Knowl Inf Syst (2008) 17:217–240
DOI 10.1007/s10115-008-0123-9

REGULAR PAPER

Determining error bounds for spectral filtering based reconstruction methods in privacy preserving data mining

Songtao Guo · Xintao Wu · Yingjiu Li

Received: 28 November 2006 / Revised: 15 November 2007 / Accepted: 22 December 2007 /
Published online: 22 February 2008

Abstract Additive randomization has been a primary tool for hiding sensitive private information. Previous work empirically showed that individual data values can be approximately reconstructed from the perturbed values, using spectral filtering techniques. This poses a serious threat of privacy breaches. In this paper we conduct a theoretical study on how the reconstruction error varies, for different types of additive noise. In particular, we first derive an upper bound for the reconstruction error using matrix perturbation theory. Attackers who use spectral filtering techniques to estimate the true data values may leverage this bound to determine how close their estimates are to the original data. We then derive a lower bound for the reconstruction error, which can help data owners decide how much noise should be added to satisfy a given threshold of the tolerated privacy breach.

Keywords Privacy preserving · Spectral filtering · Disclosure analysis · Error bound analysis

1 Introduction

Randomization has been a primary tool to hide sensitive private information. The random perturbation techniques aim to distort the sensitive individual values while allowing estimation of the underlying distribution parameters. For all randomization based approaches, there are two fundamentally conflicting requirements: privacy for the individual data values and utility of the perturbed data values.

Consider a data set U with m records of n attributes and a noise data set V with same dimensions as U. The random value perturbation techniques generate a perturbed data matrix

S. Guo · X. Wu (✉)
Department of Software and Information Systems, University of North Carolina at Charlotte, Charlotte, NC 28223, USA
e-mail: xwu@uncc.edu

Y. Li
School of Information Systems, Singapore Management University, Singapore 28223, Singapore

$\tilde{U} = U + V$. Let $\hat{U}$ denote an estimate of the original data which users (or attackers) can achieve. To preserve utility, certain aggregate characteristics (i.e., mean and covariance matrices for numerical data, or marginal totals in contingency table for categorical data) of U should remain basically unchanged in the perturbed data $\tilde{U}$ or can be restored from the reconstructed data $\hat{U}$. In other words, distributions of U can be approximately reconstructed from the perturbed data $\tilde{U}$ using distribution reconstruction approaches (e.g., [4,3]) when some a-priori knowledge (e.g., distribution, statistics etc.) about the noise V is available.

To preserve privacy, not only the difference between $\tilde{U}$ and U but also that between $\hat{U}$ and U should be greater than some tolerated threshold. Here we follow the tradition of using the difference as a measure to quantify how much privacy is preserved. A key element in preserving privacy and confidentiality of sensitive data is the ability to evaluate the extent of all potential disclosures for the released data. In other words, we need to answer to what extent the confidential information in the perturbed data can be compromised by attackers. Hence, we should consider not only the perturbed data, $\tilde{U}$, which is released directly, but also the reconstructed data, $\hat{U}$, which attackers may exploit various reconstruction methods to obtain.

The reconstruction methods investigated in Agrawal and Agrawal [3], and Agrawal and Srikant [4] only focused on how to reconstruct the distribution of the original data from the perturbed data but did not consider the issue that attackers may reconstruct the individual data values through various means. The authors, in [17], argued that randomization schemes might not be secure as attackers may apply a random matrix based spectral filtering technique to retrieve original data values from the perturbed data. Recently, Huang et al. in [15], investigated a similar method based on the principal component analysis (PCA), which exploits correlations among attributes to reconstruct original data values. Their results show that accurate individual data values can be estimated from the perturbed data.

The previous work in [15,17,18] exploited spectral properties of the data and showed that the noise may be separated from the perturbed data and, as a result, privacy could be seriously compromised. Although they empirically assessed effects of the perturbation on the accuracy of the reconstructed individual values, one major question is what the explicit form of the relation between the reconstruction error and the noise may exist. In other words, what bounds of the reconstruction error can be achieved by attackers using spectral filtering based techniques.

In this paper we theoretically explore the problem which originates from the usage of additive noise for privacy preservation. We explicitly assess effects of perturbation on the accuracy of the reconstructed values and give an explicit relation on how the reconstruction error $\hat{U} - U$ varies with the additive noise V. We bound the reconstruction error and the perturbations in terms of matrix norms. In particular, we first derive an upper bound for the reconstruction error using the matrix perturbation theory [21]. Attackers who use spectral filtering techniques to estimate true data values may leverage this bound to determine how close their estimates are to the original data. We then derive a lower bound for the reconstruction error, which can help data owners decide how much noise should be added to satisfy a given threshold of tolerated privacy breach. Since the traditional matrix perturbation theory [21] mainly focused on how the eigenvalues and the angle between eigenvectors (or invariance subspaces) of a perturbed matrix $\tilde{A}$ are upper bounded by the perturbation, we cannot borrow their results directly to derive the lower bound. We present a singular value decomposition (SVD) based reconstruction method and derive a lower bound for the reconstruction error. Since the spectral filtering based approach is equivalent to the SVD based approach, as a result the achieved lower bound of SVD based approach can also be considered as the lower bound of the spectral filtering based approach.

Table 1 Summary of symbols

Symbol	Definition
U	The original data set of m records of n variables
V	The noise data set of m records of n variables
$\tilde{U}$	The perturbed data set $\tilde{U} = U + V$
$\hat{U}$	The estimated data set from a given $\tilde{U}$
A	A $n \times n$ covariance matrix of the original data
E	A $n \times n$ covariance matrix of perturbation
$\tilde{A}$	The perturbed $n \times n$ covariance matrix of A
λ_i	The ith eigenvalue
e_i	The ith eigenvector

The rest of this paper is organized as follows. In Sect. 2 we give definitions of matrix norms and introduce some preliminary background on matrix perturbation theory which will be used in our analysis. In Sect. 3 we revisit the spectral filtering reconstruction methods and present measures used in this paper for both privacy and utility. In Sect. 4 we present in detail the formal analysis on the upper bound and the lower bound with different types of additive noise. We present experimental results in Sect. 5. Finally we discuss the related work in Sect. 6 and offer our concluding remarks in Sect. 7.

2 Preliminaries

We use the tilde conventions to denote perturbations and use the hat conventions to denote estimations. A symbol with a tilde (or hat) over it always denotes a perturbed (or estimated) quantity. The original quantity is denoted by the same symbol without a tilde or hat. Specifically, lower-case variables, e.g., x, represent vectors; italic upper-case alphabets, e.g., A, refer to matrices. For instance, $\tilde{A} = A + E$ denotes a perturbation of A. Let $\Lambda(A) = \{\lambda_1, \ldots, \lambda_n\}$ be the eigenvalues of A and let $[e_1, \ldots, e_n]$ be their corresponding eigenvectors, where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. Similarly, let $\Lambda(\tilde{A}) = \{\tilde{\lambda}_1, \ldots, \tilde{\lambda}_n\}$ and $[\tilde{e}_1, \ldots, \tilde{e}_n]$ be those of $\tilde{A}$, respectively. Table 1 summarizes our notations used in this paper.

Definition 1 Let $A \in \mathcal{R}^{m \times n}$. The Frobenius norm of A is the number

$$\|A\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2}.$$

The 2-norm of A is

$$\|A\|_2 = \underbrace{\max}_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2}$$

where $\|x\|_2$ is for the 2-norm (Euclidean norm) of a vector.

Definition 1 shows the mathematical form of the Frobenius norm and the 2-norm, which will be used in this paper. We will cast much of our analysis in terms of absolute and relative errors of the Frobenius norm, instead of point-wise bounds. The use of absolute and relative

errors in the form of the Frobenius norm gives perturbation bounds a simplicity that makes them easier to interpret. Basically, the Frobenius norm is used to measure the magnitude of data values in total while the 2-norm is used to denote the largest singular value of a matrix.

We list some properties of matrix norms which will be used in our proofs as below. Refer to linear algebra books (e.g., [21]) for more details.

1. $\|AB\|_F \leq \|A\|_F \|B\|_F$ and $\|AB\|_2 \leq \|A\|_2 \|B\|_2$, when $A \in \mathcal{R}^{m\times n}$ and $B \in \mathcal{R}^{n\times q}$.
2. $\|A\|_2 \leq \|A\|_F \leq \sqrt{n}\|A\|_2$.
3. $\|A\|_2 = \sqrt{\lambda_{\max}(A^T A)}$, the square root of the largest eigenvalue of $A^{\mathrm{T}}A$.
4. if A is symmetric, then $\|A\|_2 = \lambda_{\max}(A)$, the largest eigenvalue of A.

3 Spectral analysis of reconstruction methods

3.1 Spectral filtering revisited

Consider a noise matrix V with same dimensions as U. The random value perturbation techniques generate a perturbed data matrix $\tilde{U} = U + V$. The objective of the spectral filtering based approach is to derive the estimation $\hat{U}$ of U from the perturbed data $\tilde{U}$ based on random matrix theory. An explicit filtering procedure is shown below.

1. Calculate the covariance matrix of $\tilde{U}$ by $\tilde{A} = \tilde{U}^T \tilde{U}$.
2. Since the covariance matrix is symmetric and positive semi-definite, we apply spectral decomposition on $\tilde{A}$ to get $\tilde{A} = \tilde{Q}\Lambda\tilde{Q}^{\mathrm{T}}$, where $\tilde{Q}$ is orthogonal matrix whose column vectors are eigenvectors of $\tilde{A}$, and $\tilde{\Lambda}$ is the diagonal matrix with the corresponding eigenvalues on its diagonal.
3. Derive information of the eigenvalues from the covariance matrix of the noise V.
4. Extract the first k components of $\tilde{A}$ as the principal components by comparing $\tilde{\lambda}_i$ with eigenvalues of the noise. $\tilde{\lambda}_1 \geq \tilde{\lambda}_2 \geq \cdots \geq \tilde{\lambda}_k$ are the first k largest eigenvalue of $\tilde{A}$ and $\tilde{e}_1, \tilde{e}_2, \ldots, \tilde{e}_k$ are the corresponding eigenvectors. These eigenvectors form an orthonormal basis of a subspace $\tilde{\chi}$. Let $\tilde{X} = [\tilde{e}_1, \tilde{e}_2, \ldots, \tilde{e}_k]$. The orthogonal projection on to $\tilde{\chi}$ is $P_{\tilde{\chi}} = \tilde{X}\tilde{X}^{\mathrm{T}}$.
5. Obtain the estimated data set using $\hat{U} = \tilde{U}P_{\tilde{\chi}}$.

The authors, in [17], focused on the scenario where only a small number of instances exists in the data set. Furthermore, the noise matrix V considered in [17] is generated using one i.i.d. Gaussian distribution with zero mean and known variance. Based on the random matrix theory, we can derive the theoretical bounds of the eigenvalues corresponding to the noise matrix V as $\lambda_{V_{\min}} = \sigma^2(1 - 1/\sqrt{Q})^2$ and $\lambda_{V_{\max}} = \sigma^2(1 + 1/\sqrt{Q})^2$, where Q is linear to the ratio between the number of records and the number of attributes. As in most data mining applications, the number of records far exceeds that of attributes (hence Q is large), we can see $\lambda_{V_{\min}} \approx \lambda_{V_{\max}} \approx \sigma^2 = \lambda_V$. In this paper, we focus on scenarios with a large number of instances in data sets.

3.2 Other methods

Several similar reconstruction methods have also been investigated. For example, a PCA based reconstruction method was investigated in [15] and a SVD based one was investigated in [14]. Since the SVD based reconstruction method can help to derive the lower bound of reconstruction error using the well-known Mirsky Theorem [21], we briefly present the SVD

method and show the equivalence relationship between the spectral filtering and SVD based method.

Singular Value Decomposition decomposes a matrix $U \in \mathcal{R}^{m\times n}$ (say $m \geq n$) into the product of two unitary matrices, $L \in \mathcal{R}^{m\times m}$, $R \in \mathcal{R}^{n\times n}$, and a pseudo-diagonal matrix $D = \text{diag}(d_1, \ldots, d_n) \in \mathcal{R}^{m\times n}$, such that $U = LDR^{\mathrm{T}}$ or $U = \sum_{i=1}^{n} d_i l_i r_i^{\mathrm{T}}$. The diagonal elements d_i of D are referred to as *singular values*, which are, by convention, sorted in descending order: $d_1 \geq d_2 \geq \cdots \geq d_n \geq 0$. The columns l_i and r_i of L and R are, respectively, called the left and right *singular vectors* of U. Similarly let $\tilde{U} = U + V$ be a perturbation of U and let $\tilde{U} = \tilde{L}\tilde{D}\tilde{R}^{\mathrm{T}}$ be a SVD of $\tilde{U}$. The SVD reconstruction method simply reconstructs U approximately as $\hat{U} = \tilde{U}_k = \tilde{L}_k\tilde{D}_k\tilde{R}_k = \sum_{i=1}^{k} \tilde{d}_i\tilde{l}_i\tilde{r}_i^{\mathrm{T}}$ where $\tilde{D}_k$ is the diagonal matrix with k principal singular values of $\tilde{U}$ and $\tilde{L}_k$ ($\tilde{R}_k$) contains the corresponding left (right) singular vectors.

Result 1 The estimated data from the spectral filtering $\hat{U}_{\text{SF}} = \tilde{U}P_{\tilde{\chi}} = \tilde{U}\tilde{Q}_k\tilde{Q}_k^{\mathrm{T}}$ is the same as that from SVD $\hat{U}_{\text{SVD}} = \tilde{L}_k\tilde{D}_k\tilde{R}_k^{\mathrm{T}}$.

Proof We prove these two methods are equivalent. Since $\tilde{R}_k = R\begin{pmatrix} I_k \\ 0 \end{pmatrix}$,

$$\tilde{U}\tilde{R}_k = \tilde{U}\tilde{R}\begin{pmatrix} I_k \\ 0 \end{pmatrix} = (\tilde{L}\tilde{D}\tilde{R}^{\mathrm{T}})\tilde{R}\begin{pmatrix} I_k \\ 0 \end{pmatrix} = \tilde{L}\tilde{D}\begin{pmatrix} I_k \\ 0 \end{pmatrix} = \tilde{L}_k\tilde{D}_k.$$

Since the columns of right singular vectors ($\tilde{R}$) are the eigenvectors of $\tilde{U}^{\mathrm{T}}\tilde{U}$, that is $\tilde{Q} = \tilde{R}$. Then

$$\hat{U}_{\text{SF}} = \tilde{U}\tilde{R}_k\tilde{R}_k^{\mathrm{T}} = \tilde{L}_k\tilde{D}_k\tilde{R}_k^{\mathrm{T}} = \hat{U}_{\text{SVD}}.$$

□

The non-zero singular values for U are precisely the square roots of the non-zero eigenvalues of the positive semi-definite matrix UU^{T}, and these are precisely the square roots of the non-zero eigenvalues of $U^{\mathrm{T}}U$. Furthermore, the columns of L are eigenvectors of UU^{T} and the columns of R are eigenvectors of $U^{\mathrm{T}}U$.

We can observe that all spectral based methods reconstruct the original data by projecting the perturbed data onto the projection subspaces which are determined by the first k eigenvectors for the spectral filtering method or by the first k singular vectors for the SVD method. In Sect. 4, we shall discuss the strategies of determining k.

3.3 Quantification of privacy and utility

All the above spectral filtering based methods aim to reconstruct individual data directly. They are different from those distribution reconstruction methods [4,3]. In [4], the authors use a measure that defines privacy as follows: if the original value can be estimated with c confidence to lie in the interval $[x_\alpha, x_\beta]$, then the interval width $x_\beta - x_\alpha$ defines the amount of privacy at c confidence level. Since spectral based methods can recover individual data, attackers tend to use the reconstructed data value as an estimate of the original one.

Definition 2 The absolute error of $\hat{U}$, which is regarded as an estimate of U, is defined as

$$ae(U, \hat{U}) = \|\hat{U} - U\|_F.$$

If $\|U\|_F \neq 0$, then the relative error of $\hat{U}$ is defined as

$$re(U, \hat{U}) = \frac{\|\hat{U} - U\|_F}{\|U\|_F}.$$

The use of absolute and relative errors in terms of the Frobenius norm gives the evaluation of the perturbation a simplicity that makes them easier to interpret. In the remainder of this paper, we shall cast much of our bound analysis in terms of Frobenius norm with measures defined in Definition 2 and briefly discuss how to derive a probabilistic upper error bound for a tuple (row) of the data set in Sect. 4.1.

The utility of the data, at the end of the privacy preserving process, is another important issue. The measure used to evaluate the utility usually depends on the specific data mining techniques with respect to which a privacy algorithm is performed. For example, for a classification problem, the authors in [4] measure the inaccuracy in distribution reconstruction by examining the effects on the misclassification rate. In this paper, we apply the universal information loss defined in [3] as the metric for utility loss since the specific data mining task may be unknown. We know the more the noises are made to the data, the less the data reflects the domain of interest. Therefore, an evaluation parameter for the data utility can be the amount of information that is lost after the application of privacy preserving process.

Given the perturbed data, it is (in general) not possible to reconstruct the original density function $f_U(x)$ with an arbitrary precision. The greater the variance of the perturbation, the lower the precision in estimating $f_U(x)$. The universal information loss $\mathcal{I}(f_U, \hat{f}_U)$ denotes the lack of precision in estimating $f_U(x)$ in terms of distribution.

$$\mathcal{I}(f_U, \hat{f}_U) = \frac{1}{2} E\left[\int_{\Omega_X} |f_U(x) - \hat{f}_U(x)| \, \mathrm{d}x\right]. \tag{1}$$

Note that the applied metric is universal in the sense that it can be applied to any reconstruction algorithm since it depends only on the original density $f_U(x)$, and its estimate $\hat{f}_U(x)$. The information loss $\mathcal{I}(f_U, \hat{f}_U)$ lies between zero and one with zero as perfect reconstruction while with one as no overlap between the original density distribution and the reconstructed one. As spectral based methods have reconstructed individual data, we can estimate the density distributions f_U and $\hat{f}_U$ by using multi-dimensional histogram on the original U and the reconstructed $\hat{U}$.

4 Bound analysis

As the previous work in [15,17] only empirically assesses the effects of perturbation on the accuracy of the estimated individual value, in this section, we explore the explicit relation between $\hat{U} - U$ and the noise V and give the upper and lower bounds of $\|\hat{U} - U\|_F$ in terms of $\|V\|_F$.

4.1 Upper bound analysis

The traditional matrix perturbation theory [21] focuses on how the perturbation E affects the matrix A. Specifically, it provides precise upper bounds on the eigenvalues, the angle between eigenvectors, or invariance subspaces of a matrix A and that of its perturbation $\tilde{A} = A + E$, in terms of the norms of the perturbation matrix E. In our scenario, A is the derived covariance

matrix of the original data U while E is the derived perturbation on A caused by V. Hence, it is more significant to consider how the primary perturbation V affects the data matrix U rather than how the derived perturbation E affects the covariance matrix A.

Since the difference between the estimated data and the original one is determined by the invariant subspaces P_χ ($P_{\tilde{\chi}}$) of A ($\tilde{A}$), we first need to assess the bias between these subspaces.

Proposition 1 *Let $A \in \mathcal{R}^{n\times n}$ be a symmetric positive definite matrix, and let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ be its eigenvalues and $e_1, e_2, \ldots, e_n$ be corresponding n eigenvectors. Let the matrix $X \in \mathcal{R}^{n\times(n-k)}$ be defined according to $X = [e_1 e_2 \cdots e_k]$ and $Y = [e_{k+1} \cdots e_n]$, so that the matrix $[XY] \in \mathcal{R}^{n\times n}$ is orthogonal and unitary. Given a perturbation E, let $\tilde{A} = A + E$ and $\epsilon = ||E||_F$. Let χ and $\tilde{\chi}$ be the invariant subspace of A and $\tilde{A}$, respectively. χ is spanned by X. P_χ and $P_{\tilde{\chi}}$ are the corresponding orthogonal projection onto these invariant subspaces. Define eigengap $\delta = \lambda_k - \lambda_{k+1}$. There exists a matrix P satisfying*

$$||P||_F \leq \frac{\sqrt{2}\epsilon}{\delta - \sqrt{2}\epsilon}$$

so that the columns of $\tilde{X} = (X + YP)$ form an orthonormal basis for the subspace spanned by the first k eigenvectors of $\tilde{A}$

$$||P_{\tilde{\chi}} - P_\chi||_F \leq \frac{2\epsilon}{\delta - \sqrt{2}\epsilon}. \tag{2}$$

Proof See Appendix. □

The difference between the invariant subspace of the original data and that of the perturbed data shown in Eq. 2 depends on the eigengap $\delta = \lambda_k - \lambda_{k+1}$ which is determined by the spectrum of the original data. Note that the spectrum of the original data is unknown to attackers. In the following, we show how to estimate this eigengap using the spectrums of the perturbed data and the noise.

Proposition 2 *Given a symmetric matrix $A \in \mathcal{R}^{n\times n}$ and a symmetric perturbation E, let $\tilde{A} = A+E$. Let the eigenvalues of E be $\varepsilon_1 \geq \varepsilon_2 \geq \cdots \geq \varepsilon_n$. Let λ_k and $\tilde{\lambda}_k$ are the eigenvalues of A and $\tilde{A}$, respectively, where $k = 1, \ldots, n$, and let $\tilde{\delta} = \tilde{\lambda}_k - \tilde{\lambda}_{k+1}$, $\delta_E = \varepsilon_1 - \varepsilon_n$, then*

$$\delta \in [\tilde{\delta} - \delta_E, \tilde{\delta} + \delta_E]$$

Proof From Corollary 4.9 in [21], we have:

$$\begin{aligned} \lambda_k &\in [\tilde{\lambda}_k - \varepsilon_1, \tilde{\lambda}_k - \varepsilon_n] \\ \lambda_{k+1} &\in [\tilde{\lambda}_{k+1} - \varepsilon_1, \tilde{\lambda}_{k+1} - \varepsilon_n]. \end{aligned} \tag{3}$$

Since $\delta = \lambda_k - \lambda_{k+1}$,

$$\begin{aligned} \delta &\geq (\tilde{\lambda}_k - \varepsilon_1) - (\tilde{\lambda}_{k+1} - \varepsilon_n) = (\tilde{\lambda}_k - \tilde{\lambda}_{k+1}) - (\varepsilon_1 - \varepsilon_n) = \tilde{\delta} - \delta_E \\ \delta &\leq (\tilde{\lambda}_k - \varepsilon_n) - (\tilde{\lambda}_{k+1} - \varepsilon_1) = (\tilde{\lambda}_k - \tilde{\lambda}_{k+1}) + (\varepsilon_1 - \varepsilon_n) = \tilde{\delta} + \delta_E \end{aligned}$$

□

Result 2 (Upper bound) Given a data set $U \in \mathcal{R}^{m\times n}$ and a perturbation noise set $V \in \mathcal{R}^{m\times n}$, let $\tilde{U} = U + V$ and $\hat{U}$ to denote the estimate obtained from the spectral filtering technique. We have

$$\|\hat{U} - U\|_F \leq \|\tilde{U}\|_F \frac{2||E||_F}{(\tilde{\delta} - \delta_E) - \sqrt{2}||E||_F} + \|V P_\chi\|_F \tag{4}$$

where $E = V^{\mathrm{T}}U + U^{\mathrm{T}}V + V^{\mathrm{T}}V$ is the derived perturbation on covariance matrix $A = U^{\mathrm{T}}U$.

Proof Consider the covariance matrix of $\tilde{U}$:

$$\tilde{A} = \tilde{U}^{\mathrm{T}}\tilde{U} = (U+V)^{\mathrm{T}}(U+V) = U^{\mathrm{T}}U + V^{\mathrm{T}}U + U^{\mathrm{T}}V + V^{\mathrm{T}}V$$

We denote $\tilde{A}$ as $A + E$ where $E = V^{\mathrm{T}}U + U^{\mathrm{T}}V + V^{\mathrm{T}}V$. Since

$$\hat{U} - U \approx \tilde{U}P_{\tilde{\chi}} - UP_{\chi} = \tilde{U}P_{\tilde{\chi}} - (\tilde{U} - V)P_{\chi} = \tilde{U}(P_{\tilde{\chi}} - P_{\chi}) + VP_{\chi} \tag{5}$$

hence we have,

$$\begin{aligned}
\|\hat{U} - U\|_F &\approx \|\tilde{U}(P_{\tilde{\chi}} - P_{\chi}) + VP_{\chi}\|_F \\
&\leq \|\tilde{U}(P_{\tilde{\chi}} - P_{\chi})\|_F + \|VP_{\chi}\|_F \\
&\leq \|\tilde{U}\|_F \|(P_{\tilde{\chi}} - P_{\chi})\|_F + \|VP_{\chi}\|_F \\
&\leq \|\tilde{U}\|_F \frac{2\epsilon}{\delta - \sqrt{2}\epsilon} + \|VP_{\chi}\|_F \quad Proposition\ 1 \\
&\leq \|\tilde{U}\|_F \frac{2||E||_F}{(\tilde{\delta} - \delta_E) - \sqrt{2}||E||_F} + \|VP_{\chi}\|_F \quad Proposition\ 2
\end{aligned}$$

□

The upper bound given in Result 2 determines how close the estimated data achieved by attackers is to the original one when the spectral filtering technique is exploited. This represents a serious threat of privacy breaches as attackers know exactly how close their estimates are.

From Eq. 4, we can observe that both the eigen gap $\tilde{\delta}$ and the projection space P_{χ} depend on the determination of the number of principal components k. The original spectral filtering algorithm [17] suggested the following strategy to determine the first k eigen components.

Strategy 1 $k = \max\{i | \tilde{\lambda}_i \geq \lambda_V\}$ where λ_V denotes the largest eigenvalue of noise ϵ_1. □

Strategy 1 aims to include all significant eigen components (with $\lambda_i > 0$) in the projection space for reconstruction. From Eq. 3, we can derive $\tilde{\lambda}_i \leq \lambda_i + \epsilon_1$. So if $\tilde{\lambda}_i \leq \epsilon_1$, λ_i might not be considered as a principal component.

The noise considered in the additive perturbation can be either independent or correlated with the original data. In the following, we show the upper bound of the reconstruction error for different cases.

Corollary 1 *When the noise is completely correlated with the original data, the upper bound of the reconstruction error can be expressed as*

$$\|\hat{U} - U\|_F \leq \|\tilde{U}\|_F \frac{2||E||_F}{(\tilde{\delta} - \delta_E) - \sqrt{2}||E||_F} + \|V\|_F. \tag{6}$$

When the noise is independent with the original data, the upper bound of the reconstruction error can be expressed as

$$\|\hat{U} - U\|_F \leq \|\tilde{U}\|_F \frac{2||V||_F^2}{(\tilde{\delta} - \delta_E) - \sqrt{2}||V||_F^2} + \|VP_{\chi}\|_F. \tag{7}$$

Proof When the noise is completely correlated with the original data, we have $\|VP_\chi\|_F \approx \|V\|_F$ as k represents the number of principal components. Then Eq. 4 becomes Eq. 6.

When the data and noise are uncorrelated, we have $V^TU = U^TV = 0$. Hence $E = V^TU + U^TV + V^TV$ can be simplified as $E = V^TV$. In terms of Frobenius norm, we have $\|E\|_F = \|V^TV\|_F \leq \|V\|_F^2$. By replacing $\|E\|_F$ with $\|V\|_F^2$ in Eq. 4, we have Eq. 7. □

In [17], the noise is assumed as following one i.i.d. Gaussian distribution $N(0, \Sigma)$, where the covariance matrix $\Sigma = \text{diag}(\sigma^2, \ldots, \sigma^2)$. This represents the scenario where the noise is completely independent with original data. One example of this scenario is the online collection of customer's individual data (as other customers' data is unknown during the data collection). For i.i.d. noise, we have the following result.

Corollary 2 *When the noise is generated by an i.i.d. Gaussian distribution with zero mean and known variance σ^2, the upper bound of the reconstruction error can be expressed as*

$$\|\hat{U} - U\|_F \leq \|\tilde{U}\|_F \frac{2||V||_F^2}{\tilde{\delta} - \sqrt{2}||V||_F^2} + \sqrt{k/n}||V||_F \tag{8}$$

where $||V||_F = \sqrt{\sigma^2 mn}$.

Proof When the noise matrix is generated by an i.i.d. Gaussian distribution with zero mean and known variance σ^2, the square error of VP_χ is $\delta^2 = \sigma^2\frac{k}{n}$ [15] and $||V||_F$ is $\sqrt{\sigma^2 mn}$. We have

$$\|VP_\chi\|_F = \sqrt{\delta^2 mn} = \sqrt{\sigma^2 \frac{k}{n} mn} = \sqrt{\frac{k}{n}}\,||V||_F.$$

Furthermore, when the noise is generated by an i.i.d. Gaussian distribution, $\varepsilon_1 \approx \varepsilon_n$. In other words, $\delta_E = \varepsilon_1 - \varepsilon_n$ is close to zero. By replacing $\|VP_\chi\|_F$ and δ_E, Eq. 7 becomes Eq. 8. □

Upper bounds given in Eqs. 4, 7, and 8 interpret privacy loss at the aggregate level. Attackers may be interested in exploring the error bound for each individual tuple. From Eq. 5, we can have

$$\hat{u}_i - u_i \approx \tilde{u}_i(P_{\tilde{\chi}} - P_\chi) + v_i P_\chi$$

where v_i denotes the added noise on the u_i. By incorporating Eq. 2, we have

$$||\hat{u}_i - u_i||_F \leq \tilde{u}_i \frac{2\epsilon}{\delta - \sqrt{2\epsilon}} + \|v_i P_\chi\|_F. \tag{9}$$

Similarly we can derive the error bounds at the tuple level for the completely correlated noise and the i.i.d. Gaussian distributed noise, respectively. However, one problem here is that v_i in Eq. 9 is not available to attackers. The Frobenius norm of this vector, as a function with n variables, represents the vector length in the Hilbert space. Therefore, the distribution of such function may be derived or approximated from the distribution of the noise. As a result, a probabilistic bound of $||v_i||_F$ based on its corresponding distribution can be obtained by attackers. By replacing the $||v_i||_F$ with the upper probabilistic bound, we can derive the probabilistic error bound for an individual tuple.

Strategy 1 can be generally used to determine the number of eigen components in the projection space. However, it cannot guarantee to achieve an optimal result. The reason is that it aims to include all significant eigen components (with $\lambda_i > 0$) in the projection

space for reconstruction. However, since the inclusion of one eigen component also brings additional noise projected on that eigenvector, the benefit due to inclusion of one insignificant eigen component may be diminished by the side effect due to the additional noise projected on this eigenvector.

For i.i.d noise, since the effect of the projection on any vector is the same, we can propose a better strategy (as shown in Strategy 2) which can achieve an optimal estimation. Strategy 2 only includes the ith eigen component when the benefit due to inclusion of the ith component is greater than the loss due to the noise projected on the ith component, i.e., $\lambda_i \geq \lambda_V$. When the noise is independent to the data and also has no correlation among the noise, we have $\tilde{\lambda}_i = \lambda_i + \lambda_V \geq 2\lambda_V$.

Strategy 2 For i.i.d. noise, the estimated data $\hat{U} = \tilde{U} P_{\tilde{\chi}} = \tilde{U} \tilde{Q}_k \tilde{Q}_k^{\mathrm{T}}$ is approximately optimal by using $k = \max\{i | \tilde{\lambda}_i \geq 2\lambda_V\}$.

Proof See Appendix for proof details. □

4.2 Lower bound analysis

In Sect. 3.2, we have presented the SVD based reconstruction method and shown the equivalence between the SVD based method and the spectral filtering method. Hence the derived lower bound from SVD based method can also be considered as the lower bound of the spectral filtering method. Recall that the SVD based reconstruction method simply estimates U as $\hat{U} = \tilde{U}_k = \tilde{L}_k \tilde{D}_k \tilde{R}_k = \sum_{i=1}^{k} \tilde{d}_i \tilde{l}_i \tilde{r}_i^{\mathrm{T}}$ where $\tilde{D}_k$ is the diagonal matrix with k principal singular values of $\tilde{U}$ and $\tilde{L}_k$ ($\tilde{R}_k$) contains the corresponding left (right) singular vectors. In this section, we derive a lower bound using the well-known Mirsky Theorem for SVD decomposition [21].

Result 3 (Lower bound) Consider $\hat{U} = \tilde{U}_k = \tilde{L}_k \tilde{D}_k \tilde{R}_k^{\mathrm{T}}$ as the reconstruction of the original data set U. The reconstruction error between $\hat{U}$ and U has its lower bound:

$$||\hat{U} - U||_F \geq ||U_k - U||_F$$

where $U_k = L_k D_k R_k$.

Proof $\tilde{U}$ and U are matrices of the same dimensions with singular values

$$\tilde{d}_1 \geq \tilde{d}_2 \geq \cdots \geq \tilde{d}_n$$
$$d_1 \geq d_2 \geq \cdots \geq d_n.$$

Since $\tilde{U}_k = \tilde{L}_k \tilde{D}_k \tilde{R}_k^T$, we set

$$\tilde{d}_{k+1} = \cdots = \tilde{d}_n = 0.$$

By Mirsky's theorem [21]

$$||\hat{U} - U||_F^2 \geq \sum_{i=1}^{n} |\tilde{d}_i - d_i|^2 \geq d_{k+1}^2 + \cdots + d_n^2 = ||U_k - U||_F^2.$$

The relationship between the reconstruction error and perturbation (especially the lower bound) will, in turn, guide us to add noise into the original data set. The lower bound gives data owners the worst case security assurance since it is bounded by any matrix B of rank no greater than k derived by attackers. In order to preserve privacy, data owners need to

make sure $||\hat{U} - U||_F/||U||_F$ is greater than the privacy threshold τ specified before the perturbation.

Based on the derived lower bound,

$$\tau||U||_F \leq ||U_k - U||_F = d_{k+1}^2 + \cdots + d_n^2.$$

Hence k which might be chosen by attackers can be determined by

$$k = \max\{i|\tau \leq (d_{i+1}^2 + \cdots + d_n^2)/||U||_F\}. \tag{10}$$

For i.i.d. noise, based on strategy 2, $\lambda_i \geq \lambda_V$, the data owner should generate V such that the eigenvalue of $(V^{\mathrm{T}}V)$ satisfies

$$\lambda_{k+1} < \lambda_V \leq \lambda_k. \tag{11}$$

Since λ_V is the eigenvalue of $V^{\mathrm{T}}V$, the variance of the noise can be derived $\sigma^2 = \lambda_V/(m-1)$, where m is the number of rows in V.

5 Experimental results

5.1 Experiment setting

In our experiment, we use two data sets. The first one is an artificial dataset, as specified similarly in [17]. We increase the size of instances from 300 to 30,000 since we focus on the scenario with a large number of instances in data sets. To better compare the difference of reconstruction error between Strategy 1 and 2, we add two more features which are independent with the previous four features. Specifically, U is a highly correlated data set with 35 variables which are generated from 6 independent features. Each feature has a specific trend like sinusoidal, square, or triangular shape and there is no dependency between any two features. The second one is the numerical part of the Adult data set [5], with 6 attributes and 32,561 instances. To better illustrate the performance of the reconstruction method with different strategies and different types of noises, we normalize each attribute to its 0, 1 domain range. In this paper, we consider three different types of additive noise.

- Type 1. V is an additive noise following one i.i.d. Gaussian distribution $N(0, \Sigma)$, where the covariance matrix $\Sigma = \mathrm{diag}(\sigma^2, \sigma^2, \ldots, \sigma^2)$ (The same as in [17]).
- Type 2. V is an additive noise following one Gaussian distribution $N(0, \Sigma)$, where the covariance matrix $\Sigma = c \times \mathrm{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2)$. Here each feature is applied with a separate Gaussian distribution with its variance linear with the variance of the original data.
- Type 3. V is an additive noise following Gaussian distribution $N(0, \Sigma)$, where the covariance matrix $\Sigma = c \times A$. A is the covariance matrix of the original data set. Here the covariance matrix of noise is linear with that of the original data.

Type 1 represents the scenario where the noise is completely independent with the original data. Type 2 represents the scenario where the variance of the original data is a-priori known while type 3 represents the scenario where the whole covariance matrix of the original data is used to generate noise. Note that in all the above three scenarios, we assume that the noise is generated with a Gaussian distribution and its associated mean vector is zero. This assumption is generally true in privacy preserving data mining applications as the change of the mean values will significantly affect the accuracy of data mining results.

In our following experiments, we perturb the original data by different levels of noise, which are generated by varying the covariance matrix Σ. For each level, we keep the same noise-to-signal ratio ($\|V\|_F/\|U\|_F$). For type 1, based on $||V||_F \approx \sqrt{\sigma^2 mn}$, we can derive $\sigma^2 = \|V\|_F/\sqrt{mn}$. While for both type 2 and 3, since $||V||_F \approx \sqrt{c(\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2)m}$, we can derive

$$c \approx \frac{||V||_F^2}{(\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2)m}. \tag{12}$$

For each perturbed data, we use our spectral filtering technique with Strategy 1 and 2 to reconstruct the point-wise data, respectively. We also show how the reconstruction error is affected by varying k, the number of eigen components included in the projection space. We use the relative error $re(U, \hat{U}) = \frac{\|\hat{U}-U\|_F}{\|U\|_F}$ to measure the privacy loss and use the universal information difference $\mathcal{I}(f_U, \hat{f}_U)$ (Eq. 1) to measure the utility loss. Since the artificial data set is very sparse (35 features), we cannot accurately derive its density distribution by applying the multi-dimensional histogram technique. Here we only give evaluations on the utility loss for the Adult data set (6 features).

5.2 Artificial data set

Table 2 shows our experimental results on the relative error $re(U, \hat{U})$ with three types of additive noise (type 1, 2, and 3 noises) for the artificial data set.

5.2.1 Effect of varying k

For i.i.d. noise, we have presented two strategies on how to determine k by examining the eigenvalues of the covariance matrix of the perturbed data and the eigenvalues of the covariance matrix of the additive noise. It is easy to see that different k values lead to different reconstruction errors [which are measured by $re(U, \hat{U})$], as shown in each column of Table 2. There are two phases of reconstruction error changes when we increase k. We take column V1 (with the variance of the noise as 0.213 and the relative noise strength as 0.628) as an example. In the first phase (i.e., from $k = 1$ to $k = 6$), the reconstruction error $re(U, \hat{U})$ is decreased from 0.821 to 0.260 when more principal components are included in reconstruction. This is because that the gain of inclusion of significant principal components are greater than the loss due to the inclusion of noise projected on those components. In the second phase (i.e., $k \geq 6$), the reconstruction error $re(U, \hat{U})$ is increased since the gain of inclusion of any additional component (component with small eigenvalue or insignificant component) is diminished by the loss due to the inclusion of noise on that component. When we examine the original data, there exist 6 principal components as the data is highly correlated among 35 features. Since the added noise V1 is relatively small, both strategies incur the same reconstruction error by incorporating all six principal components.

5.2.2 Effect of varying noise

In the next experiment, we vary the variance of the added noise from 0.213 (V1) to 4.814 (V9) as shown in Table 2. We denote the values with $*$ as the results following Strategy 2, while the values with † as the results following Strategy 1. For each noise data set, we also show all the relative reconstruction errors by varying k values. In each column, the value in bold font highlights the best result which could be achieved by comparing k reconstruction errors.

Table 2 The relative error $re(U, \hat{U})$ verses varying V under three three types of additive noise (Type 1, 2, and 3) for the artificial data set

Noise		V1	V2	V3	V4	V5	V6	V7	V8	V9
$\|\|V\|\|_F/\|\|U\|\|_F$		0.628	0.786	0.954	1.178	1.366	1.677	1.944	2.121	2.985
Variance		0.213	0.333	0.491	0.750	1.007	1.524	2.040	2.430	4.814
	$k=1$	0.821	0.825	0.830	0.839	0.847	0.863	0.877	0.890	0.960
	$k=2$	0.649	0.659	0.671	0.692	0.711	0.750	0.783	0.810	***0.956**
Type 1	$k=3$	0.440	0.461	0.488	0.529	0.565	0.636	0.694	***0.739**	0.964
$re(U, \hat{U})$	$k=4$	0.297	0.337	***0.383**	***0.450**	***0.506**	***0.607**	***0.687**	0.748	1.032
	$k=5$	0.271	***0.324**	0.383	0.465	0.532	0.651	0.745	0.816	1.141
	$k=6$	*†**0.260**	†0.325	†0.395	†0.489	†0.567	†0.699	†0.805	†0.883	†1.245
	$k=7$	0.282	0.353	0.428	0.530	0.614	0.757	0.873	0.956	1.348
c		0.402	0.630	0.927	1.415	1.903	2.864	3.850	4.583	9.080
	$k=1$	0.826	0.832	0.841	0.854	0.868	0.897	0.926	0.945	†**1.072**
	$k=2$	0.654	0.667	0.684	0.709	0.748	0.819	0.876	0.911	1.125
Type 2	$k=3$	0.452	0.479	0.513	0.564	0.613	0.697	†**0.778**	†**0.830**	1.120
$re(U, \hat{U})$	$k=4$	0.309	0.353	†**0.405**	†**0.479**	†**0.544**	†**0.652**	0.900	0.967	1.317
	$k=5$	0.279	†**0.345**	0.462	0.552	0.631	0.761	1.008	1.085	1.487
	$k=6$	†**0.255**	0.391	0.512	0.616	0.706	0.856	1.103	1.190	1.634
	$k=7$	0.294	0.431	0.558	0.673	0.774	0.939	1.190	1.286	1.777
c		0.402	0.630	0.927	1.415	1.903	2.864	3.850	4.583	9.080
	$k=1$	0.893	0.935	0.989	**1.067**	**1.140**	**1.276**	†**1.398**	†**1.485**	†**1.926**
	$k=2$	0.800	0.879	0.977	1.117	†1.240	†1.455	1.644	1.769	2.420
Type 3	$k=3$	0.702	†0.824	†0.964	†1.156	1.318	1.593	1.830	1.981	2.779
$re(U, \hat{U})$	$k=4$	†0.650	0.797	0.961	1.177	1.358	1.659	1.918	2.083	2.943
	$k=5$	0.636	0.788	0.956	1.177	1.361	1.667	1.930	2.097	2.968
	$k=6$	**0.627**	**0.783**	**0.955**	1.179	1.366	1.675	1.940	2.109	2.987
	$k=7$	0.627	0.783	0.955	1.179	1.366	1.675	1.940	2.109	2.987

The values with $*$ denote the results following Strategy 2, while the values with † denote the results following Strategy 1. The bold values indicate those best estimates achieved by the spectral filtering technique

From Table 2, we can see that Strategy 2 can achieve optimal results (least reconstruction error) for all perturbations from V1 to V9 while Strategy 1 suffers when relative large perturbations are added. The reason is that Strategy 1 always include all six principal components in the projection space across all nine noise data sets. On the contrary, Strategy 2 compares the magnitude of the principal components with the magnitude of additive noise to determine k. For example, the best k value for noise V4 is four as shown in Table 2. We can observe that the magnitudes of the last two principal components are not as significant as those of noise projected on the corresponding components. Hence, the gain of inclusion of the last two (not very significant) principal components is diminished by the loss due to the inclusion of noise projected on those components.

Quality of the data reconstruction depends upon the relative noise contained in the perturbed data. As the noise added to the actual value increases, the reconstruction error increases. Figure 1 shows point-wise data distributions of reconstruction for feature two (we get a sample of 300 data records) when we vary noise levels. We can see when the noise-to-signal

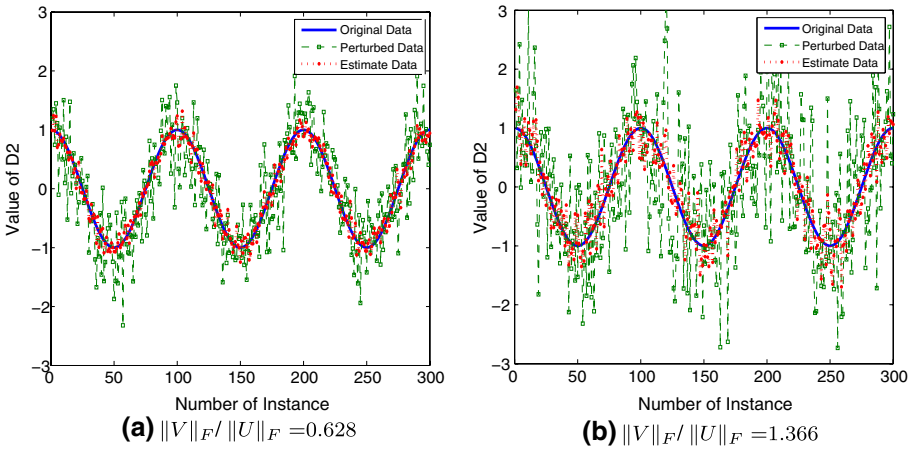

Fig. 1 Reconstruction error (point-wise data distribution for attribute 2) with best k verses varying noise magnitude

ratio $\|V\|_F/\|U\|_F$ is 0.628, (the corresponding variance $\sigma^2 = 0.213$), the spectral filtering can achieve relatively accurate estimates because the effects due to the noise projection on the remaining 29 components are safely filtered. When we increase the noise-to-signal ratio to 1.366 (the corresponding noise variance is $\sigma^2 = 1.007$), the reconstruction error increases as shown in Fig. 1(b). The reasons are twofold. First, much larger noise exists in the projection space. Second, information contained in those principal components which are excluded from the projection space is lost since the large noise tends to affect the determination of k.

5.2.3 Effect of different types of noise

Since Strategy 2 is not designed for scenarios (type 2 and 3 noises) with correlated noises, we only show the reconstruction errors of Strategy 1 in the second and third blocks of Table 2. For type 2, the spectral filtering with Strategy 1 can achieve the optimal estimates. However, it generally cannot achieve good results for type 3 where the covariance matrix of the noise is linear with the covariance matrix of the signal. As the noise is not randomly generated, the spectral filtering technique, which is based on random matrix perturbation, cannot satisfactorily separate noise from data as they share the same distribution pattern.

Figure 2 compares the reconstruction error for one single feature (attribute 2) in three cases. Each case has different type of the noise, however, with the same magnitude ($\|V\|_F/\|U\|_F = 0.628$). We can see the spectral filtering performs best for the completely random perturbation (type 1) while performs worst for the completely correlated perturbation (type 3).

5.3 Adult data set

Table 3 shows our experimental results on the relative error $re(U, \hat{U})$ with three types of additive noise (type 1, 2, and 3) for the Adult data set. The values with $*$ denote the results following Strategy 2, while the values with † denote the results following Strategy 1. The bold values indicate those best estimates achieved by the spectral filtering technique. We have similar observations as those on the previous artificial data set. For example, Strategy 2 can always achieve optimal estimates for the i.i.d. noise (type 1) while Strategy 1 usually

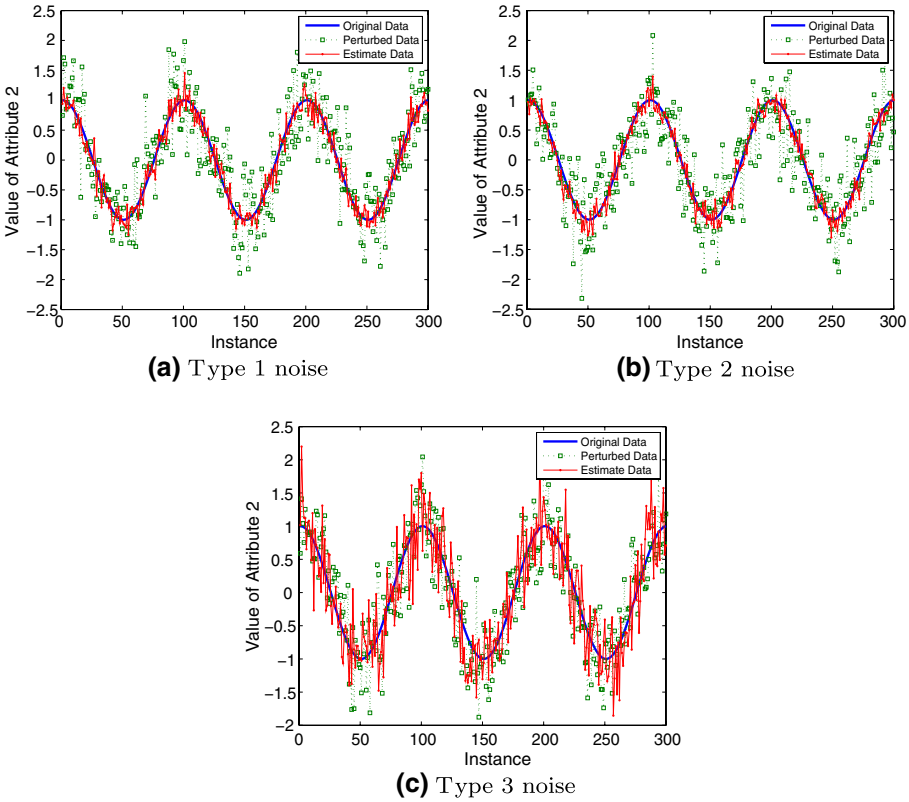

Fig. 2 Reconstruction error (data distribution for attribute 2) with $||V||_F/||U||_F = 0.213$ under three types

incurs more inaccuracies since it tends to include all principal components (6 in this data set) without considering the side effect incurred by the inclusion of noise. We can also observe that the more noise we add, the greater the reconstruction error. This observation is held across all three types of noises.

To measure the utility, we apply the universal information loss $\mathcal{I}(f_U, \hat{f}_U)$ as defined in Eq. 1. Recall that the universal information loss only depends on the original density $f_U(x)$, and its estimate $\hat{f}_U(x)$. To derive the density distribution $f_U(x)$ of the original data and $\hat{f}_U(x)$ of the corresponding reconstructed data, we equally divide each dimension into 5 bins and compare the multidimensional histograms based on the frequency information contained in those 5^6 six-dimensional bins.

Table 4 shows our results on the utility loss of the reconstructed Adult data with different levels of type 1 noise (V1–V9). The values with $*$ denote the results following Strategy 2, while the values with † denote the results following Strategy 1. The bold values indicate those best estimates achieved by the spectral filtering technique. From Table 4, we can observe that Strategy 2 always outperforms Strategy 1 in terms of utility preservation. Note that the information loss $\mathcal{I}(f_U, \hat{f}_U)$ lies between 0 and 1 with 0 as perfect reconstruction while with 1 as no overlap between the original density distribution and the reconstructed one. Another observation is that the utility of reconstructed data decreases for both Strategy 1 and Strategy 2 when the magnitude of added noise increases.

Table 3 The relative error $re(U, \hat{U})$ verses varying V under three types of noises for the Adult data set

Noise		V1	V2	V3	V4	V5	V6	V7	V8	V9
$\|\|V\|\|_F/\|\|U\|\|_F$		0.172	0.176	0.188	0.218	0.231	0.243	0.266	0.297	0.326
Variance		0.005	0.0052	0.006	0.008	0.009	0.01	0.012	0.015	0.018
	$k=1$	0.267	0.268	0.269	0.273	0.275	0.276	0.280	0.285	0.290
	$k=2$	0.212	0.213	0.217	0.226	0.230	0.234	0.243	0.254	*__0.266__
Type 1	$k=3$	0.185	0.186	0.192	0.207	*__0.214__	*__0.221__	*__0.234__	*__0.254__	0.269
$re(U, \hat{U})$	$k=4$	0.176	0.178	*__0.186__	*__0.206__	0.216	0.225	0.241	0.265	0.286
	$k=5$	0.173	*__0.176__	0.186	0.211	0.223	0.233	0.253	0.281	0.306
	$k=6$	*†__0.172__	†0.176	†0.188	†0.218	†0.231	†0.243	†0.266	†0.297	†0.326
c		0.302	0.312	0.362	0.604	1.200	3.028	6.037	12.15	30.26
	$k=1$	0.274	0.275	0.276	0.283	0.286	0.289	0.294	0.303	__0.312__
	$k=2$	0.231	0.233	0.238	0.254	0.260	0.267	0.281	0.299	†0.317
Type 2	$k=3$	0.207	0.210	†0.218	†0.240	†0.249	†0.258	†0.276	†0.300	0.324
$re(U, \hat{U})$	$k=4$	†0.193	†0.196	0.205	0.231	0.241	0.252	0.272	0.299	0.325
	$k=5$	0.182	0.186	0.196	0.224	0.235	0.246	0.269	0.297	0.325
	$k=6$	__0.172__	__0.176__	__0.187__	__0.218__	__0.231__	__0.242__	0.266	__0.297__	0.326
c		0.302	0.312	0.362	0.604	1.200	3.028	6.037	12.15	30.26
	$k=1$	0.276	0.276	0.279	0.286	0.289	0.292	0.298	0.308	__0.317__
	$k=2$	0.233	0.235	0.240	0.256	0.263	0.271	0.284	0.302	†0.321
Type 3	$k=3$	0.208	0.210	†0.218	†0.239	†0.249	†0.259	†0.276	†0.300	0.323
$re(U, \hat{U})$	$k=4$	†0.193	†0.196	0.206	0.230	0.242	0.253	0.272	0.298	0.325
	$k=5$	0.183	0.186	0.196	0.223	0.236	0.248	0.269	0.297	0.325
	$k=6$	__0.172__	__0.176__	__0.188__	__0.217__	__0.231__	__0.243__	__0.266__	__0.296__	0.326

The values with $*$ denote the results following Strategy 2, while the values with † denote the results following Strategy 1. The bold values indicate those best estimates achieved by the spectral filtering technique

Table 4 Utility of reconstructed Adult data with type 1 noise

Noise		V1	V2	V3	V4	V5	V6	V7	V8	V9
$\|\|V\|\|_F/\|\|U\|\|_F$		0.172	0.176	0.188	0.218	0.231	0.243	0.266	0.297	0.326
	$k=1$	__0.137__	__0.137__	0.152	0.309	0.089	__0.137__	0.306	0.152	0.088
	$k=2$	0.292	0.232	0.233	0.078	__0.011__	0.292	0.152	0.297	*0.156
	$k=3$	0.261	0.270	0.084	0.051	*0.221	*0.196	*0.143	*0.289	0.226
Utility loss	$k=4$	0.259	0.213	*0.082	*0.050	0.125	0.170	0.141	0.125	0.213
$\mathcal{I}(f_U, \hat{f}_U)$	$k=5$	0.254	*0.208	__0.077__	__0.045__	0.116	0.164	__0.137__	__0.119__	__0.086__
	$k=6$	*†0.515	†0.501	†0.521	†0.512	†0.499	†0.466	†0.466	†0.462	†0.479

The values with $*$ denote the results following Strategy 2, while the values with † denote the results following Strategy 1. The bold values indicate those best estimates achieved by the spectral filtering technique

To further evaluate how different types of noise (type 1, 2, and 3) affect the utility of the reconstructed data. We show one result on the relationship between the utility verses varying three types of noises in Fig. 3. We can observe that the spectral filtering best preserves the utility with type 1 noise (i.i.d.) while it incurs the largest utility loss with type 3 noise (completely correlated). This is because the completely correlated noise cannot be well

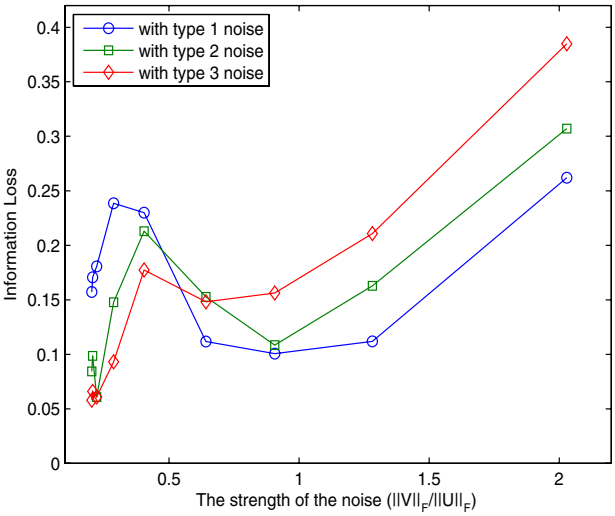

Fig. 3 Utility verses varying magnitudes of three types of additive noise

filtered out by the spectral based reconstruction method although some statistical properties (e.g., the covariance matrix) can be fully preserved.

6 Related work

The field of statistical databases has developed various perturbation based methods to prevent the disclosure of confidential individual data while satisfying requests for aggregate information. The perturbation family includes swapping values between records, replacing the original database by a sample from the same distribution, adding noise to the values in the databases, adding noise to the results of a query and sampling the result of a query [1]. There are various approaches to assess risk of identity disclosure and most of them relate to the inadvertent release of small counts in the full k-way table or data cubes [7,22,23]. We should point out that the privacy consideration in the current literature for statistical database is not enough for many general environments which contain many numerical attributes. In most statistical database literatures, the privacy concerned is about the re-identification of some specific entries in the database.

A considerable amount of work on privacy preserving data mining has been reported in recent years. The random noise addition methods have been well investigated (e.g., [4,3, 10,15,17,18]). The objective of randomization based privacy-preserving data mining is to prevent the disclosure of confidential individual values while preserving general patterns and rules. Agrawal and Srikant [4] proposed the development of data mining techniques that incorporate privacy concerns and this work has been extended in [3]. In [4], the randomization schema is to add a random number v_i which is drawn from some known distribution, to u_i, the value of a sensitive attribute. Then, the randomized value $\tilde{u}_i = u_i + v_i$ is released for data mining. They also show an approach to recovering the distribution of the original data given the distribution of random noises and apply this approach for decision tree learning. Agrawal and Agrawal [3] have provided an expectation-maximization (EM) algorithm for reconstructing the distribution of the original data from perturbed observations. They provide

information theoretic measures to quantify the amount of privacy provided by a randomization approach. Some recent work [3,13,16] has explored whether the reconstructed distribution or data mining results can be exploited by attackers to breach individual privacy.

The authors in [11,14,15,17,18] have investigated point-wise reconstruction methods (spectral filtering, PCA based, and SVD based), which may be exploited by attackers to breach individual privacy. Those techniques investigated how correlations among attributes affect the privacy of a data set disguised via the additive random perturbation scheme. In [15], Huang et.al. also introduced a Bayes approach based on maximum a posteriori (MAP) estimation, which considers both priori and posterior knowledge via Bayes' theorem to estimate original data. However, strong assumptions are imposed on this method such as the original data and the noise are multi-variate normal distributed and both distributions are available to attackers.

In addition to the previous additive noise perturbation approach, the random rotation based perturbation approach ($Y = RX$ where R is an orthogonormal random matrix) has recently been investigated in [6,20]. The idea is to preserve the multidimensional geometric properties (vector length, inner products and distance between a pair of vectors) by perturbing the original data set through geometric rotation transformation. Hence, data mining results on the rotated data can achieve perfect accuracy. However, one important issue is that this approach is also subject to some specific attacks (e.g., a-priori knowledge PCA based attack [19] and ICA based attack [12]).

The condensation based perturbation approach [2] aims at preserving the covariance matrix for multiple columns by partitioning the original data into k-record groups and regenerating a set of k records to approximately preserve the distribution and covariance. This approach will not significantly sacrifice the accuracy of data mining results obtained from the perturbed data. However, since the difference between the regenerated records and their nearest neighbor in original data are very small, the original data records can be estimated from the perturbed data with high confidence [6].

There have been active researches on defining the right privacy measure. However, the problem how to quantify and evaluate the tradeoffs between model accuracy and privacy is still open [8,9]. Paper [3] suggests to measure privacy using Shannon's information theory. The average amount of information in the non-randomized attribute X depends on its distribution and is measured by its differential entropy. The average amount of information that remains in X after the randomized attribute Z is disclosed can be measured by the conditional differential entropy. The average information loss for X that occurs by disclosing Z can be measured in terms of the difference between the two entropies. The notion of privacy breaches captures rare disclosures. The problem with the definition of privacy breaches from [10] is that we have to specify which properties are privacy-sensitive, whose probabilities must be kept below breach level. In this paper, we use F-norm to quantify the relative amount of noise added to actual data.

7 Conclusion and future work

Spectral filtering based techniques have recently been investigated as a major means of point-wise data reconstruction [15,17,18]. It was empirically shown that those techniques may be exploited by attackers to breach the privacy protection offered by the additive randomization based privacy preserving data mining methods. This paper presented a theoretical study on evaluating privacy breaches when the spectral filtering techniques are applied. We gave an explicit upper bound of the reconstruction error. This upper bound may be exploited by

attackers to determine how close their estimates are to the original data using the spectral filtering techniques. We also derived an explicit lower bound of the reconstruction error. This lower bound can help users determine how much and what kind of noise should be added when a tolerated privacy breach threshold is given. We empirically evaluated the trade-off between privacy preservation and utility loss using one artificial data set and one real data set with three types of additive noise. Our findings showed that the i.i.d. additive noise can be mostly filtered out by the spectral filtering based techniques when strong correlations exist in the original data. As a result, individual privacy may be compromised. Another observation is that using correlated additive noise generally incurs larger utility loss in terms of the universal information measure although it can better preserve individual privacy. In the future we will explore how other types of additive noise(e.g., generated from distributions like uniform, Poisson, etc.) affect the performance of spectral filtering based techniques. We are also interested in exploring the relationship between the accuracy of data mining results and the universal information loss.

Acknowledgments This work was supported in part by US National Science Foundation CCR-0310974 and IIS-0546027. Part of this work has been presented in SAC'06 [11] and PKDD'06 [14]. All opinions, findings, conclusions and recommendations in this paper are those of the authors and do not necessarily reflect the views of the funding agencies. We would like to thank the anonymous reviewers for their insightful comments on our SAC'06 and PKDD'06 submissions and this journal submission, which help to improve the quality of the paper.

Appendix: Proof

Lemma 1 *Let $A \in \mathcal{R}^{n\times n}$ be a symmetric positive definite matrix, and let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ be its eigenvalues and $e_1, e_2, \ldots, e_n$ be corresponding eigenvectors. Let $X = [e_1 e_2, \ldots, e_k]$, $Y = [e_{k+1} \cdots e_n]$ so that the matrix $[XY] \in \mathcal{R}^{n\times n}$ is orthogonal and unitary. Given a perturbation E, let $\tilde{A} = A + E$, $\epsilon = ||E||_F > 1/2$, and define $\delta = \lambda_k - \lambda_{k+1}$.*

If $\delta > 2\sqrt{2}\epsilon$, then there is a matrix P satisfying $||P||_F \leq 2\frac{\epsilon}{\delta - \sqrt{2}\epsilon}$ so that the columns of $\tilde{X} = X + YP$ form an orthonormal basis for the subspace spanned by the first k eigenvectors of $\tilde{A}$, $\tilde{e}_1, \tilde{e}_2, \ldots, \tilde{e}_k$.

Proof Since A is a symmetric positive definite matrix, we can apply spectral decomposition on A:

$$[XY]^{\mathrm{T}} A[XY] = \begin{bmatrix} L_1 & 0 \\ 0 & L_2 \end{bmatrix}$$

where $L_1 = \mathrm{diag}(\lambda_1, \ldots, \lambda_k)$, and $L_2 = \mathrm{diag}(\lambda_{k+1}, \ldots, \lambda_n)$. Also, let

$$\tilde{E} = [X\ Y]^{\mathrm{T}} E[X\ Y] = \begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{bmatrix}.$$

From Theorem V.2.8 of [21], there exists a matrix P satisfying

$$||P|| \leq 2\frac{||F_{21}||}{\delta - ||F_{11}|| - ||F_{22}||}.$$

Since [X Y] is unitary, $\epsilon = ||E||_F$, it holds true that

$$||\tilde{E}||_F = ||[X\ Y]^T E[X\ Y]||_F = ||E||_F = \epsilon.$$

Moreover, since $||F_{11}||_F^2 + ||F_{12}||_F^2 + ||F_{21}||_F^2 + ||F_{22}||_F^2 = ||\tilde{E}||_F^2$ and $\tilde{E}$ is symmetric, we have

$$\begin{aligned}
||F_{21}||_F^2 = ||F_{12}||_F^2 &\leq \frac{1}{2}||\tilde{E}||_F \\
||F_{21}||_F = ||F_{12}||_F &\leq \frac{1}{\sqrt{2}}\epsilon \\
(||F_{11}||_F + ||F_{22}||_F)^2 &\leq 2(||F_{11}||_F^2 + ||F_{22}||_F^2) \\
&\leq 2||\tilde{E}||_F^2 \\
&= 2||E||_F^2 \\
||F_{11}||_F + ||F_{22}||_F &\leq \sqrt{2}||E||_F \\
\delta - ||F_{11}||_F - ||F_{22}||_F &\geq \delta - \sqrt{2}\epsilon.
\end{aligned}$$

Hence,

$$||P||_F \leq \sqrt{2}\frac{\epsilon}{\delta - \sqrt{2}\epsilon} \tag{13}$$

so that the columns of $\tilde{X} = (X + YP)$ form an orthonormal basis for the subspace spanned by three eigenvectors of $\tilde{A}$. The representation of $\tilde{A}$ with respect to $\tilde{X}$ is

$$\tilde{L}_1 = L_1 + F_{11} + F_{12}P.$$

The eigenvalues associated with these k eigenvectors are the eigenvalues of $\tilde{L}_1$, and the eigenvalues associate with the rest of $\tilde{A}$'s eigenvectors are the eigenvalues of

$$\tilde{L}_2 = L_2 + F_{22} - PF_{12}.$$

Thus, to complete the proof of the Lemma, it suffices to verify that the eigenvalues of $\tilde{L}_1$ are all (strictly) greater than the eigenvalues of $\tilde{L}_2$.
Since $\delta > 2\sqrt{2}\epsilon$, we have

$$||P||_F \leq \sqrt{2}\frac{\epsilon}{\delta - \sqrt{2}\epsilon} < 1.$$

Similarly, we can derive

$$||F_{11}||_F + ||F_{12}||_F \leq \sqrt{2}||E||_F.$$

Then we have

$$\begin{aligned}
||F_{11} + F_{12}P||_F &\leq ||F_{11}||_F + ||E_{12}P||_F \\
&\leq ||F_{11}||_F + ||F_{12}||_F||P||_F \\
&\leq ||F_{11}||_F + ||F_{12}||_F \\
&\leq \sqrt{2}||E||_F \\
&= \sqrt{2}\epsilon.
\end{aligned}$$

By the same argument, we also have

$$||E_{22} - PE_{12}||_F \leq \sqrt{2}\epsilon.$$

Since the Forbenius norm upper bounds the Spectral norm, this also implies

$$\begin{aligned}
||F_{11} + F_{12}P||_2 &\leq \sqrt{2}\epsilon. \\
||F_{22} - PF_{12}||_2 &\leq \sqrt{2}\epsilon.
\end{aligned}$$

Let the eigenvalues of $\tilde{L}_1$ are $\tilde{\lambda}_1, \tilde{\lambda}_2, \ldots, \tilde{\lambda}_k$, and those of $\tilde{L}_2$ are $\tilde{\lambda}_{k+1}, \ldots, \tilde{\lambda}_n$. The spectral variation of $\tilde{L}_1$ with respect to L_1 is

$$sv_{L_1}(\tilde{L}_1) = \max_{i=1}^{k} \min_{j=1}^{k} |\tilde{\lambda}_i - \lambda_j|.$$

The spectral variation of $\tilde{L}_2$ with respect to L_2 is

$$sv_{L_2}(\tilde{L}_2) = \max_{i=k+1}^{n} \min_{j=k+1}^{n} |\tilde{\lambda}_i - \lambda_j|.$$

From Corollary IV.3.4 of [21]:

$$sv_{L_1}(\tilde{L}_1) \leq ||F_{11} + F_{12}P||_2 \leq \sqrt{2}\epsilon$$
$$sv_{L_2}(\tilde{L}_2) \leq ||F_{22} - PF_{12}||_2 \leq \sqrt{2}\epsilon.$$

The above conditions ensure that those eigenvalues of $\tilde{L}_1$ lie in the interval $[\lambda_k - \sqrt{2}\epsilon, \lambda_1 + \sqrt{2}\epsilon]$, and those of $\tilde{L}_2$ lie in the interval $[\lambda_n - \sqrt{2}\epsilon, \lambda_{k+1} + \sqrt{2}\epsilon]$. As we know

$$\lambda_k - \lambda_{k+1} = \delta > 2\sqrt{2}\epsilon$$

so we have

$$\lambda_k - \sqrt{2}\epsilon > \lambda_{k+1} + \sqrt{2}\epsilon$$

which implies that all of $\tilde{L}_1$'s eigenvalues are strictly greater than all of $\tilde{L}_2$'s eigenvalues. □

Proof of Proposition 1 We can find an invariant subspace $R(\tilde{X})$ of $\tilde{A}$, and its corresponding orthogonal projection $P_{\tilde{X}} = \tilde{X}\tilde{X}^{\mathrm{T}}$. Our aim is to bound $||\tilde{X} - X||_F$, as well as $||P_X - P_{\tilde{X}}||_F$.

Let $M = P^{\mathrm{T}}P$, then $||M||_F \leq ||P||_F^2 \leq \frac{2\epsilon^2}{\tilde{\delta}^2} < 1$, where $\tilde{\delta} = \delta - \sqrt{2}\epsilon$.

$$\begin{aligned}
||\tilde{X} - X||_F &= ||(X + YP)(I + P^{\mathrm{T}}P)^{-1/2} - X||_F \\
&= ||(X + YP)(I - I + (I + P^{\mathrm{T}}P)^{-1/2}) - X||_F \\
&= ||X + YP - (X + YP)(I - (I + M)^{-1/2}) - X||_F \\
&\leq ||YP||_F + ||(X + YP)(I - (I + M)^{-1/2})||_F \\
&= ||P||_F + ||(X + YP)(I - (I + M)^{-1/2})||_F \\
&\leq ||P||_F + (||X||_F + ||YP||_F)||(I - (I + M)^{-1/2})||_F \\
&\leq ||P||_F + (||X||_F + ||YP||_F)\frac{2\epsilon^2}{\tilde{\delta}^2} \\
&= ||P||_F + (||X||_F + ||P||_F)\frac{2\epsilon^2}{\tilde{\delta}^2} \\
&\leq \frac{\sqrt{2}\epsilon}{\tilde{\delta}} + \left(\sqrt{k} + \frac{\sqrt{2}\epsilon}{\tilde{\delta}}\right)\frac{2\epsilon^2}{\tilde{\delta}^2} \\
&< \frac{\sqrt{2}\epsilon}{\tilde{\delta}} + (\sqrt{k} + 1)\frac{\sqrt{2}\epsilon}{\tilde{\delta}} \\
&= (\sqrt{k} + 2)\frac{\sqrt{2}\epsilon}{\tilde{\delta}}.
\end{aligned}$$

According to [21, pp. 232], we can derive:

$$\begin{aligned} ||P_X - P_{\tilde{X}}||_F &\leq 2\sqrt{2}\frac{||F_{12}||_F}{\delta - ||F_{11}||_F - ||F_{22}||_F} \\ &\leq 2\sqrt{2}\frac{\frac{1}{\sqrt{2}}\epsilon}{\delta - \sqrt{2}\epsilon} \\ &= \frac{2\epsilon}{\delta - \sqrt{2}\epsilon}. \end{aligned} \tag{14}$$

□

Proof of Result 3 In the spectral filtering method, when we select the first k components, the error matrix can be expressed as

$$\begin{aligned} f(k) &= \hat{U} - U \\ &= (U+V)\tilde{Q}_k\tilde{Q}_k^{\mathrm{T}} - U \\ &= (U+V)\tilde{Q}\begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix}\tilde{Q}^{\mathrm{T}} - U \\ &= V\tilde{Q}\begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix}\tilde{Q}^{\mathrm{T}} - U\left[\tilde{Q}I\tilde{Q}^{\mathrm{T}} - \tilde{Q}\begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix}\tilde{Q}^{\mathrm{T}}\right] \\ &= V\tilde{Q}\begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix}\tilde{Q}^{\mathrm{T}} - U\tilde{Q}\begin{pmatrix} 0 & 0 \\ 0 & I_{n-k} \end{pmatrix}\tilde{Q}^{\mathrm{T}}. \end{aligned} \tag{15}$$

Similarly, when we select the first $k+1$ components, the error matrix becomes

$$\begin{aligned} f(k+1) &= V\tilde{Q}\begin{pmatrix} I_{k+1} & 0 \\ 0 & 0 \end{pmatrix}\tilde{Q}^{\mathrm{T}} - U\tilde{Q}\begin{pmatrix} 0 & 0 \\ 0 & I_{n-k-1} \end{pmatrix}\tilde{Q}^{\mathrm{T}} \\ &= V\left[\tilde{Q}\begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix}\tilde{Q}^{\mathrm{T}} + \tilde{e}_{k+1}\tilde{e}_{k+1}^{\mathrm{T}}\right] - U\left[\tilde{Q}\begin{pmatrix} 0 & 0 \\ 0 & I_{n-k} \end{pmatrix}\tilde{Q}^{\mathrm{T}} - \tilde{e}_{k+1}\tilde{e}_{k+1}^{\mathrm{T}}\right] \\ &= \left(V\tilde{Q}\begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix}\tilde{Q}^{\mathrm{T}} - U\tilde{Q}\begin{pmatrix} 0 & 0 \\ 0 & I_{n-k} \end{pmatrix}\tilde{Q}^{\mathrm{T}}\right) + V\tilde{e}_{k+1}\tilde{e}_{k+1}^{\mathrm{T}} + U\tilde{e}_{k+1}\tilde{e}_{k+1}^{\mathrm{T}} \\ &= f(k) + V\tilde{e}_{k+1}\tilde{e}_{k+1}^{\mathrm{T}} + U\tilde{e}_{k+1}\tilde{e}_{k+1}^{\mathrm{T}}. \end{aligned} \tag{16}$$

The last two parts in Eq. 16 are the projections of noise and data on the $(k+1)$th eigenvector. When the magnitude of the added noise is small compared with that of the original data, we have $\tilde{e}_i \approx e_i$. The strength of the data projection can be approximated as

$$\begin{aligned} ||U\tilde{e}_{k+1}\tilde{e}_{k+1}^{\mathrm{T}}||_F^2 &\approx ||Ue_{k+1}e_{k+1}^{\mathrm{T}}||_F^2 \\ &= Tr[(Ue_{k+1}e_{k+1}^{\mathrm{T}})^{\mathrm{T}}(Ue_{k+1}e_{k+1}^{\mathrm{T}})] \\ &= Tr\left(e_{k+1}e_{k+1}^{\mathrm{T}}U^{\mathrm{T}}Ue_{k+1}e_{k+1}^{\mathrm{T}}\right) \\ &= Tr\left[e_{k+1}e_{k+1}^{\mathrm{T}}\left(\sum_{i=1}^{n}\lambda_i e_i e_i^{\mathrm{T}}\right)e_{k+1}e_{k+1}^{\mathrm{T}}\right] \\ &= Tr(\lambda_{k+1}e_{k+1}e_{k+1}^{\mathrm{T}}) \\ &= \lambda_{k+1}. \end{aligned}$$

For i.i.d noise, the effect of the projection on any vector should be the same. Thus,

$$||V\tilde{e}_{k+1}\tilde{e}_{k+1}^{\mathrm{T}}||_F^2 = \lambda_V.$$

Hence, we include the ith component only when $\lambda_i \geq \lambda_V$. The benefit due to inclusion of the ith eigen component is greater than the loss due to the noise projected along the ith eigen component. Since $\tilde{\lambda}_i = \lambda_i + \lambda_V \geq 2\lambda_V$, hence

$$k = \max\{i | \tilde{\lambda}_i \geq 2\lambda_V\}.$$

□

References

1. Adam NR, Wortman JC (1989) Security-control methods for statistical databases. ACM Comput Surv 21(4):515–556
2. Aggarwal C, Yu P (2004) A condensation approach to privacy preserving data mining. In: Proceedings of international conference on extending database technology, LNCS, vol 2992, pp 183–199
3. Agrawal D, Agrawal C (2001) On the design and quantification of privacy preserving data mining algorithms. In: Proceedings of the 20th symposium on principles of database systems, Santa Barbara, pp 247–255
4. Agrawal R, Srikant R (2000) Privacy-preserving data mining. In: Proceedings of the ACM SIGMOD international conference on management of data, Dallas, pp 439–450
5. Asuncion A, Newman DJ (2007) UCI machine learning repository. University of California, School of Information and Computer Science, Irvine. http://www.ics.uci.edu/~mlearn/MLRepository.html
6. Chen K, Liu L (2005) Privacy preserving data classification with rotation perturbation. In: Proceedings of the 5th IEEE international conference on data mining, Houston pp 589–592
7. Dobra A, Fienberg SE (2001) Bounds for cell entries in contingency tables induced by fixed marginal totals with applications to disclosure limitation. Stat J United Nations ECE 18, pp 363–371
8. Evfimievski A (2002) Randomization in privacy preserving data mining. SIGKDD Explor 4(2):43–48
9. Evfimievski A, Gehrke J, Srikant R (2003) Limiting privacy breaches in privacy preserving data mining. In: Proceedings of the 22nd ACM SIGMOD-SIGACT-SIGART symposium on principles of database system, pp 211–222
10. Evfimievski A, Srikant R, Agrawal R, Gehrke J (2002) Privacy preserving mining of association rules. In: Proceedings of the 8th ACM SIGKDD international conference on knowledge discovery and data mining, Edmonton, pp 217–228
11. Guo S, Wu X (2006) On the use of spectral filtering for privacy preserving data mining. In: Proceedings of the 21st ACM symposium on applied computing, pp 622–626
12. Guo S, Wu X (2007) Deriving private information from arbitrarily projected data. In: Proceedings of the 11th Pacific-Asia conference on knowledge discovery and data mining, pp 84–95
13. Guo S, Wu X, Li Y (2006a) Deriving private information from perturbed data using iqr based approach. In: Proceedings of the 22nd international conference on data engineering workshops, pp 92–101
14. Guo S, Wu X, Li Y (2006b) On the lower bound of reconstruction error for spectral filtering based privacy preserving data mining. In: Proceedings of the 10th European conference on principles and practice of knowledge discovery in databases, Berlin, pp 520–527
15. Huang Z, Du W, Chen B (2005) Deriving private information from randomized data. In: Proceedings of the ACM SIGMOD conference on management of data, Baltimore, pp 37–48
16. Kantarcioglu M, Jin J, Clifton C (2004) When do data mining results violate privacy? In: Proceedings of the 10th ACM SIGKDD international conference on knowledge discovery and data mining, pp 599–604
17. Kargupta H, Datta S, Wang Q, Sivakumar K (2003) On the privacy preserving properties of random data perturbation techniques. In: Proceedings of the 3rd international conference on data mining, pp 99–106
18. Kargupta H, Datta S, Wang Q, Sivakumar K (2005) Random-data perturbation techniques and privacy-preserving data mining. Knowl Inf Syst 7(4):387–414
19. Liu K, Giannella C, Kargupta H (2006) An attacker's view of distance preserving maps for privacy preserving data mining. In: Proceedings of the 10th European conference on principles and practice of knowledge discovery in databases (PKDD'06), Berlin, pp 297–308
20. Liu K, Kargupta H, Ryan J (2006) Random projection based multiplicative data perturbation for privacy preserving distributed data mining. IEEE Trans Knowl Data Eng 18(1):92–106
21. Stewart G, Sun J (1990) Matrix perturbation theory. Academic Press, London
22. Sung SY, Liu Y, Xiong H, Ng PA (2006) Privacy preservation for data cubes. Knowl Inf Syst 9(1):38–61
23. Wang K, Fung BCM, Yu PS (2007) Handicapping attacker's confidence: an alternative to k-anonymization. Knowl Inf Syst 11(3):345–368

Author Biographies

Songtao Guo received his Ph.D. degree in Information Technology from the Department of Software and Information Systems, University of North Carolina at Charlotte in 2007. He earned his BS degree in Computer Communication in 2000 and his ME in Signal and Information Processing in 2003 from the University of Electronic Science and Technology of China. His major research interests include privacy preserving data mining and information security.

Xintao Wu is an Associate Professor in the Department of Software and Information Systems at the University of North Carolina at Charlotte, USA. He got his Ph.D. degree in Information Technology from George Mason University in 2001. He received his BS degree in Information Science from the University of Science and Technology of China in 1994, an ME degree in Computer Engineering from the Chinese Academy of Space Technology in 1997. His major research interests include data mining and knowledge discovery, data privacy and security, and bioinformatics.

Yingjiu Li is currently an Assistant Professor in the School of Information Systems at Singapore Management University, Singapore. He received his Ph.D. degree in Information Technology from George Mason University in 2003. His research interests include applications security, privacy protection, and data rights management. He has published over 40 technical papers in the refereed journals and conference proceedings. Yingjiu Li is a member of the ACM and the IEEE. The URL for his web page is http://www.mysmu.edu/faculty/yjli/.