Int. J. Appl. Math. Comput. Sci., 2013, Vol. 23, No. 2, 447–461
DOI: 10.2478/amcs-2013-0034

RANDOM PROJECTIONS AND HOTELLING'S T^2 STATISTICS FOR CHANGE DETECTION IN HIGH–DIMENSIONAL DATA STREAMS

EWA SKUBALSKA-RAFAJŁOWICZ

Institute of Computer Engineering, Automation and Robotics
Technical University of Wrocław, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland
e-mail: ewa.rafajlowicz@pwr.wroc.pl

The method of change (or anomaly) detection in high-dimensional discrete-time processes using a multivariate Hotelling chart is presented. We use normal random projections as a method of dimensionality reduction. We indicate diagnostic properties of the Hotelling control chart applied to data projected onto a random subspace of $\mathbb{R}^n$. We examine the random projection method using artificial noisy image sequences as examples.

Keywords: change detection, multidimensional control charts, dimensionality reduction, random projections, process monitoring.

1. Introduction

Detection of abnormal measurements, when carried out on independent and identically distributed (i.i.d) observations, is usually referred to as outlier detection and is in many ways equivalent to density estimation. We assume that the data, d-dimensional vectors, are independent random samples drawn from an unknown distribution, let us say f_0. If the distribution changes from f_0 to any other distribution f_1 (incidentally or permanently), we should be able to detect this event with a possibly highest probability. Furthermore, in many cases it is assumed that only some parameters of the distribution, typically moments, may change and thus only these have to be monitored (Montgomery, 1996; Zorriassatine *et al.*, 2003).

In many industrial applications, it has become more and more important to monitor the behavior of complex systems using multivariate measurements. High-dimensional data: large images, DNA microarray data and multidimensional time-series are obtained during process monitoring. Many chemical and process plants (semiconductor manufacturing, for example) maintain this process and its quality by using hundreds and thousands of variables. Manufacturing databases contain streams of images, hyperspectral images or videos where the observations have dimensions ranging from thousands to billions. Statistical monitoring of high-dimensional processes for faulty state detection is becoming increasingly important. In fact, it is an open problem, especially when measurements, images etc. form data sets of extremely large dimensions. A control chart based on the Hotelling statistics, i.e., the Hotelling control chart, is the most popular method for handling changes of multivariate process data (Montgomery, 1996). Traditional methods used in statistical process monitoring are restricted to small-size samples with dimensionality not greater than hundreds (Wang and Jiang, 2009).

The problem of dimensionality reduction (Lee and Verleysen, 2007; Wang, 2012) is especially important when the number of observation, however large, is much smaller than the dimension of the data. It is obvious that dimensionality reduction is a major technique in dealing with high dimensional processes.

Principal Component Analysis (PCA) (Jolliffe, 1986; Lee and Verleysen, 2007), Partial Least Squares (PLS) (Wold, 1966) or Independent Component Analysis (ICA) (Hyvärinen *et al.*, 2001) are, among others, the most popular methods used to extract a smaller set of variables that are representative enough for process monitoring (Tsung and Wang, 2010). Unfortunately, if the dimension of the data is really large, straightforward usage of these methods is impossible. Some efforts in this direction have been made by Skubalska-Rafajłowicz (2006; 2008), who proposed high-dimensional novelty detection methods in nonparametric settings.

The goal of this paper is to show that random

projections (see Achlioptas, 2001; Ailon and Chazelle, 2006; Arriaga and Vempala, 1999; Biau *et al.*, 2008; Dasgupta and Gupta, 2003; Frankl and Maehara, 1987; Indyk and Motwani, 1998; Indyk and Naor, 2007; Li *et al.*, 2006a; Matousek, 2008; Vempala, 2004; Wang, 2012) can be efficiently used for change detection during statistical monitoring of a high-dimensional process. We propose here a new approach to monitoring high-dimensional time series that is based on the Hotelling statistic (see Hotelling, 1931; Mason *et al.*, 1992; Mason and Young, 2002; Montgomery, 1996; Rao, 1973) but performed on projected observations. Thus, the proposed method can be used for really high-dimensional problems. We have assumed that the variance-covariance matrix is not known and due to a high dimensionality its estimation is impossible even if the number of reference observations is formally sufficient. Furthermore, we develop some new properties of random projections when applied to multivariate independent observations under normality assumptions.

It should be indicated that these methods can also be applied to fault diagnosis in dynamical systems (Korbicz *et al.*, 2004). In discrete time stochastic processes we seek to detect whether a new observation abnormally deviates from the usual dynamics of the system. This problem can be reduced to density monitoring when a suitable representation of the system observation is Markovian (Cuturi *et al.*, 2009) and a conditional probability density, describing the current behavior of the system given past observation, is known.

Our approach is addressed mostly to problems of high-dimensional process monitoring with spatial temporal or spatio-temporal dependencies, as, for example, to distributed sensor networks based monitoring, and it is designated for detection of changes in the mean that is composed of many small changes of a vector's coordinates.

Preliminary results of using random projection for change detection in high-dimensional data streams were presented at the Conference on *Diagnostics of Processes and Systems 2011* (Skubalska-Rafajlowicz, 2011).

In the next section some prerequisites concerning statistical process monitoring and control charts are given. Section 3 presents linear random projections as a method of dimensionality reduction. The new method of monitoring multidimensional normal data using random projections and the Hotelling statistic is introduced in Section 4, and in the next section its diagnostic properties are summarized. Experiments presented in Section 6 illustrate the performance of the proposed method used for mean shifts detection. A discussion and final conclusions are given in the last section.

2. Statistical process monitoring and control charts

Control charts are one of the most popular and powerful techniques of Statistical Process Control (SPC). Control charts are used for statistical process monitoring and thereafter process improvement (Mason and Young, 2002; Montgomery, 1996). Statistical control charts are designed in order to detect abnormalities (out-of control states) in the process under consideration. The most common abnormalities are mean shifts, variance or covariance matrix changes which indicate process faults. A typical Shewhart-type control chart (Mason and Young, 2002; Montgomery, 1996) has control limits, which are set at values such that if the process is in control, nearly all points of a selected statistic (for example, sample means) will be situated between the upper control limit and the lower control limit. In other words, control limits form an acceptance interval (region) for monitored statistic values. Usually choosing the control limits is equivalent to setting up the critical region for testing hypothesis. The control chart only detects assignable causes and an engineering action will be necessary to eliminate them. However, control charts provide valuable diagnostic information, are effective in defect prevention and are a reliable technique for improving quality. A control chart can distinguish (in a statistical sense) between a noise in an in-control state and abnormal variations, indicating a fault in a final product, in a part of a production process, in a separate tool or in materials (resources). Thus, control charts provide diagnostic information which can be used on different decision levels.

The most familiar multivariate process monitoring is the Hotelling T^2 control chart (Mason and Young, 2002; Wang and Jiang, 2009) which is based on the Hotelling statistic (Rao, 1973; Srivastava, 2009). The basis of the T^2 statistic is knowledge of the covariance matrix Σ or its good estimate.

Hotelling's T^2 chart is designed for monitoring the mean vector of the process. It is more sensitive to large mean shifts than to small shifts; while some other schemes, such as the Multivariate Exponentially Weighted Moving Average (MEWMA) chart and the Multivariate CUmulative SUM (MCUSUM) chart, are more sensitive to shifts with small sizes. All these charts are based on an analysis of the Mahalanobis distance or its approximations (Montgomery, 1996; Sullivan and Woodall, 2000).

2.1. Hotelling control chart. Suppose that $X_1, X_2, \ldots$ are independent identically distributed (i.i.d.) normal random vectors observed sequentially, i.e., each X_i is an independent copy of a normal random vector $X \sim \mathcal{N}_d(0, \Sigma)$. If $\Sigma = I_d$, then each $||X_i||^2$ is random variable which is χ^2_d distributed with the expectation d

and the variance $2d$. In the general, non-standardized case, i.e., when $\Sigma \neq I_d$, $||X_i||^2$ is a linear combination of at most d random variables which are independent and each follows χ^2 distribution with one degree of freedom. It is well known that after standardization (if Σ is a full rank matrix), $X_i^T \Sigma^{-1} X_i = ||\Sigma^{-1/2} X_i||^2$ follows the χ^2_d distribution. Further, we will assume that Σ is a full rank matrix. T^2 statistic in $\mathbb{R}^d$ is

$$T^2 = X^T C^{-1} X,$$

where the covariance matrix Σ is not known and

$$C = \frac{1}{N} \sum_{i=1}^{N} X_i X_i^T$$

is its sample based estimate (Rao, 1973).

It is known that in our case

$$T^2 \sim \frac{dN}{N-d+1} F(d, N-d+1), \quad N \geq d,$$

where F is the Snedecor distribution (Forbes *et al.*, 2011).

The $F(\nu_1, \nu_2)$ distribution is the ratio of two chi-square distributions with degrees of freedom ν_1 and ν_2, respectively, where each chi-square has first been divided by its degrees of freedom. Furthermore, if the dimensionality of data is large, one can approximate this distribution using a chi-square distribution and then a univariate normal distribution. If

$$T^2 < h_\alpha = \frac{dN}{N-d+1} F_\alpha(d, N-d+1),$$

where α is a significance level, we decide that the process is not faulty (in-control). The h_α value forms the upper control limit. A lower control limit is generally not used in the Hotelling control chart.

If the sample size $N \ll d$ and the covariance matrix Σ is not known, its sample based estimate $C = \frac{1}{N} \sum_{i=1}^{N} X_i X_i^T$ is a singular matrix (with probability one). Thus, it is not possible to use Hotelling's T^2 in d dimensions (Srivastava, 2009). One of the possibilities to deal with this problem is to reduce the dimension of the data. Even if we have a sufficient number of observations, the sample covariance matrix may be too large to be estimated.

3. Linear random projections for dimensionality reduction

Random projections (Johnson and Lindenstrauss, 1984; Frankl and Maehara, 1987; Indyk and Motwani, 1998; Arriaga and Vempala, 1999; Achlioptas, 2001; Dasgupta and Gupta, 2003; Vempala, 2004; Ailon and Chazelle, 2006) are widely considered to be one of the most potential methods of dimensionality reduction (Biau *et al.*, 2008; Lee and Verleysen, 2007; Wang, 2012), i.e., the transformation of data from a high-dimensional vector space to a space of lower dimensions, when the prior dimension is very large. The techniques of random projections can be traced back to Milman's proof of the Dvoretsky theorem (Milman, 1971).

In the random projection method, the original high-dimensional observations are projected onto a lower-dimensional space using a suitably scaled random matrix with independent, typically, normally distributed entries.

A random projection from d dimensions to k dimensions is a linear transformation represented by a $d \times k$ matrix $S \in \mathbb{R}^{k \times d}$, i.e., a matrix whose entries are i.i.d. samples of a certain random variable.

3.1. Projection. The projection $x \mapsto v = Sx$, $\mathbb{R}^d \to \mathbb{R}^k$, $k \ll d$, is defined by a projection matrix S:

$$\begin{bmatrix} s_{11} & s_{12} & \dots & s_{1d} \\ s_{21} & s_{22} & \dots & s_{2d} \\ \vdots & \vdots & \dots & \vdots \\ s_{k1} & s_{k2} & \dots & s_{kd} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_d \end{bmatrix} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_k \end{bmatrix}. \quad (1)$$

Random projections are closely related to the Johnson–Lindenstrauss lemma (Johnson and Lindenstrauss, 1984), which states that any set A, say, of N points in an Euclidean space can be embedded in an Euclidean space of a lower dimension ($\sim O(\log N)$) with a relatively small distortion of the distances between any pair of points from A. The Johnson–Lindenstrauss lemma has been shown to be useful in applications in computer science and engineering (Achlioptas, 2001; Ailon and Chazelle, 2006; Arriaga and Vempala, 1999; Biau *et al.*, 2008; Donoho, 2000; Indyk and Motwani, 1998; Li *et al.*, 2006a; 2006b; Matousek, 2008; Vempala, 2004; Skubalska-Rafajłowicz, 2008; 2009), among many others.

The main idea of a random projection is that we can estimate the distance between two points (two vectors), let us say u and z, in the d-dimensional Euclidean space $D^2(u,z) = ||u-z||^2 = (u-z)^T(u-z)$, $u,\ z \in \mathbb{R}^d$, from the sample squared distances as follows:

$$\hat{D}^2(u,z) = \frac{1}{k} \sum_{j=1}^{k} (s_j(u-z))^2 = \frac{1}{k} ||Su - Sz||^2, \quad (2)$$

where s_j is the j-th row of S, i.e., an individual projection.

Thus, for any chosen pair of vectors $u, v \in \mathbb{R}^d$ $E(\hat{D}^2) = D^2$, $\mathrm{var}(\hat{D}^2) = \frac{2}{k} D^4$ and $\frac{k\hat{D}}{D^2} \sim \chi^2_k$. These facts lead to the conclusion that

$$\Pr\left\{ \frac{|\hat{D}^2 - D^2|}{D^2} \geq \varepsilon \right\} \leq 2 \exp\left(-\frac{k}{4}\varepsilon^2 + \frac{k}{6}\varepsilon^3 \right), \quad (3)$$

where $\varepsilon \in (0,1)$ (for details, see, e.g., Dasgupta and Gupta, 2003; Vempala, 2004).

Moreover, it is easy to see that, because the random projection defined by $S \in \mathbb{R}^{k\times d}$ is a linear mapping, we obtain

$$S\Big(\frac{1}{N}\sum_{i=1}^{N} z_i\Big) = \frac{1}{N}\sum_{i=1}^{N} Sz_i, \tag{4}$$

where $z_1, \ldots, z_N \in \mathbb{R}^d$ are any given column vectors. This means that any change in the projected vectors $Sz_1, \ldots, Sz_N$ mean indicates a change in the original average $\frac{1}{N}\sum_{i=1}^{N} z_i$.

The above simple property can be easily generalized as a mean invariance under a linear random projection property, which is a basis for the method proposed in this paper. Since, if $X_i,\ i = 1, \ldots, N$ are i.i.d., finite first moment ($E(X_i) = E(X) < \infty,\ i = 1, \ldots, N$) random vectors, then

$$E\{S\frac{1}{N}\sum_{i=1}^{N} X_i | S\} = SE(X),$$

where $E|S$ stands for the conditional expectation, given S.

3.2. Sparse random projections. We do not have to use $s_{ij} \sim \mathcal{N}(0,1)$ for dimension reduction in a space with Euclidean norm. For example, we can sample s_{ij} from any sub-Gaussian tails distributions (Ailon and Chazelle, 2006; Matousek, 2008). We say that a random variable X has a sub-Gaussian tail if both X and $-X$ have sub-Gaussian upper tails, i.e., if there exists a constant $a > 0$ such that, for all $\varepsilon > 0$,

$$\Pr\{|X| > \varepsilon\} \le \exp(-a\varepsilon^2). \tag{5}$$

To speed up the computations, one can generate a sparse random projection matrix. In particular, Achlioptas (2001) proved that the entries of the projection matrix S can be chosen as independent random variables with values $+1,\ -1$ (each with probability $1/2$).

More generally, a projection matrix composed from independent entries of the form

$$s_{ij} = \sqrt{c}\begin{cases} 1 & \text{with probability } 1/2c, \\ 0 & \text{with probability } 1-1/c, \\ -1 & \text{with probability } 1/2c, \end{cases} \tag{6}$$

leads to the projection distribution with sub-Gaussian tails (at least up to a suitable threshold).

An Achlioptas variant of this result (with $c = 3$) has the entries attaining a value of 0 with probability $2/3$ and values $+1$ and -1 with probability $1/6$ each. This setting allows computing the projection about three times faster than the Gaussian. Since S is sparse, only about one third of the entries are nonzero numbers.

The proof of the Johnson–Lindenstrass lemma using a projection matrix with the entries to be arbitrary independent random variables with zero mean, unit variance, and a sub-Gaussian tails is given by Matousek (2008). To avoid misunderstanding, we will distinguish between the random projection matrix $\mathbf{S}$ as a matrix of independent random variables and a matrix of real numbers $S \in \mathbb{R}^{k\times d}$, a sample of $\mathbf{S}$.

4. Monitoring multidimensional normal data streams using random projection

We restrict here our attention to normal random projections of data. However, almost all results presented in this section can be generalized to random projections with sub-Gaussian tails.

So, we assume that of the elements the projection matrix S are obtained as independent, identically distributed samples of a normal random variable (or a variable with sub-Gaussian tails), having zero mean and unit variance.

Throughout this paper we assume that observations are random vectors X which follow a multivariate Gaussian distribution, i.e., that multidimensional data points, are i.i.d. normal $\mathcal{N}_d(0,\Sigma)$. Σ is a symmetric, nonnegative-definite covariance matrix, but it is not known.

4.1. Design phase: Analysis of historical data. We assume that there are fewer observations than the dimension of the observation vector X, i.e., the collected sample size N is less than the dimension d.

In the design phase we project the collected data $X_1, X_2, \ldots, X_N$ into a k-dimensional ($k < N \ll d$) space using random linear projections with a projection matrix S. However, once generated projection matrix S is kept constant (the same for all data).

Since X_i are i.i.d. random normal vectors $X_i \sim \mathcal{N}_d(0,\Sigma)$, we obtain k-dimensional random vectors V which also follow a normal distribution. More precisely, for any constant projection matrix S,

$$V_i = S\,X_i, \quad V_i|S \sim \mathcal{N}_k(0,\, S\,\Sigma\,S^T), \quad i = 1, \ldots, N. \tag{7}$$

Note that for the sake of simplicity we assume that $E(X_i) = 0$. Thus, we also have $E(V_i) = 0$.

The next step is to compute the new covariance matrix estimate for random vector V, based on projected samples:

$$C_S = \frac{1}{N}\sum_{i=1}^{N} V_i V_i^T. \tag{8}$$

If the new dimension $k < N$, we can expect that the matrix C_S is nonsingular. C_S is a $k \times k$, full rank matrix

(with probability 1 with respect to S), even if the original covariance matrix Σ is singular.

Note that for any projection S C_S is a random (with respect to observations) matrix

$$C_S = \frac{1}{N}\sum_{i=1}^{N} SX_iX_i^TS^T = SCS^T, \tag{9}$$

where

$$C = \frac{1}{N}\sum_{i=1}^{N} X_iX_i^T$$

is a singular estimate of the covariance matrix Σ.

Furthermore, for a constant projection matrix S, $N\,C_S|S$ is distributed as a Wishart distribution of dimensionality k with N degrees of freedom and the associated parameter matrix $S\,\Sigma\,S^T$ (see, e.g., the works of Mathai and Provost (1992) as well as Rao (1973) for known results):

$$N\,C_S|S \sim W_k(N, S\,\Sigma\,S^T). \tag{10}$$

Note that the previous results are true for any fixed projection matrix of a rank k.

4.2. Detection phase. Now, a new independent observation $X \sim \mathcal{N}_d(0,\Sigma)$ is projected using the projection matrix S, the same as previously. So, we obtain $V = S\,X|S$. $V|S$ is a normal random vector which also follows $\sim \mathcal{N}_k(0, S\Sigma S^T)$. We define the new Hotelling statistic for projected data as follows:

$$T_S^2 = (SX)^T C_S^{-1} SX|S \sim \frac{kN}{N-k+1}F(k, N-k+1),$$

where F is the Snedecor distribution with k and $N-k+1$ (numerator and denominator) degrees of freedom. Note that we have assumed that $E[X] = 0$. Otherwise, one should replace X in the formula for T_S^2 by the zero mean random vector $X - E[X]$.

When $E[X]$ (or $E[V] = SE[X]$) is not known and should be estimated, $V - \frac{1}{N}\sum_{i=1}^{N} V_i$ follows the normal distribution $\mathcal{N}_k(0, \frac{N+1}{N}S\Sigma S^T)$ and the corrected statistic T_S^2 follows the Snedecor distribution with k and $N-k$. More precisely,

$$T_S^2 \sim \frac{kN}{N-k}F(k, N-k),$$

since $V - \frac{1}{N}\sum_{i=1}^{N} V_i$ follows the normal distribution $\mathcal{N}_k(0, (1+\frac{1}{N})S\Sigma S^T)$ and the distribution of a covariance matrix estimate follows the Wishart distribution $\frac{1}{N}W_k(N-1, S\,\Sigma\,S^T)$. Notice that the distribution of T_S^2 in the in–control state does not depend on a projection matrix S, but it depends on k and N, i.e., on the dimensionality of the data after projection and on a number of historical data, respectively.

It is well known (see, e.g., Srivastava, 2009) that the Hotelling statistic is invariant under transformations by an element of the general linear group of the degree k, i.e., the set of $k \times k$ invertible matrices. Thus, if C_S is a full rank matrix, then for any nonsingular matrix $B \in \mathbb{R}^{k\times k}$ we have

$$\begin{aligned}
&(BV)^T\left(\sum_i (BV_i)(BV_i)^T\right)^{-1} BV \\
&\quad = V^TB^T\left(B(\sum_i V_iV_i^T)B^T\right)^{-1} BV \\
&\quad = V^TB^T(B^T)^{-1}\left(\sum_i V_iV_i^T\right)^{-1} B^{-1}BV \\
&\quad = V^T\left(\sum_i V_iV_i^T\right)^{-1} V.
\end{aligned}$$

In the out-of-control state, when a change in the mean occurred, i.e., $E(X) = m$, we have $V = SX|S \sim \mathcal{N}_k(Sm, S\Sigma S^T)$ and

$$T_S^2 \sim \frac{kN}{N-k+1}F(k, N-k+1, (Sm)^T(S\Sigma S^T)^{-1}Sm), \tag{11}$$

i.e., it follows a noncentral F-Snedecor distribution with non-centrality parameter $\lambda^2 = (Sm)^T(S\Sigma S^T)^{-1}Sm$ and with the degrees of freedom k and $N-k+1$, respectively. Properties of the T_S^2 in the out-of-control state are examined in greater detail in the next section.

If $T_S^2 < h_\alpha$, where α is a significance level, we decide that the process is not faulty (in-control). Otherwise, a value of the T_S^2 statistic larger than h_α indicates that the new observation does not follow the assumed distribution and some undesirable changes may occur. The appropriate h_α value is equal to $\frac{kN}{N-k+1}G_{(k,N-k+1)}(1-\alpha)$, where $G_{(k,N-k+1)}(1-\alpha)$ is the inverse of the Snedecor probability distribution function (quantiles) at probability level $1-\alpha$.

Since $aF(a,b)$ tends to the chi-squared variate χ_a^2 as b tends to infinity, we can use $1-\alpha$ quantiles of the chi-squared distribution to obtain h_α for projections of the dimensionality greater than 50.

Here α is equal to the probability of Type I errors. It controls the false alarm rate when the monitored process is in the normal state (in-control) and no changes occur in the distribution of observation X and consequently, in the distribution of V.

Summarizing, we propose the following simple procedure:

- *Phase I (Design)*

  1. Collect historical data X_i, $1,\ldots,N$. 2. Select k and choose the projection matrix S at random,

i.e., each entry $s_{ij},\ i = 1,\ldots,k,\ j = 1,\ldots,d$, of the matrix S is drawn (independently) from the standard normal distribution $\mathcal{N}(0,1)$ (or according to (6)). Once generated, the projection matrix S is kept fixed. Select α and h_α using the Snedecor or χ^2_k distribution tables.

3. Estimate C_S matrix according to (8).

- *Phase II (Detection)*

  When a new observation X appears, evaluate the T^2_S statistic and then decide about an alarm if $T^2_S > h_\alpha$.

Remark 1. Notice that the T^2_S statistic is scalar invariant with respect to the projection matrix S, i.e., $T^2_S = T^2_{aS}$ for any scalar constant a.

5. Diagnostic properties of the Hoteling chart after data projection

In this section we examine diagnostic properties of the Hotelling control chart applied to data projected onto random subspace of $\mathbb{R}^d$.

The main question is whether we can detect on the basis of T^2_S a change in the mean of the multivariate normal distribution, i.e., the detect that a new observation X follows the normal distribution $\mathcal{N}_d(m,\Sigma)$, where m is a change in the mean. In this case, $V = SX|S \sim \mathcal{N}_k(Sm, S\Sigma S^T)$ and

$$T^2_S \sim \frac{kN}{N-k+1} F(k, N-k+1, (Sm)^T (S\Sigma S^T)^{-1} Sm)$$

i.e., noncentral F-Snedecor distribution with degrees of freedom k and $N-k+1$, respectively.

Since S defines a singular linear transformation, a Type II error (no alarm despite the fact that a change occurred) depends on the new vector Sm, but not on m directly.

Remark 2. The matrix $S^T(S\Sigma S^T)^{-1}S$ is a $d \times d$ matrix of rank k, i.e., it is always rank deficient. The matrix $S\Sigma S^T$ is symmetric and positive definite (assuming that $N > k$ and the rank of Σ is $\geq k$), since it is a full rank $k \times k$ matrix with probability one (similarly, the matrix C_S is also nonsingular with probability one, see Eqns. (8) and (9)). Thus, the quadratic form $(Sm)^T(S\Sigma S^T)^{-1}(Sm) = 0$ if and only if $Sm = 0$. This means that, for $m \neq 0$, $Sm = 0$ iff m is orthogonal to all k row vectors of S, but this event has zero probability.

Moreover, Σ is not known and C_S, which estimates $S\Sigma S^T$, contains only information about the total variance of the multivariate normal distribution, i.e., the trace of Σ. If $k < d$, then $S^T(S\Sigma S^T)^{-1}S \neq \Sigma^{-1}$, because S is a $k \times d$ matrix.

Observe that

$$E_{X_1,\ldots,X_N}[C_S] = S\Sigma S^T. \tag{12}$$

The expectation of a $k \times k$ matrix $\mathbf{S}\Sigma\mathbf{S}^T$ over the space of projections S equals

$$M_k = E_{\mathbf{S}}[\mathbf{S}\Sigma\mathbf{S}^T] = [E\{s_i^T \Sigma s_j\}]_{i,\,j=1,\ldots,k}, \tag{13}$$

and since all entries of $\mathbf{S}$ are zero mean, unit variance i.d.d. random variables and $S \in \mathbb{R}^{k\times d}$, we get

$$M_k = \Big[E\{\sum_{r=1}^{d}\sum_{l=1}^{d} s_{il}\sigma_l\sigma_r s_{jr}\}\Big]_{ij} = \mathrm{trace}[\Sigma]\ I_k. \tag{14}$$

This means that random projections lead to eccentricity reduction of the distribution under consideration.

Note that the expected value of $T^2_S = V^T C_S^{-1} V$ with respect to X, conditioned on data $X_1,\ldots,X_N$ and the projection matrix S, is given by

$$\begin{aligned} &E_X\{V^T C_S^{-1} V|S,\ X_1,\ldots,X_N\} \\ &\quad = E_X\{X^T S^T C_S^{-1} SX|S,\ X_1,\ldots,X_N\}. \end{aligned} \tag{15}$$

Thus, $X^T S^T C_S^{-1} SX$ is a random quadratic form with the symmetric matrix $B = S^T C_S^{-1} S$ which depends only on S and historical data $X_1,\ldots,X_N$. Using the moment generation functions method (see, e.g., Mathai and Provost, 1992), we can obtain $E\{X^T BX|B\} = m^T Bm + \mathrm{trace}[B\,\Sigma]$. The full form formula

$$m^T\,(S^T\,C_S^{-1}\,S)\,m + \mathrm{trace}[S^T\,C_S^{-1}\,S\,\Sigma] \tag{16}$$

consists of two parts. The first one, a quadratic form $m^T\,(S^T\,C_S^{-1}\,S)\,m$, points out the mean changes. Furthermore, we can also detect some changes in the covariance matrix Σ, because the last part of the summand is of the form $\mathrm{trace}[S^T\,C_S^{-1}\,S\,\Sigma]$.

We can use known properties of quadratic forms in multivariate normal variables (Mathai and Provost, 1992) and obtain the variance of T^2_S for $X \sim \mathcal{N}_d(m,\Sigma)$:

$$\begin{aligned} &\mathrm{var}_X\{X^T S^T ASX|S\} \\ &\quad = 2\mathrm{Trace}[S^T AS\Sigma S^T AS\Sigma] \\ &\qquad + 4m^T S^T AS\Sigma S^T ASm, \end{aligned} \tag{17}$$

where $A = C_S^{-1}$ is a symmetric, positive definite $k \times k$ matrix.

We shall discuss properties $m^T(S^T C_S^{-1} S)m$ and the eccentricity parameter of the Snedecor distribution of T^2_S in an out of control state, i.e., $(Sm)^T(S^T\Sigma S)^{-1}(Sm)$, because of their influence on the probability of detection of a change of the mean.

The classical Hotelling control chart (i.e., without projections) in an out-of-control state is usually investigated by simulation (Montgomery, 1996), as it is impossible to approximate with sufficient accuracy the distribution of non-central F-Snedecor statistics. The non-centrality parameter $m\Sigma^{-1}m$ varies as follows:

$$\frac{||m||^2}{e_{\min}(\Sigma)} \geq m^T\Sigma^{-1}m \geq \frac{||m||^2}{e_{\max}(\Sigma)},$$

where $e_{\max}(\Sigma)$ and $e_{\min}(\Sigma)$ stand for the largest and the smallest eigenvalue of matrix Σ, respectively (see, for example, Mathai and Provost, 1992). This means that the detectability of a change in the mean depends not only on the magnitude of the change, but also on the eigenstructure of the covariance matrix.

A similar effect is obtained for T_S^2. It is important to note that the covariance matrix SCS^T is the same in the in-control and in the out-of-control state. This matrix is very regular, since all its diagonal elements have the same distribution with the mean equal to $\mathrm{trace}[\Sigma]$. Similarly, all the out off-diagonal elements have the same zero-mean, finite variance distribution (see the also (13) and (14)). Thus, the largest eigenvalue of the covariance matrix slowly grows with its size k and the smallest eigenvalue of the covariance matrix slowly decreases with k (see appropriate Gumbel distributions (Forbes *et al.*, 2011)). Furthermore, for $k \ll d$, only a very small part of the true covariance structure Σ can be recovered from $S^T\Sigma S$. Summarizing, although the influence of $S^T\Sigma S$ on $\lambda^2 = (Sm)^T(S^T\Sigma S)^{-1}(Sm)$, exists, it cannot be controlled by increasing k, unless k is close to d.

The greater $||Sm||$ (or $||Sm||^2 = m^T S^T Sm$), the more "visible" the change of a mean represented by vector m after projections. Thus, the most important component affecting the loss in the accuracy is $||Sm||$. Notice that $||Sm||$ depends only on inner products between the projection vectors and m (see (2)). If Sm is very small, then the change in the mean m is very hard to detect, even if $||m||$ is large.

Recall that we have

$$E_S[m^T S^T Sm] = km^T m = k||m||^2.$$

Estimating the norm of a vector using random projections we have to use the factor $1/k$ (see (2)). So it is critical to guarantee that projection S does not reduce $1/\sqrt{k}\,||Sm||$ too much in comparison with $\|m\|$.

How large should the dimension k of the projection defined by S be? Obviously, k should be smaller than the number of learning samples N. Of course, it is impossible to use the Johnson–Lindenstrauss lemma to assure the low distortion of norms for every possible observation, but we can try to bound k with respect to the estimate $\|m\|^2$. If our aim is to obtain the general detectability of the mean change, we can only expect that the norm of a projection of vector m would not be too small, as has been previously explained in detail. By (2) it is only guaranteed that the norm of m will be approximately preserved after projection in the mean sense.

Using chi-squared tail Chernoff bounds (see (3) for comparisons), one can obtain the bound on k, which guarantees that the distortion, defined as $\frac{1}{k}||Sm||_k^2/||m||_d^2$, will be smaller than $(1-\varepsilon)$ with a prescribed probability $\delta > 0$, where $\varepsilon > 0$ determines the acceptable level of distortion. Namely, from

$$\Pr\left\{\frac{1}{k}||Sm||_k^2 < (1-\varepsilon)\,||m||_d^2\right\} \leq \exp\left(-\frac{k}{4}\varepsilon^2 + \frac{k}{6}\varepsilon^3\right) \quad (18)$$

one can conclude that, if

$$k > \frac{-4\ln\delta}{\varepsilon^2 - 2\varepsilon^3/3}, \quad (19)$$

then $\Pr[\frac{1}{k}||Sm||^2 < (1-\varepsilon)||m||_d^2]$ is smaller than $\delta > 0$. Note that this probability does not depend on vector m. Alternatively, one can obtain the bound on the accuracy level ε for a given k, i.e., $\varepsilon^2(1 - 2/3\varepsilon) > -4\ln\delta/k$.

For example, assuming that the acceptable distortion of $||Sm||^2$ is $\varepsilon = 0.2$ and the probability of distortion $\delta = 0.1$, we obtain that the number of projections k should be greater than 265.

The bound provided by (19) should be considered the bound for the worst case scenario, i.e., it is only an upper bound on the smallest lower bound $k_{\min}$ which is not known. As follows from other applications of random projections (see, e.g., Li *et al.*, 2006a; 2006a; Skubalska-Rafajłowicz, 2008), it usually suffices to use a smaller number of projections than suggested by (19). Our experiments, as presented in the next section, also validate such conclusions.

6. Computational examples

In this section we want to show that the method proposed in this paper can be used in practice even if the data dimensionality is really large. Although we do not concentrate on image processing problems, we can examine the random projection method using an artificial noisy image sequence as an example, when changes in the data are easy to be controlled visually and the dimensionality of the observed images, treated as vectors, can be huge. Three images $I0$, $I1$ and $I2$, of size 100×100 each (i.e., of dimensionality $d = 10000$), correspond to the in-control state and two different out-of-control states, respectively. A correlated Gaussian vector sequence was generated as a spatial autoregression process, without writing down explicitly and memorizing the corresponding covariance matrix, which would be too large to be processed efficiently by a computer. We have used normal random projections and sparse random projections, according to (6) with parameter $c = 3$.

6.1. Detection of changes in images with non-correlated noise. Assume that we observe a sequence of noise images with a cross in the center of an image. The normal image $I0$ consists of 10000 pixels, and each image

pixel is corrupted independently by a zero mean normal noise $\mathcal{N}(0, 0.1)$. $I0$ (the mean) and a random example of the noisy $I0$ rearranged as 100×100 arrays are depicted in Fig. 1.

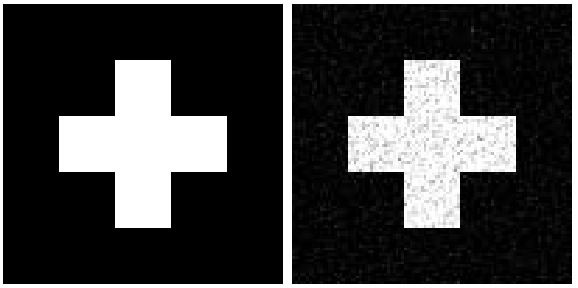

Fig. 1. Image $I0$, i.e., a 10000 dimensional vector rearranged as a 100×100 array and image $I0$ corrupted by an uncorrelated normal noise.

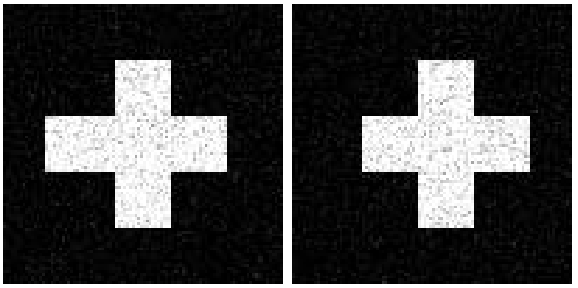

Fig. 2. Noise faulty images: $I1$ (left panel) and $I2$ (right panel).

Figure 2 shows two faulty images. Namely, in the first image $I1$ the left cross arm is two pixels longer; the second faulty image $I2$ contains the same cross as in the $I0$, but moved two pixels to the right.

Notice that for the non-correlated image noise $n_i \sim \mathcal{N}(0, 0.1)$, $i = 1, \ldots, 10000$, i.e., the noise with the individual variance 0.01 and a common covariance matrix $0.01I$, the sum $N^2_{\text{noise}} = \sum n_i^2 \sim 0.01\chi^2_{10000}$. Thus, the expected value of the squared vector noise length N^2_{noise} is $0.01d = 0.01 \cdot 10000 = 100$ and its variance equals $(0.01)^2 2d = (0.01)^2 20000 = 2$ (see Forbes *et al.*, 2011). The total variance of the image noise is 100 in this case. Furthermore, the squared distance between $I0$ and $I1$ treated as vectors is equal to 100 and the squared distance between $I0$ and $I2$ (treated as vectors) equals 600. Note that for "in-control" data we have assumed that the true mean equals a zero vector. Thus, all observations are translated adequately.

We can approximate the distribution of the random variable N^2_{noise} by the normal distribution $\mathcal{N}(100, \sqrt{2})$. Thus, with the significance level 0.01 we can assume that the value of the squared vector image length is less than 103.352. So, if we know that the noise is non-correlated (white Gaussian noise), it is very easy to distinguish between the images. In our experiments based on 1000 independent trials, every noisy image $I1$ and $I2$ produces the squared vector length greater than the limit 103.352. Thus, the change in the vector image mean (with respect to $I0$ treated as zero) was immediately detected. Notice that very similar results are obtained for random projections of dimensionality 100 (see Table 3), without any knowledge of the correlation matrix.

This means that the problem of detection is easy in this case, provide that the covariance matrix of the image noise is known. However, we can treat the problem as a test problem and assume that the noise covariance matrix is not known. So, we will examine the diagnostic power of the proposed projection method inferring about changes in the mean (changes of images) on the basis of projected noisy observations only, i.e., using the T_S^2 statistic for projections of dimension $k = 5,\ 10,$ and $k = 100$.

Figures 3–5 show T_S^2 statistic values using normal or sparse (Fig. 5) random projections of dimensionality $k = 5$ for 1000 independent observations of in-control data ($I0$) and out of control data ($I1$ and $I2$) corrupted by a Gaussian white noise, respectively. The mean for T_S^2 statistic in Fig. 3, i.e., for in-control data, is approximately equal to 5. The graphs for T_S^2 statistic in other dimensions look similar to the graph presented in Fig. 3, but their scale depends on k. Figure 6 illustrates the T_S^2 statistic for "out-of-control" data ($I1$) with a Gaussian white noise and sparse random projections of dimensionality $k = 100$.

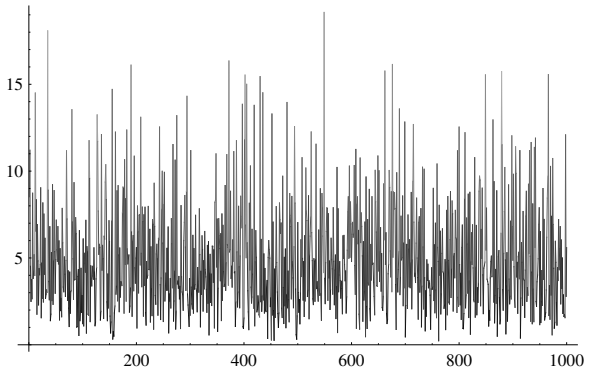

Fig. 3. T_S^2 statistic for in-control data ($I0$ corrupted by a Gaussian white noise) with normal random projections of dimensionality $k = 5$; 1000 independent observations.

Tables 1–3 show false alarm rates and the probability of changes detection for separate observations of an average of over 1000 test samples for images corrupted by Gaussian white noise with variance $\sigma^2 = 0.01$ and 10 different normal random projections and sparse random projections of dimensionality $k = 5$, $k = 10$ and $k = 100$, respectively. The average total variance of projected data, i.e., the averaged trace of estimated covariance matrices was 98.87 with standard deviation 2.44 and 99.67 with

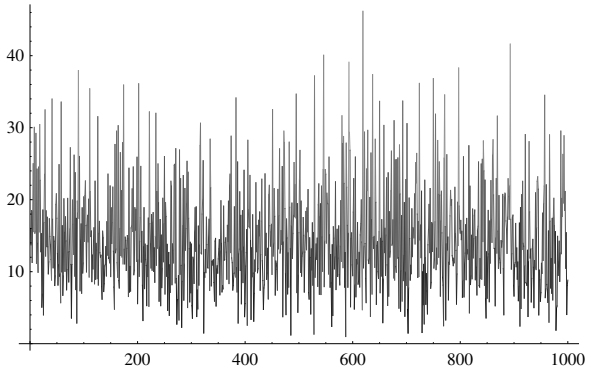

Fig. 4. T_S^2 statistic for out-of-control data (faulty image $I1$ corrupted by a Gaussian white noise) with normal random projections of dimensionality $k = 5$; 1000 independent observations.

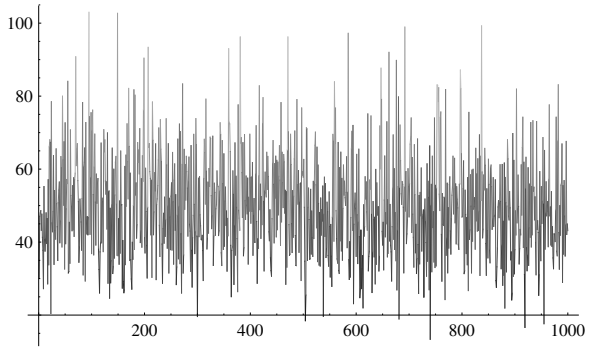

Fig. 5. T_S^2 statistic for out-of-control data (faulty image $I2$ corrupted by a Gaussian white noise) with sparse random projections of dimensionality $k = 5$; 1000 independent observations.

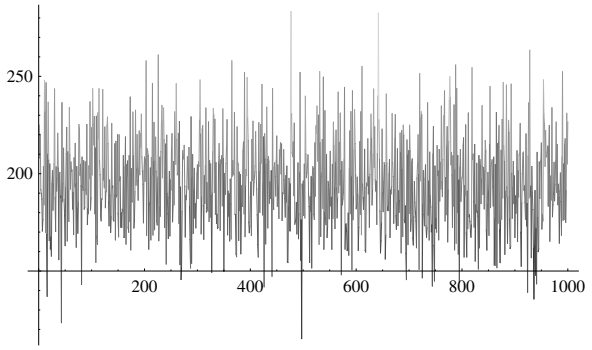

Fig. 6. T_S^2 statistic for out-of-control data (faulty image $I1$) with a Gaussian white noise and sparse random projections of dimensionality $k = 100$; 1000 independent observations.

standard deviation 1.69 for sparse and normal projections of dimensionality $k = 5$, respectively, 99.76 with standard deviation 0.322 and 100.03 with standard deviation 0.535 for sparse and normal projections of dimensionality $k = 100$, respectively.

The last rows of the average simulation results of the tables are also presented. Threshold values for T_S^2 statistic are chosen according to $\frac{kN}{d-k+1}G_{(k,N-k+1)}(1-\alpha)$ with $\alpha = 0.01$. Thus, we have assumed that the theoretical false alarm rate is equal to 0.01. The alarm level (upper control limit) for projection dimension $k = 100$ was selected as $h_{0.01} = 138.0$. This value is closer to the threshold obtained from χ^2_{100} distribution than that from the Snedecor one, because, due to numerical round-off errors, the variability of the practically estimated covariance matrix was smaller than the variability of the theoretical covariance estimate following the Wishart distribution. Table 3 demonstrates that even now the attained false alarm rate was about twice as small as the assumed probability 0.01. The probability of detection of faulty images using random projections of dimensionality $k = 100$ was close to 1.0 with the range from 0.982 to 1.000. But even when one uses only several projections, namely, for $k = 5$ or $k = 10$, the probability of mean change detection was relatively high. Detailed results are summarized in Tables 1–3.

6.2. Detection of changes in images corrupted by a correlated noise. The reference image is also $I0$, but the image noise is generated by a random dynamic process independently not only for each observation, but also for each image row, i.e.,

$$n_i = \varepsilon_i \sim \mathcal{N}(0,1), \quad i = 1 \text{mod}(100),$$

$$n_i = 0.8\varepsilon_{i-1} + \varepsilon_i,$$

$$\varepsilon_i \sim \mathcal{N}(0,0.1), \quad i = 2,\ldots,100 \text{mod}(100).$$

In such a case the noise covariance matrix is a tridiagonal matrix and consists of 100 identical segments (same for every image row) with variances equalling 1, 0.65 and 0.0164 (98 times in each diagonal segment). Thus, the total variance of the noise is 325.72 and is 3.25 times greater than in the non-correlated noise case. The image defects are as previously defined by $I1$ and $I2$, respectively.

Detailed simulation results are summarized in Tables 4, 5, 7, 8 and 6, 9 for sparse random projections and normal random projections, respectively. For correlated normal data and random projections of dimensionality 100, the probability of detecting the change from $I0$ to $I1$ was between 0.63 to 0.91, depending on a randomly chosen projection (see Tables 8 and 9). This means that (in the case of a permanent mean shift) the average time to change detection (average run length of the ARL chart) is smaller than 2. The average total variance of projected data, in comparison with the true total variance 325.72, equals 329.29 with a standard deviation of 18.04 for sparse projections of dimensionality $k = 5$, and 324.46 with standard deviation 2.87 for sparse projections of dimensionality $k = 100$, respectively.

Table 1. False alarm probability and probability of mean change detection averaged over 1000 test samples for images corrupted by white Gaussian noise with variance $\sigma^2 = 0.01$ and 10 different normal random projections and sparse random projections of dimensionality $k = 5$. The alarm level (upper control limit) $h_{0.01} = 15.16$.

Projection	Correct $I0$ false alarm rate	Probability of detection for $I1$	Probability of detection for $I2$
normal	0.004	0.308	0.970
sparse	0.006	0.392	1.000
normal	0.009	0.075	0.899
sparse	0.010	0.085	0.817
normal	0.009	0.016	1.000
sparse	0.011	0.067	0.333
normal	0.011	0.137	0.995
sparse	0.007	0.068	0.885
normal	0.009	0.125	0.987
sparse	0.006	0.358	0.553
normal	0.007	0.310	0.938
sparse	0.010	0.336	0.909
normal	0.007	0.141	0.436
sparse	0.015	0.104	0.996
normal	0.010	0.103	1.000
sparse	0.012	0.271	0.903
normal	0.012	0.029	0.952
sparse	0.010	0.519	0.228
normal	0.012	0.063	0.649
sparse	0.005	0.111	0.786
avg. normal	0.0090	0.1307	0.6826
avg. sparse	0.0092	0.2311	0.6410

Table 2. False alarm probability and probability of a mean change detection averaged over 1000 test samples for images corrupted by white Gaussian noise with variance $\sigma^2 = 0.01$ and 10 different normal random projections and sparse random projections of dimensionality $k = 10$. The alarm level $h_{0.01} = 23.21$.

Projection	Correct $I0$ false alarm rate	Probability of detection for $I2$	Probability of detection for $I3$
normal	0.010	0.589	1.000
sparse	0.014	0.369	0.997
normal	0.015	0.131	1.000
sparse	0.007	0.492	0.681
normal	0.009	0.500	0.902
sparse	0.011	0.188	0.999
normal	0.008	0.408	1.000
sparse	0.007	0.060	1.000
normal	0.012	0.713	0.999
sparse	0.013	0.084	0.998
normal	0.009	0.111	0.996
sparse	0.008	0.225	0.982
normal	0.007	0.073	0.991
sparse	0.008	0.455	0.915
normal	0.013	0.115	1.000
sparse	0.007	0.246	1.000
normal	0.008	0.062	1.000
sparse	0.003	0.468	1.000
normal	0.011	0.604	1.000
sparse	0.010	0.321	1.000
avg. normal	0.0102	0.2818	0.9968
avg. sparse	0.0087	0.2908	0.9572

Table 3. False alarm probability and probability of a mean change detection averaged over 1000 test samples for images corrupted by white Gaussian noise with variance $\sigma^2 = 0.01$ and 10 different normal random projections and sparse random projections of dimensionality $k = 100$, respectively. The alarm level $h_{0.01} = 138.0$.

Projection	Correct $I0$ false alarm rate	Probability of detection for $I1$	Probability of detection for $I2$
normal	0.002	1.000	1.000
sparse	0.008	0.999	1.000
normal	0.006	0.999	1.000
sparse	0.004	1.000	1.000
normal	0.005	1.000	1.000
sparse	0.006	0.982	1.000
normal	0.004	1.000	1.000
sparse	0.007	1.000	1.000
normal	0.005	1.000	1.000
sparse	0.001	1.000	1.000
normal	0.003	1.000	1.000
sparse	0.008	1.000	1.000
normal	0.004	0.997	1.000
sparse	0.002	0.999	1.000
normal	0.002	0.999	1.000
sparse	0.007	0.997	1.000
normal	0.007	1.000	1.000
sparse	0.006	0.998	1.000
normal	0.002	0.998	1.000
sparse	0.005	1.000	1.000
avg. normal	0.0040	0.9993	1.000
avg. sparse	0.0054	0.9971	1.000

Table 4. False alarm probability and probability of a mean change detection averaged over 1000 test samples for images corrupted by correlated Gaussian noise and 10 different sparse random projections of dimensionality $k = 5$.

Correct $I0$ false alarm rate	Probability of detection for $I1$	Probability of detection for $I2$
0.01	0.073	0.374
0.008	0.027	0.193
0.009	0.026	0.266
0.006	0.050	0.661
0.008	0.038	0.146
0.006	0.037	0.206
0.010	0.026	0.462
0.013	0.021	0.320
0.011	0.062	0.435
0.010	0.033	0.806
avg. 0.0091	avg. 0.0387	avg. 0.3867

6.3. Additional comments. It should be mentioned that one can easily calculate the moments of T_S^2 in the in-control state ($m = 0$ no change in the mean) and in the out-of-control state, as moments of a central and a non-central Snedecor distribution, respectively. The first moment of

$$\frac{kN}{N-k+1}F(k, N-k+1)$$

equals

$$\frac{k}{1-1/N-k/N}$$

and the first moment of

$$\frac{kN}{N-k+1}F(k, N-k+1, (Sm)^T(S\Sigma S^T)^{-1}Sm)$$

equals

$$\frac{k+(Sm)^T(S\Sigma S^T)^{-1}Sm}{1-1/N-k/N}.$$

In all simulations we have performed, the averaged values of the T_S^2 statistics were very close to the theoretical means.

Table 5. False alarm probability and probability of a mean change detection averaged over 1000 test samples for images corrupted by correlated Gaussian noise and 10 different sparse random projections of dimensionality $k = 10$.

Correct $I0$ false alarm rate	Probability of detection for $I1$	Probability of detection for $I2$
0.010	0.012	0.312
0.006	0.091	0.967
0.008	0.027	0.193
0.007	0.104	0.899
0.010	0.075	0.363
0.010	0.059	0.661
0.014	0.028	0.668
0.010	0.066	0.918
0.011	0.066	0.977
0.007	0.146	0.673
avg. 0.0097	avg. 0.0674	avg. 0.6651

Table 6. False alarm probability and probability of a mean change detection averaged over 1000 test samples for images corrupted by correlated Gaussian noise and 10 different normal random projections of dimensionality $k = 10$.

Correct $I0$ false alarm rate	Probability of detection for $I1$	Probability of detection for $I2$
0.009	0.218	0.511
0.013	0.025	0.780
0.009	0.037	0.824
0.008	0.033	0.617
0.009	0.023	0.840
0.010	0.023	0.730
0.004	0.057	0.565
0.010	0.057	0.672
0.005	0.040	0.509
0.004	0.041	0.653
avg. 0.0081	avg. 0.0554	avg. 0.6601

Table 7. False alarm probability and probability of a mean change detection averaged over 1000 test samples for images corrupted by correlated Gaussian noise and 10 different sparse random projections of dimensionality $k = 50$

Correct $I0$ false alarm rate	Probability of detection for $I1$	Probability of detection for $I2$
0.001	0.151	1.000
0.004	0.187	1.000
0.001	0.235	0.998
0.003	0.119	1.000
0.002	0.329	1.000
0.005	0.410	1.000
0.018	0.458	0.999
0.005	0.246	1.000
0.005	0.254	0.999
0.008	0.306	1.000
avg. 0.0052	avg. 0.2595	avg. 0.9996

7. Concluding remarks

We have proposed a new approach to fault analysis of high dimensional multi-variate and inter-correlated measurements based on dimensionality reduction via random projections. We assumed the normality of the data but, as in the practice of the statistical quality control area, this assumption can be weakened.

Random projections may be treated as weighted averages of multivariate observations. This approach is very simple, in contrast to a laborious, or even impossible in higher dimensions, analysis of data structures using PCA or ICA.

Table 8. False alarm probability and probability of a mean change detection averaged over 1000 test samples for images corrupted by correlated Gaussian noise and 10 different sparse random projections of dimensionality $k = 100$.

Correct $I0$ false alarm rate	Probability of detection for $I1$	Probability of detection for $I2$
0.005	0.810	1.000
0.005	0.862	1.000
0.007	0.806	1.000
0.005	0.886	1.000
0.009	0.914	1.000
0.004	0.797	1.000
0.005	0.654	1.000
0.006	0.696	1.000
0.003	0.717	1.000
0.006	0.889	1.000
avg. 0.0055	avg. 0.8031	avg. 1.000

Table 9. False alarm probability and probability of a mean change detection averaged over 1000 test samples for images corrupted by correlated Gaussian noise and 10 different normal random projections of dimensionality $k = 100$.

Correct $I0$ false alarm rate	Probability of detection for $I1$	Probability of detection for $I2$
0.002	0.717	1.000
0.009	0.686	1.000
0.008	0.630	1.000
0.011	0.698	1.000
0.006	0.675	1.000
0.005	0.844	1.000
0.001	0.780	1.000
0.005	0.678	1.000
0.005	0.670	1.000
0.003	0.764	1.000
avg. 0.0055	avg. 0.7142	avg. 1.000

Concentration properties used in the inequality (18) (see also the proof of the Johnson–Lindenstrauss lemma (Dasgupta and Gupta, 2003)) favor high-dimensional data being "the blessing of dimensionality" (Donoho, 2000).

Using random projections can be considered an implication of the Cramer-Wold theorem (Cuesta-Albertos *et al.*, 2007; Cramer and Wold, 1936). The theorem states that a Borel probability measure on $\mathbb{R}^d$, where $d \geq 2$, is uniquely determined by its one-dimensional projections.

This paper addresses the problem of how many projections are really needed to determine changes in the mean of the probability distribution, when only projections of observations are available.

Note that the method proposed here can be also used for a total variance change detection; however, the covariance change detection problem is even more complicated (also in low dimensions) (Montgomery, 1996) and is beyond the scope of this paper.

It should be mentioned that a different approach to projections of multivariate data is presented in the works by Runger (1996), Runger *et al.* (2007) as well as Bodnar and Schmid (2005), where projections are given in advance (latent variable model of the process) and simultaneously the dimensionality of the observation is relatively small.

Since Shewhart-type charts, including the multivariate Hotellings T^2 chart, are more sensitive to large shifts than to small shifts, in the case of small changes the T_S^2 chart should be combined with MEWMA or MCUSUM control charts (Sullivan and Woodall, 2000).

References

Achlioptas, D. (2001). Database friendly random projections, *Proceedings of the 20th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, Santa Barbara, CA, USA*, pp. 274–281.

Ailon, N. and Chazelle, B. (2006). Approximate nearest neighbors and the fast Johnson–Lindenstrauss transform, *Proceedings of the 38th Annual ACM Symposium on Theory of Computing, Seattle, WA, USA*, pp. 557–563.

Arriaga, R. and Vempala, S.(1999). An algorithmic theory of learning: Robust concepts and random projection, *Pro-*

ceedings of the 40th Annual IEEE Symposium on the Foundations of Computer Science, New York, NY, USA, pp. 616–623.

Biau, G. and Devroye, L. and Lugosi, G. (2008). On the performance of clustering in Hilbert spaces *IEEE Transactions on Information Theory* **54**(2): 781–790.

Bodnar, O. and Schmid, W. (2005). Multivariate control charts based on a projection approach *Allgemeines Statistisches Archiv* **89**(1): 75–93.

Chandola, V., Banerjee, A. and Kumar, V. (2009). Anomaly detection: A survey, *ACM Computing Surveys* **41**(3): 15:1–15:58.

Cramer, H. and Wold, H.(1936). Some theorems on distribution functions, *Journal of the London Mathematical Society* **11**(2): 290–295.

Cuesta-Albertos, J.A., del Barrio, E., Fraiman, R. and Matran, C. (2007). The random projection method in goodness of fit for functional data, *Computational Statistics and Data Analysis* **51**(10): 4814–4831.

Cuturi, M., Vert, J-P. and dAspremont, A. (2009). White functionals for anomaly detection in dynamical systems, *in* Y. Bengio, D. Schuurmans, J. Lafferty, C.K.I. Williams and A. Culotta (Eds.), *Advances in Neural Information Processing Systems*, Vol. 22, MIT Press, Vancouver, pp. 432–440.

Dasgupta, S. and Gupta, A. (2003). An elementary proof of a theorem of Johnson and Lindenstrauss, *Random Structures and Algorithms* **22**(1): 60–65.

Donoho D.L. (2000). High-dimensional data analysis: The curses and blessings of dimensionality, *Technical report*, Department of Statistics, Stanford University, Stanford, CA.

Frankl, P. and Maehara, H. (1987). The Johnson–Lindenstrauss lemma and the sphericity of some graphs, *Journal of Combinatorial Theory A* **44**(3): 355–362.

Forbes, C., Evans, M. and Hastings, N. and Peacock, B. (2011). *Statistical Distributions*, 4th Edn., John Wiley and Sons, Inc., Hoboken, NJ.

Hyvärinen, A., Karhunen, J. and Oja, E. (2001). *Independent Component Analysis,* Wiley, New York, NY.

Hotelling, H. (1931). The generalization of Student's ratio *The Annals of Mathematical Statistics* **2**(3): 360–378.

Indyk, P. and Motwani, R. (1998). Approximate nearest neighbors: Towards removing the curse of dimensionality, *Proceedings of the 30th Annual ACM Symposium on the Theory of Computing, Dallas, TX, USA*, pp. 604–613.

Indyk, P. and Naor, A.(2007). Nearest neighbor preserving embeddings, *ACM Transactions on Algorithms* **3**(3): 31:1–31:12.

Jolliffe, I.T. (1986). *Principal Component Analysis*, Springer-Verlag, New York, NY.

Johnson, W.B. and Lindenstrauss, J.(1984). Extensions of Lipschitz mapping into Hilbert space, *Contemporary Mathematics* **26**: 189–206.

Korbicz, J., Kościelny, J.M., Kowalczuk, Z. and Cholewa, W. (Eds.) (2004). *Fault Diagnosis. Models, Artificial Intelligence, Applications.* Springer Verlag, Berlin/Heidelberg/New York, NY.

Lee, J.A. and Verleysen, M. (2007). *Nonlinear Dimensionality Reduction,* Springer, New York, NY.

Li, P., Hastie, T.J. and Church, K.W. (2006a). Nonlinear estimators and tail bounds for dimension reduction in L1 using Cauchy random projections, *Technical report*, Department of Statistics, Stanford University, Stanford, CA.

Li, P., Hastie, T.J. and Church, K.W. (2006b). Sub-Gaussian random projections, *Technical report*, Department of Statistics, Stanford University, Stanford, CA.

Mason, R.L., Tracy, N.D. and Young, J.C., (1992). Multivariate control charts for individual observations, *Journal of Quality Technology* **24**(2): 88–95.

Mason, R.L. and Young, J.C. (2002). *Multivariate Statistical Process Control with Industrial Application*, SIAM, Philadelphia, PA.

Mathai, A.M. and Provost, S.B. (1992). *Quadratic Forms in Random Variables: Theory and Applications*, Marcel Dekker, New York, NY.

Matoušek, J.(2008). On variants of the Johnson–Lindenstrauss lemma, *Random Structures and Algorithms* **33**(2): 142–156.

Milman, V.(1971). A new proof of the theorem of A. Dvoretzky on sections of convex bodies, *Functional Analysis and Its Applications* **5**(4): 28–37, (English translation).

Montgomery, D.C. (1996). *Introduction to Statistical Quality Control*, 3rd Edn., John Wiley and Sons, New York, NY.

Qin, S.J.(2003). Statistical process monitoring: Basics and beyond *Journal of Chemometrics* **17**(8–9): 480–502.

Rao, C.R. (1973). *Linear Statistical Inference and Its Applications*, John Wiley and Sons, New York, NY/London/Sydney/Toronto.

Runger, G.C. (1996). Projections and the U-squared multivariate control chart, *Journal of Quality Technology* **28**(3): 313–319.

Runger, G., Barton, R., Del Castillo, E. and Woodall, W.H. (2007). Optimal monitoring of multivariate data for fault patterns, *Journal of Quality Technology* **39**(2): 159–172.

Skubalska-Rafajłowicz, E. (2006). RBF neural network for probability density function estimation and detecting changes in multivariate processes, *in* L. Rutkowski, R. Tadeusiewicz, L.A. Zadeh and J. Żurada (Eds.), *Artificial Intelligence and Soft Computing*, Lecture Notes in Computer Science, Vol. 4029, Springer-Verlag, Berlin/Heidelberg, pp. 133–141.

Skubalska-Rafajłowicz, E. (2008). Random projection RBF nets for multidimensional density estimation, *International Journal of Applied Mathematics and Computer Science* **18**(4): 455–464, DOI: 10.2478/v10006-008-0040-9.

Skubalska-Rafajłowicz, E. (2009). Neural networks with sigmoidal activation functions dimension reduction using normal random projection, *Nonlinear Analysis* **71**(12): e1255–e1263.

Skubalska-Rafajłowicz, E. (2011). Fast and efficient method of change detection in statistically monitored high-dimensional data streams, *Proceedings of the 10th International Science and Technology Conference on Diagnostics of Processes and Systems, Zamość, Poland*, pp. 256–260.

Srivastava, M.S. (2009). A review of multivariate theory for high dimensional data with fewer observations, *in* A. SenGupta (Ed.), *Advances in Multivariate Statistical Methods*, Vol. 9, World Scientific, Singapore, pp. 25–52.

Sulliva, J.H. and Woodall, W.H. (2000). Change-point detection of mean vector or covariance matrix shifts using multivariate individual observations, *IIE Transactions* **32**(6): 537–549.

Tsung F. and Wang K. (2010). Adaptive charting techniques: Literature review and extensions, *in* H.-J. Lenz, P.-T. Wilrich and W. Schmid (Eds.), *Frontiers in Statistical Quality Control*, Vol. 9, Springer-Verlag, Berlin/Heidelberg, pp. 19–35.

Vempala, S. (2004). *The Random Projection Method*, American Mathematical Society, Providence, RI.

Wang, K. and Jiang, W. (2009). High-dimensional process monitoring and fault isolation via variable selection, *Journal of Quality Technology* **41**(3): 247–258.

Wang, J. (2012). *Geometric Structure of High-Dimensional Data and Dimensionality Reduction*, Higher Education Press, Beijing/Springer-Verlag, Berlin/Heidelberg.

Wold, H. (1966). Estimation of principal components and related models by iterative least squares *in* P. Krishnaiaah (Ed.), *Multivariate Analysis,* Academic Press, New York, NY, pp. 391–420.

Zorriassatine, F., Tannock, J.D.T. and O'Brien, C. (2003). Using novelty detection to identify abnormalities caused by mean shifts in bivariate processes, *Computers and Industrial Engineering* **44**(3): 385–408.

Ewa Skubalska-Rafajłowicz, Ph.D. in 1981 and D.Sc. in 2001 from the Wrocław University of Technology. Her fields of research include neural networks, pattern and image recognition and control charts, and use of space-filling curves and random projections for the dimensionality reduction. She has published almost 100 research papers on the above topics of expertise including those in *IEEE Transactions on Neural Networks*, *IEEE Transactions on Evolutionary Computing*, and the *International Journal of Applied and Mathematics and Computer Science*. She is also a member of the editorial board of *ISRN Applied Mathematics*. She has been invited as a visiting professor to Concordia University and Manitoba University in Canada as well as to Free University Berlin-West, Tübingen University and Otto-von-Guerricke University in Magdeburg.

Received: 29 March 2012
Revised: 30 October 2012
Re-revised: 2 April 2013