data set has no outlying point, and the PCA provides a satisfactory decomposition of the two orthogonal sources of variation in the data. On the right-hand side, an outlying point has been added. The effect of adding this point is to rotate the principal axes substantially such that the axes are misaligned with most of the data. This is an illustration of the failure of standard PCA to be robust against outliers.

## LINEAR METHODS

Consider the $n \times D$ data matrix $\mathbf{X}$ which has been centered such that the column averages are all 0. Let $||\mathbf{u}||$ denote the Euclidean norm of the column vector $\mathbf{u}$. Let $\mathbf{S} = (n-1)^{-1}\mathbf{X}'\mathbf{X}$ be the sample covariance matrix. The first step of PCA is to find the first principal axis of the data as the solution of the optimization:

$$\underset{||\mathbf{u}||=1}{\operatorname{argmax}} ||\mathbf{S}^{1/2}\mathbf{u}||^2. \tag{1}$$

The $k$th principal axis is calculated by a similar maximization constrained to be orthogonal to the first $k-1$ principal axes.

In seeking to robustify PCA, we can examine two nonrobust aspects of this optimization problem. Firstly, the covariance matrix $\mathbf{S}$ that appears inside the norm is sensitive to outliers and could be replaced by a robust alternative. A second robustification of the optimization problem could be achieved by replacing the Euclidean norm by a norm less sensitive to outliers. The first of these leads to robust PCA using a robust estimate of the covariance matrix. The second of these leads to robust PCA using robust projection pursuit.

We review the literature on the robust covariance matrix approach first. This approach dates back to Maronna[17] and Campbell[18] who proposed using affine equivariant M-estimators of the covariance matrix. However, M-estimators have breakdown value at most $1/D$. Therefore, they can only handle a small proportion of outliers when $D$ is sufficiently large (see Ref 19). Croux and Haesbroeck[20] proposed using high-breakdown affine equivariant estimators of the covariance matrix such as the minimum volume ellipsoid (MVE) and minimum covariance determinant (MCD) methods of Rousseeuw[21,22] as well as S-estimators of Davies[23] and Rousseeuw and Leroy.[24] Although they are very robust, the problem of these methods is that they are computationally expensive and therefore limited only to small to moderate dimensions. The fastest algorithms to date can only handle up to about 100 dimensions (see Ref 25).

A second approach to make PCA robust is projection pursuit. In this approach, one maximizes a robust univariate measure of dispersion in successive orthogonal directions. Doing this, we bypass the need to robustly estimate the covariance matrix. Some papers on this approach are Li and Chen,[26] Croux and Ruiz-Gazen,[27] Hubert et al.,[28] Maronna.[29] Hubert and Rousseeuw[25] combined the advantages of both projection pursuit and high-breakdown covariance estimators. They proposed to first use projection pursuit to reduce the dimensionality to some moderate size and then to apply PCA using MCD estimators of the covariance matrix.

Let $\mathbf{X} = \mathbf{L} + (\mathbf{X} - \mathbf{L})$, where $\mathbf{L}$ is a matrix of rank $d$. Then PCA can also be written as a matrix

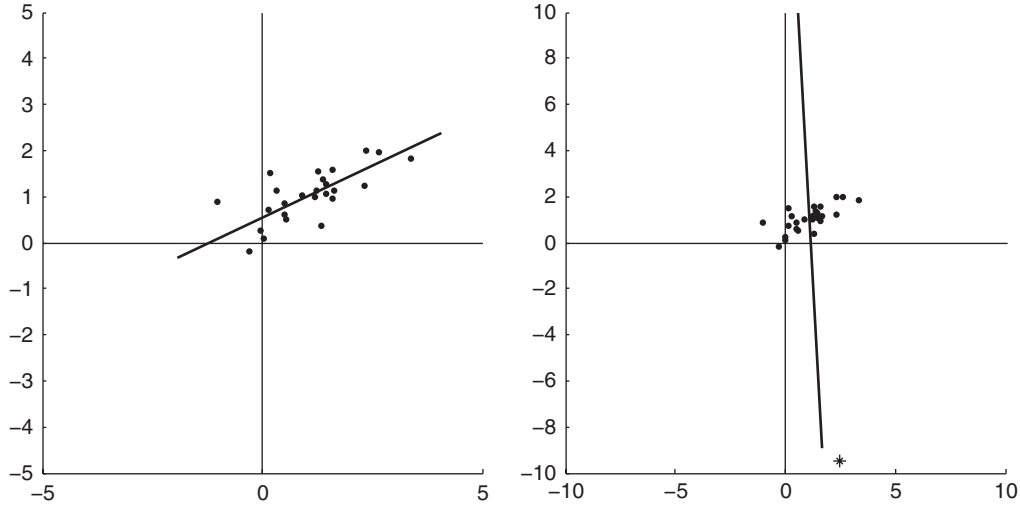



**FIGURE 1 |** The effect of a single outlier on principal component analysis (PCA).

approximation problem. It is the solution to the minimization problem:

$$\arg\min_{\mathbf{L}} ||\mathbf{X}-\mathbf{L}||_F \quad \text{subject to} \quad \text{rank}(\mathbf{L}) = d, \qquad (2)$$

where the matrix norm $||\cdot||_F$ is the Frobenius norm. In practice, however, the rank $d$ is not known and is usually estimated by the number of large singular values of $\mathbf{X}$. This estimated rank is often referred as the *numerical rank* of $\mathbf{L}$. In this minimization, we can robustify PCA by choosing a norm less sensitive to outliers.

Candés et al.[30] and Chandrasekaran et al.[31] independently considered a robustification of the formulation in Eq. (2). But unlike the residual matrix in traditional PCA, they assumed that the matrix $\mathbf{Z}=\mathbf{X}-\mathbf{L}$ is sparse (with entries that can be arbitrarily large in magnitude) in addition to the assumption that $\mathbf{L}$ is low-rank. In this setting, they considered an idealized version of Robust PCA by recovering $\mathbf{L}$ and $\mathbf{Z}$ from $\mathbf{X}$ using a tractable convex optimization, with a cost not much higher than that of the classical PCA. The optimization problem they considered is:

$$\arg\min_{\mathbf{L},\mathbf{Z}} \left\{||\mathbf{L}||_* + \lambda||\mathbf{Z}||_1\right\} \quad \text{subject to} \quad \mathbf{L}+\mathbf{Z}=\mathbf{X}. \qquad (3)$$

Note that in this optimization $\mathbf{L}$ is not constrained to be low-rank and $\mathbf{Z}$ is not constrained to be sparse. These respective properties are obtained at the solution of the optimization. In this expression:

$$||\mathbf{G}||_* = \sum_i \sigma_i(\mathbf{G}) \qquad \text{and} \qquad ||\mathbf{G}||_1 = \sum_{jk}|G_{jk}|$$

where $\sigma_i(\mathbf{G})$ is the $i$th singular value of the matrix $\mathbf{G}=\left(G_{jk}\right)$. Here, $||\mathbf{G}||_*$ is called the nuclear norm and $||\mathbf{G}||_1$ is called the 1-norm of the matrix $\mathbf{G}$. There are some identifiability issues regarding the decomposition $\mathbf{X}=\mathbf{L}+\mathbf{Z}$ that need to be addressed. The decomposition is not identifiable if $\mathbf{L}$ is both low-rank and sparse. To avoid this case, Candés et al.,[30] imposed the condition that the low-rank matrix $\mathbf{L}$ is not sparse. To do this they borrowed the notion of incoherence in Candés and Recht,[32] for the matrix completion problem, to implement this assumption. Another type of identifiability issue arises when the sparse matrix $\mathbf{Z}$ is both low-rank and sparse. In this case Candés et al.[30] assumed that the sparsity pattern of $\mathbf{Z}$ is selected uniformly at random.

Under the aforementioned conditions, and some other rather weak assumptions, they showed that, quite surprisingly, the solution of Eq. (3) is in fact the exact recovery of the low-rank matrix $\mathbf{L}$ and the sparse matrix $\mathbf{Z}$. The approach by Candés et al.[30] also allows for missing entries in $\mathbf{X}$.

In what sense is this method robust? Suppose that $\mathbf{L}$ is any low-rank matrix, and $\mathbf{Z}$ is sparse. Let us write $\mathbf{X}_c=\mathbf{L}+c\mathbf{Z}$, where $c>0$. Letting $c\to\infty$ has the effect of increasing the sparse noise matrix such that its nonzero entries are arbitrarily large in size, while the matrix remains sparse. Let $n$ be the smaller of the two dimensions of $\mathbf{L}$. The surprising conclusion of Ref 30 is that, using $\lambda=1/\sqrt{n}$, the exact decomposition from Eq. (3) can be obtained with high probability universally over all choices of $c$. In this sense, the method is robust against the size of the errors.

## NONLINEAR METHODS

While many nonlinear dimensionality reduction methods have been developed, the robustness of these

methods has drawn insufficient attention so far. The most obvious robustness problem is the sensitivity to the presence of outliers. The development of robust PCA procedures has motivated researchers to extend the idea of robust PCA into a nonlinear framework using robust Kernel PCA.

Kernel PCA[33] is a framework that includes a variety of nonlinear dimensionality reduction methods such as locally linear embedding, ISOMAP, Laplacian Eigenmaps (LEM), and some others. It borrows the idea of a kernel from reproducing kernel Hilbert spaces (RKHS). Kernel PCA performs PCA in a feature space that is related to the original input space by some implicit nonlinear mapping. It is hoped that the structure of the observed data can be studied using linear methods in this high-dimensional feature space.

Assume that there exists a map $\Phi : \mathbb{R}^D \to \mathcal{H}$, transforming the observed data into a Hilbert space. Define an $n \times n$ matrix $\mathbf{K}$ by:

$$\mathbf{K}_{ij} = \left\langle \Phi\left(\mathbf{x}_i\right), \Phi\left(\mathbf{x}_j\right) \right\rangle,$$

where $\langle \cdot, \cdot \rangle$ is the inner product in the space $\mathcal{H}$. The matrix $\mathbf{K}$ is positive semidefinite, and it is called a *kernel*.

The traditional PCA is then applied on the transformed data $\{\Phi(\mathbf{x}_1), \ldots, \Phi(\mathbf{x}_n)\}$. To this end, we consider the eigen-decomposition of the covariance matrix:

$$\mathbf{C}_\Phi = \frac{1}{n}\sum_{j=1}^{n} \Phi\left(\mathbf{x}_j\right) \Phi'\left(\mathbf{x}_j\right), \tag{4}$$

where we assume that $\sum_{j=1}^{n} \Phi\left(\mathbf{x}_j\right) = 0$. The low-dimensional subspace is the space spanned by eigenvectors corresponding to the $d$ largest eigenvalues of $\mathbf{C}_\Phi$.

The eigenvalues $\lambda$ and the corresponding eigenvectors $\mathbf{v}$ of $\mathbf{C}_\Phi$ are the solutions to the equation:

$$\mathbf{C}_\Phi \mathbf{v} = \lambda \mathbf{v}. \tag{5}$$

Note that all vectors $\mathbf{v}$ satisfying Eq. (5) lie in the span of $\{\Phi(\mathbf{x}_1), \ldots, \Phi(\mathbf{x}_n)\}$. Thus, we can rewrite $\mathbf{v}$ as:

$$\mathbf{v} = \sum_{i=1}^{n} \alpha_i \Phi\left(\mathbf{x}_i\right), \tag{6}$$

and the problem becomes finding the $\lambda$ and $\alpha = (\alpha_1, \ldots, \alpha_n)'$.

Substituting Eqs (4) and (6) into Eq. (5), we observe that $\lambda$ and $\alpha$ satisfy:

$$n\lambda\alpha = \mathbf{K}\alpha.$$

Therefore, the problem is equivalent to the eigen-decomposition of the kernel $\mathbf{K}$. The so-called 'kernel trick' allows one to obtain the low-dimensional representation $\mathbf{Y}_{n\times d}$ of the data $\mathbf{X}_{n\times D}$ without specifying the nonlinear map $\Phi$:

$$\mathbf{Y} = \mathbf{A}\Lambda^{\frac{1}{2}},$$

where $\Lambda$ is a diagonal matrix of the top $d$ eigenvalues of $\mathbf{K}$, and $\mathbf{A} = [\alpha_1, \ldots, \alpha_d]$ is a $n \times d$ matrix with $\alpha_j$ being the eigenvector of $\mathbf{K}$ corresponding to the $j$ th largest eigenvalue.

Different choices of the kernel $\mathbf{K}$ will result in different low-dimensional representations. It has been shown that many dimensionality reduction methods, such as MDS, ISOMAP, LLE, Laplacian Eigenmap, and diffusion maps, can all be described as special cases under the framework of Kernel PCA.[2] For example, ISOMAP is equivalent to Kernel PCA by choosing the kernel:

$$\widetilde{\mathbf{K}} = -\frac{1}{2}\left(\mathbf{I} - \mathbf{e}\mathbf{e}'\right)\mathbf{D}^{\mathcal{G}}\left(\mathbf{I} - \mathbf{e}\mathbf{e}'\right),$$

where $\mathbf{D}^{\mathcal{G}}$ is the matrix of squared pairwise geodesic distances and $\mathbf{e} = n^{-1/2}(1, \ldots, 1)'$ is the uniform vector of unit length. LLE is equivalent to Kernel PCA by choosing the kernel:

$$\mathbf{K} = \lambda_{\max}\mathbf{I} - \left(\mathbf{I} - \widehat{\mathbf{W}}\right)'\left(\mathbf{I} - \widehat{\mathbf{W}}\right),$$

$$\widetilde{\mathbf{K}} = \left(\mathbf{I} - \mathbf{e}\mathbf{e}'\right)\mathbf{K}\left(\mathbf{I} - \mathbf{e}\mathbf{e}'\right),$$

where $\widehat{\mathbf{W}}$ is the matrix of coefficients in the LLE algorithm, and $\lambda_{\max}$ is the largest eigenvalue of $\left(\mathbf{I} - \widehat{\mathbf{W}}\right)'\left(\mathbf{I} - \widehat{\mathbf{W}}\right)$.

Huang et al.[34] and Debruyne et al.[35] independently studied influence functions of the eigenvalues and eigenvectors of Kernel PCA. They showed that for unbounded kernels, these influence functions are also unbounded, indicating that Kernel PCA can be very sensitive to the presence of outliers when certain kernels are chosen.

To robustify the Kernel PCA against outliers, Huang et al.[34] proposed a class of robust Kernel PCA variants. Similar to the previous work in robust PCA, the proposed method is based on the eigen-decomposition of a robust estimator of the covariance matrix. An iterative reweighted algorithm was employed to estimate the mean function and kernel, enforcing a down-weighting of outliers. The eigen-decomposition is then performed on the weighted kernel. The influence functions of their

robust Kernel PCA are shown to be bounded, and the numerical experiments also indicate that the proposed method is more resistant to outliers than traditional Kernel PCA. Debruyne et al.[35,36] also proposed three robust Kernel PCA algorithms, which generalize spherical PCA,[37] projection pursuit,[38] and ROBPCA,[25] into the kernel feature space. In Ref 35 a simple graphical procedure was proposed to detect and visualize the influential observations in ordinary Kernel PCA. Other related papers include Wang et al.,[39] Deng et al.,[40] Pang et al.,[41] and Huang and Yeh.[42]

Another possible way to improve the robustness against outliers is by modifying the loss function. Let $\mathcal{PS}$ be the principal $d$ -dimensional subspace of the feature space $\mathcal{H}$. For a data point $\mathbf{x}$, traditional Kernel PCA is in search of a point $\mathbf{z}$ that minimizes the squared norm:

$$\arg\min_{\mathbf{z}} \|\Phi(\mathbf{z}) - \Phi(\mathbf{x})\|^2 \quad \text{such that} \quad \Phi(\mathbf{z}) \in \mathcal{PS}.$$

The sensitivity of this type of Kernel PCA to outlying data is due to the use of squared error loss function. Thus, by modifying the loss function, we hope to obtain a more robust procedure. Bounded loss functions or an absolute error loss function can be considered. Based on this idea, several different loss functions have been proposed, including those found in Nguyen and De la Torre,[43] Alzate and Suykens,[44,45] Kwok and Tsang,[46] and Mika et al.[47] A numerical comparison between some robust Kernel PCA variants in this class is carried out in Nguyen and De la Torre.[43]

A second type of robustness is the sensitivity to the presence of noise in the input data. Recall that, each point $\mathbf{x}_i \in \mathbb{R}^D$ is related to a latent point $\mathbf{y}_i \in \mathbb{R}^d$ through an equation $\mathbf{x}_i = g(\mathbf{y}_i) + \epsilon_i$, where $g(\mathbf{y}_i) \in \mathbb{M}$, and $\epsilon_i$ is some small error vector in $\mathbb{R}^D$. We now consider the case where the dispersion of $\epsilon_1, \ldots, \epsilon_n$ is significantly different from 0, such that the data do not lie exactly on the manifold. Early research traces back to Balasubramanian and Schwartz,[48] which discusses the instability issue in the performance of ISOMAP. In Ref 48 they demonstrate via numerical examples that the ISOMAP algorithm is topologically unstable, i.e., a small amount of noise in the input data can lead to large errors in the solution. In Gerber et al.,[49] a similar phenomenon for the Laplacian Eigenmaps algorithm is illustrated. The output of LEM is constructed by using the orthogonal eigenvectors of the graph Laplacian. However, it is pointed out that the eigenvectors can be strongly correlated locally. Therefore, some features of the input data cannot be correctly recovered by LEM. The effects of noise for other nonlinear methods have been discussed in Goldberg et al.[50]

## CONCLUSION

The papers that we have reviewed have made substantial progress on the problem of robust dimension reduction. However, much work needs to be done. We believe the topic deserves greater attention in the literature than it has been given so far. In particular, much of the literature has addressed the problem of robustness against outliers. The problems of robustness against many other departures from model assumptions remain relatively unexplored.

## REFERENCES


1. Hastie T, Stuetzle W. Principal curves. *J Am Stat Assoc* 1989, 84:502–516.
2. Schölkopf B, Smola A, Müller KR. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Comput* 1998, 10:1299–1319.
3. Roweis ST, Saul LK. Nonlinear dimensionality reduction by locally linear embedding. *Science* 2000, 290:2323–2326.
4. Tenenbaum JB, Silva V, Langford JC. A global geometric framework for nonlinear dimensionality reduction. *Science* 2000, 290:2201–2372.
5. Belkin M, Niyogi P. Laplacian Eigenmaps for dimensionality reduction and data representation. *Neural Comput* 2003, 15:1373–1396.
6. Donoho D, Grimes C. Hessian Eigenmaps: new tools for nonlinear dimensionality reduction. *Proc Natl Acad Sci U S A* 2003, 100:5591–5596.
7. Zhang Z, Zha H. Principal manifolds and nonlinear dimension reduction via local tangent space alignment. *SIAM J Sci Comput* 2005, 26:313–338.
8. Chen L, Buja A. Local multidimensional scaling for nonlinear dimension reduction, graph drawing, and proximity analysis. *J Am Stat Assoc* 2009, 104: 209–219.
9. Hinton GE, Roweis S. Stochastic neighbor embedding. *Adv Neural Inf Process Syst* 2002, 15:833–840.
10. van der Maaten L, Hinton G. Visualizing data using *t*-SNE. *J Mach Learn Res* 2008, 9:2579–2605.

11. Coifman RR, Lafon S. Diffusion maps. *Appl Comput Harmon Anal* 2006, 21:5–30.
12. Brand M. Charting a manifold. *Adv Neural Inf Process Syst* 2002, 15:985–992.
13. Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. *Science* 2006, 313:504–507.
14. Hou C, Wang J, Wu Y, Yi D. Local linear transformation embedding. *Neurocomputing* 2009, 72:2368–2378.
15. Chang H, Yeung DY. Robust locally linear embedding. *Pattern Recognit* 2006, 39:1053–1065.
16. Zhang Z, Wang J. MLLE: modified locally linear embedding using multiple weights. *Adv Neural Inf Process Syst* 2007, 19:1593–1600.
17. Maronna RA. Robust M-estimators of multivariate location and scatter. *Ann Stat* 1976, 4:51–67.
18. Campbell NA. Robust procedures in multivariate analysis. I: robust covariance estimation. *J R Stat Soc C* 1980, 29:231–237.
19. Devlin SJ, Gnanadesikan R, Kettenring JR. Robust estimation of dispersion matrices and principal components. *J Am Stat Assoc* 1981, 76:354–362.
20. Croux C, Haesbroeck G. Principle components analysis based on robust estimators of the covariance or correlation matrix: influence functions and efficiencies. *Biometrika* 2000, 87:603–618.
21. Rousseeuw PJ. Least median of squares regression. *J Am Stat Assoc* 1984, 79:871–880.
22. Rousseeuw PJ. Multivariate estimation with high breakdown point. *Math Stat Appl* 1985, 8:283–297.
23. Davies L. Asymptotic behavior of S-estimators of multivariate location and dispersion matrices. *Ann Stat* 1987, 15:1269–1292.
24. Rousseeuw PJ, Leroy AM. *Robust Regression and Outlier Detection*. New York: John Wiley & Sons; 1987.
25. Hubert M, Rousseeuw PJ, Branden KV. ROBPCA: a new approach to robust principal component analysis. *Technometrics* 2005, 47:64–79.
26. Li G, Chen Z. Projection-pursuit approach to robust dispersion matrices and principal components: primary theory and Monte Carlo. *J Am Stat Assoc* 1985, 80:759–766.
27. Croux C, Ruiz-Gazen A. A fast algorithm for robust principal components based on projection pursuit. In: *COMPSTAT 1996, Proceedings in Computational Statistics*. Heidelberg: Physica-Verlag; 1996, 211–217.
28. Hubert M, Rousseeuw PJ, Verboven S. A fast method for robust principal components with applications to chemometrics. *Chemometrics Intell Lab Syst* 2002, 60:101–111.
29. Maronna RA. Principal components and orthogonal regression based on robust scales. *Technometrics* 2005, 47:264–273.
30. Candès EJ, Li X, Ma Y, Wright J. Robust principal component analysis? *J ACM* 2011, 58:Article 11. doi: 10.1145/1970392.1970395.
31. Chandrasekaran V, Sanghavi S, Parrilo P, Willsky A. Rank-sparsity incoherence for matrix decomposition. *SIAM J Optim* 2011, 21:572–596.
32. Candès EJ, Recht B. Exact matrix completion via convex optimization. *Found Comput Math* 2009, 9:717–772.
33. Schölkopf B, Smola AJ, Müller KR. Kernel principal component analysis. In: *Artificial Neural Networks-ICANN'97*. Berlin and Heidelberg: Springer; 1997, 583–588.
34. Huang SY, Yeh YR, Eguchi S. Robust kernel principal component analysis. *Neural Comput* 2009, 21:3179–3213.
35. Debruyne M, Hubert M, Van Horebeek J. Detecting influential observations in kernel PCA. *Comput Stat Data Anal* 2010, 54:3007–3019.
36. Debruyne M, Verdonck T. Robust kernel principal component analysis and classification. *Adv Data Anal Classif* 2010, 4:151–167.
37. Locantore N, Marron JS, Simpson DG, Zhang JT, Cohen KL. Robust principal component analysis for functional data. *Test* 1999, 8:1–73.
38. Croux C, Haesbroeck G. High breakdown estimators for principal components: the projection-pursuit approach revisited. *J Multivariate Anal* 2005, 95:206–226.
39. Wang L, Pang YW, Shen DY, Yu NH. An iterative algorithm for robust kernel principal component analysis. *Int Conf Mach Learn Cybern* 2007, 6:3484–3489.
40. Deng X, Yuan M, Sudjianto A. A note on robust kernel principal component analysis. *Contemp Math* 2007, 443:21–34.
41. Pang YW, Wang L, Yuan Y. Generalized KPCA by adaptive rules in feature space. *Int J Comput Math* 2010, 87:956–968.
42. Huang HH, Yeh YR. An iterative algorithm for robust kernel principal component analysis. *Neurocomputing* 2011, 74:3921–3930.
43. Nguyen MH, De la Torre F. Robust kernel principal component analysis. *Adv Neural Inf Process Syst* 2008, 1185–1192.
44. Alzate C, Suykens JA. Robust kernel principal component analysis using Huber's loss function. In: *Proceedings of the 24th Benelux Meeting on Systems and Control*, Houffalize, Belgium, 2005.
45. Alzate C, Suykens JA. The kernel component analysis using an epsilon-insensitive robust loss function. *IEEE Trans Neural Netw* 2008, 19:1583–1598.
46. Kwok JTY, Tsang IWH. The pre-image problem in kernel methods. *IEEE Trans Neural Netw* 2004, 15:1517–1525.

47. Mika S, Schölkopf B, Smola AJ, Müller KR, Scholz M, Rätsch G. Kernel PCA and De-noising in feature spaces. *Adv Neural Inf Process Syst* 1998, 11:536–542.

48. Balasubramanian M, Schwartz EL. The Isomap algorithm and topological stability. *Science* 2002, 295:7.

49. Gerber S, Tasdizen T, Whitaker R. Robust non-linear dimensionality reduction using successive 1-dimensional Laplacian Eigenmaps. In: *Proceedings of the 24th International Conference on Machine Learning*, Corvallis, OR, 2007, 281–288.

50. Goldberg Y, Zakai A, Kushnir D, Ritov YA. Manifold learning: the price of normalization. *J Mach Learn Res* 2008, 9:1909–1939.