British Journal of Mathematics & Computer Science
21(1): 1-11, 2017; Article no.BJMCS.31778
ISSN: 2231-0851

SCIENCEDOMAIN *international*
www.sciencedomain.org

Large Matrices Inversion Using the Basis Exchange Algorithm

Leon Bobrowski[1,2*]

[1]*Faculty of Computer Science, Białystok University of Technology, Poland.*
[2]*Institute of Biocybernetics and Biomedical Engineering, PAS, Warsaw, Poland.*

Author's contribution

The sole author designed, analyzed and interpreted and prepared the manuscript.

Article Information

DOI: 10.9734/BJMCS/2017/31778
Editor(s):
(1) Sergio Serrano, Department of Applied Mathematics, University of Zaragoza, Spain.
(2) Metin Basarir, Department of Mathematics, Sakarya University, Turkey.
(3) Paul Bracken, Department of Mathematics, The University of Texas-Pan American, Edinburg, TX 78539, USA.
Reviewers:
(1) Marija Milojevic Jevric, Mathamatical Institute SANU, Serbia.
(2) Carlo De Gaetani, Politecnico di Milano, Italy.
(3) Nikolaos Ploskas, Carnegie Mellon University, Pittsburgh, PA, USA.
Complete Peer review History: http://www.sciencedomain.org/review-history/18203

Original Research Article

Received: 24th January 2017
Accepted: 4th March 2017
Published: 15th March 2017

Abstract

Data exploration tasks often require inversion of large matrices. The paper presents a new method of matrices inversion, which uses the basis exchange algorithm controlled by the convex and piecewise linear (*CPL*) inversion criterion function. Using basis exchange algorithms might increase the dimension of the inverted matrices and computational efficiency of the inversion tasks. Basis exchange algorithms are based on the Gauss-Jordan transformation which is used e.g. in the famous *Simplex* algorithm applied in linear programming.

Keywords: *Data exploration; large matrices inversion; basis exchange algorithm; Gauss-Jordan transformation; convex and piecewise linear (CPL) criterion functions.*

1 Introduction

The number of large data sets is increasing rapidly at the present time. Such data sets are being transformed and explored in many ways for extracting useful information which are then used for decision support systems or in prognostic (regression) models [1,2]. Fisher's discriminant analysis is one of the fundamental methods used in the decision support systems [3]. Finding the Fisher's solution involves the inversion of the

---

**Corresponding author: E-mail: l.bobrowski@pb.edu.pl;*

covariance matrix. The inversion of the covariance matrix becomes impossible by contemporary computational procedures when the dimension of data vectors is too large. The classical regression model is also based on the data matrix inversion and such model cannot be calculated when the dimensionality of data vectors exceeds a certain limit [3].

New numerical techniques aimed at challenging inverting of large matrices are currently being developed [4,5,6]. The basis exchange algorithms may also be examined in this context [7,8,9]. The basis exchange algorithms are based on the Gauss-Jordan transformation [10]. The famous *Simplex* algorithm used in the linear programming is also based on this transformation [10]. There are many examples of linear programming problems with a great practical importance which were based on large data sets and have been solved by the *Simplex* algorithm.

The basis exchange algorithm specified directly to the task of matrices inversion is proposed and theroetically examined in the presented paper. The proposed here algorithm is controlled by the convex and piecewise linear (*CPL*) inversion criterion function which is presented and analyzed in the paper. The presented paper contains, among others, the proof of the fundamental theorem about conditions when the minimum of the *CPL* inversion criterion function becomes equal to zero.

According to our research hypothesis, the proposed method of matirces inversion should allow to increase the dimensions of the inverted matrices. It is also expected that computational efficiency of a new procedure of matrices inversion will be high. The proposed procedure of matrices inversion should also be useful in other tasks of exploratory analysis of large, high-dimensional data sets.

2 Matrices Inversion in Discriminant Analysis

Let us assume that m objects O_j (j = 1,….,m) are represented by the n-dimensional feature vectors $\mathbf{x}_j = [x_{j,1},...,x_{j,n}]^T$, or as points in the n-dimensional feature space $F[n]$ ($\mathbf{x}_j \in F[n]$). Components $x_{j,i}$ of the feature vector $\mathbf{x}_j$ represent numerical results of n measurements of different features x_i (i = 1,….,n) of the j-th object O_j ($x_{j,i} \in \{0,1\}$ or $x_{j,i} \in R$).

We assume that the feature vectors $\mathbf{x}_j$ (j = 1,......, m) have been divided into two learning sets C_1 and C_2 labelled in accordance with the objects O_j *category* (*class*) ω_k (k =1, 2). The learning set C_k contains m_k feature vectors $\mathbf{x}_j(k)$ assigned to the k-th category ω_k, where $m = m_1 + m_2$:

$$C_k = \{\mathbf{x}_j(k)\} \quad (j \in I_k) \tag{1}$$

Each learning set C_k can be characterized by the mean vector $\boldsymbol{\mu}_k$ and the covariance matrix $\boldsymbol{\Sigma}_k$:

$$\boldsymbol{\mu}_k = \Sigma_j \, \mathbf{x}_j(k) \,/\, m_k \tag{2}$$

and

$$\boldsymbol{\Sigma}_k = \Sigma_j \, (\mathbf{x}_j(k) - \boldsymbol{\mu}_k)\,(\mathbf{x}_j(k) - \boldsymbol{\mu}_k)^T / (m_k - 1) \tag{3}$$

In accordance with the Fisher′s discriminant analysis, the feature vectors $\mathbf{x}_i(k)$ from the learning set C_k (1) are projected on the line $\boldsymbol{l}(\mathbf{w})$ in the feature space $F[n]$ ($\mathbf{x} \in F[n]$) defined by the parameter vector $\mathbf{w} = [w_1,...,w_n]^T$ ($\mathbf{w} \in R^n$) of the unit lenght ($\mathbf{w}^T\mathbf{w} = 1$):

$$\boldsymbol{l}(\mathbf{w}) = \{\mathbf{x}: \mathbf{x} = t\,\mathbf{w}, \textit{ where } t \in R\} \tag{4}$$

The feature vectors $\mathbf{x}_i(k)$ from the learning set C_k (1) are projected on the points $x_i(k) = \mathbf{w}^T\mathbf{x}_j(k)$ of the line $\boldsymbol{l}(\mathbf{w})$ (4). Similarly, the mean vector $\boldsymbol{\mu}_k$ (2) is projected on the point μ_k of the line $\boldsymbol{l}(\mathbf{w})$ (4):

$$\mu_k(\mathbf{w}) = \mathbf{w}^T\boldsymbol{\mu}_k \quad (5)$$

As a result of the feature vectors $\mathbf{x}_j(k)$ projection on the line $\boldsymbol{l}(\mathbf{w})$ (4), the covariance matrix $\boldsymbol{\Sigma}_k$ (3) can be replaced by the variance $\sigma_k^2(\mathbf{w})$ of the points $x_j(k) = \mathbf{w}^T\mathbf{x}_j(k)$ projected on the line $\boldsymbol{l}(\mathbf{w})$ (4):

$$\sigma_k^2(\mathbf{w}) = \Sigma_j (x_j(k) - \mu_k)^2 / (m_k - 1) \quad (6)$$

The Fisher′s criterion used for the discriminative vector $\mathbf{w}$ (4) choice can be formulated in the below manner [2]:

$$F(\mathbf{w}) = (\mu_1(\mathbf{w}) - \mu_2(\mathbf{w}))^2 / (\sigma_1^2(\mathbf{w}) + \sigma_2^2(\mathbf{w})) \rightarrow max \quad (7)$$

In accordance with the Fisher′s criterion, the vector $\mathbf{w}$ defining the line $\boldsymbol{l}(\mathbf{w})$ (4) should be selected in such a way, that the distance $|\mu_1(\mathbf{w}) - \mu_2(\mathbf{w})|$ between the mean values $\mu_k(\mathbf{w})$ (5) is as large as possible while the sum $\sigma_1^2(\mathbf{w}) + \sigma_2^2(\mathbf{w})$ of the variances $\sigma_k^2(\mathbf{w})$ (6) is small.

The vector $\mathbf{w}_F$ defining the maximal value of the Fisher′s criterion function $F(\mathbf{w})$ (7) can be given in the below manner [3]:

$$\mathbf{w}_F = \boldsymbol{\Sigma}_p^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \quad (8)$$

where $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are the mean vectors (2) and $\boldsymbol{\Sigma}_p$ is the pooled covariance matrix (3) [2]:

$$\boldsymbol{\Sigma}_p = ((m_1 - 1)\, \boldsymbol{\Sigma}_1 + (m_2 - 1)\, \boldsymbol{\Sigma}_2) / (m_1 + m_2 - 2) \quad (9)$$

The dimensionality $n * n$ of the pooled covariance matrix $\boldsymbol{\Sigma}_p$ (9) is defined by the dimension n of the feature space $F[n]$. The inversion of the matrix $\boldsymbol{\Sigma}_p$ (9) becomes difficult or even impossible in highly dimensional feature spaces $F[n]$.

3 Matrices Inversion in the Classical Regression Model

Multivariate regression model can be based on the linear (affine) transformation of the n - dimensional feature vectors $\mathbf{x}_j = [x_{j,1},...,x_{j,n}]^T$ ($\mathbf{x}_j \in F[n]$) on the points $t_j^\wedge$ of the below line ($t_j^\wedge \in R^1$) [2]:*

$$(\forall j \in \{1,\dots,m\}) \quad t_j^\wedge = t(\mathbf{x}_j) = \mathbf{w}^T\mathbf{x}_j + w_0 \quad (10)$$

where $\mathbf{w} = [w_1,...,w_n]^T \in R^n$ and $w_0 \in R^1$.

Properties of the model (10) depend on the vector of parameters (*weights*) $\mathbf{w}$ and the *threshold* w_0. The weights w_i and the threshold w_0 are estimated from regression learning sets C_r. The regression learning sets C_r can have the below structure [3]:

$$C_r = \{\mathbf{x}_j; t_j\} = \{x_{j,1},\dots, x_{j,n},; t_j\}, \ where \ j = 1,\dots, m_r \quad (11)$$

We can assume here that each of m_r objects O_j is characterized in the above set C_r by values $x_{j,i}$ of n *independent* variables (*features*) x_i, and by the observed value t_j ($t_j \in R^1$) of the *dependent* variable t.

In case of the classical regression the parameters $\mathbf{w}$ and w_0 of the model (10) are estimated on the base of the learning set C_r (11) in accordance with the *last squares method* [2]. In this approach the sum of the squared differences $(t_j - t_j^\wedge)^2$ between the observed target variable t_j and the modelled variable $t_j^\wedge$ (10) is minimized. The optimal solution $\mathbf{v}^* = [-w_0^*, (\mathbf{w}^*)^T]^T$ of such minimization problem can be given as [2]:

$$\mathbf{v}^* = (\boldsymbol{Y}^{\mathrm{T}} \boldsymbol{Y})^{-1} \boldsymbol{Y}^{\mathrm{T}} \mathbf{t} \tag{12}$$

where the matrix $\boldsymbol{Y}$ is constituted by m_r augmented feature vectors $\mathbf{y}_j = [1, \mathbf{x}_j^{\mathrm{T}}]^{\mathrm{T}}$ of the dimensionality $n + 1$:

$$\boldsymbol{Y}^{\mathrm{T}} = [\mathbf{y}_1,\ldots, \mathbf{y}_{mr}] \tag{13}$$

and (11)

$$\mathbf{t} = [t_1,\ldots, t_{mr}]^{\mathrm{T}} \tag{14}$$

Computing (12) the optimal vector $\mathbf{v}^*$ (12) includes the inversion of the matrix $\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y}$ which has the dimension equal to $(n+1)$ $(n+1)$.

Similarly to the pooled covariance matrix $\boldsymbol{\Sigma}_p$ (9), the inversion of the matrix $\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y}$ (12) becomes difficult or even impossible in the highly dimensional feature space $F[n]$. The maximum size of the reversible matrix $\boldsymbol{\Sigma}_p$ (9) or $\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y}$ (12) depends on the used reversing method and their implementation. The use of the basis exchange algorithms gives a chance for inreasing the size of the such matrices $\boldsymbol{\Sigma}_p$ (9) or $\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y}$ (12) which can be computationally inverted in practice.

4 The Basis Exchange Algorithms

The basis exchange algorithm was initially proposed and developed as an efficient tool for designing linear classifiers and examining linear separability of large, multidimensional data sets. The first version of the basis exchange algorithm was described with details in the papers [7] and [8]. Originally, the basis exchange algorithms aimed at an efficient minimization of the perceptron criterion function. The convex and piecewise linear (*CPL*) perceptron criterion function links the linear separability concept, which is fundamental in the theory of neural networks [10,11,3]. Variety of the *CPL* criterion functions were proposed later and used for controlling different types of basis exchange algorithms [9].

The k-th basis $\mathbf{B}_k$ is the squared, non-singular matrix with the n rows $\mathbf{b}_i(k)$:

$$\mathbf{B}_k = [\mathbf{b}_1(k),\ldots, \mathbf{b}_n(k)]^{\mathrm{T}} \tag{15}$$

The dimension of each vector $\mathbf{b}_i(k)$ is equal to n.

The inverse matrix $\mathbf{B}_k^{1}$ during the k-th stage can be represented in the below manner:

$$\mathbf{B}_k^{-1} = [\mathbf{r}_1(k),\ldots, \mathbf{r}_n(k)] \tag{16}$$

The vectors $\mathbf{b}_i(k)$ and $\mathbf{r}_{i'}(k)$ fulfil the below equations:

$$(\forall i,i' \in \{1,\ldots,n\}) \quad \mathbf{b}_i(k)^{\mathrm{T}}\mathbf{r}_i(k) = 1, \textit{ and} \\ \textit{if } i' \neq i, \textit{ then } \mathbf{b}_i(k)^{\mathrm{T}}\mathbf{r}_{i'}(k) = 0 \tag{17}$$

During the k-th stage of the the basis $\mathbf{B}_k$ (15) is changed into the basis $\mathbf{B}_{k+1}$. The matrix $\mathbf{B}_{k+1}$ is created from the matrix $\mathbf{B}_k$ (15) through replacing the l-th row $\mathbf{b}_l(k)$ by the new vector $\mathbf{z}_k$ taken from a given data matrix $\boldsymbol{Z}$ which contains m' vectors $\mathbf{z}_j$ with the dimension equal to n:

$$\boldsymbol{Z} = \{\mathbf{z}_j: j = 1,\ldots, m'\} \tag{18}$$

The exchange of the l-th basis vector $\mathbf{b}_1(k)$ (15) on the entry vector $\mathbf{z}_k$ results in the new basis $\mathbf{B}_{k+1}$ (15) and the new inverse matrix $\mathbf{B}_{k+1}^{-1} = [\mathbf{r}_1(k+1),\ldots, \mathbf{r}_n(k+1)]$ (16). The Gauss-Jordan transformation allows to compute efficiently the columns $\mathbf{r}_i(k+1)$ of the matrix $\mathbf{B}_{k+1}^{-1}$ on the basis of the columns $\mathbf{r}_i(k)$ (16) of te inverse matrix $\mathbf{B}_k^{-1}$, where $k = 0, 1,..., K$ ([7], [8]):

$$\mathbf{r}_l(k+1) = (1 / \mathbf{r}_l(k)^T\mathbf{z}_{j(k)})\ \mathbf{r}_l(k) \tag{19}$$

and

$$(\forall i \neq l) \quad \mathbf{r}_i(k+1) = \mathbf{r}_i(k) - (\mathbf{r}_i(k)^T\mathbf{z}_{j(k)})\ \mathbf{r}_l(k+1) = \\ = \mathbf{r}_i(k) - (\mathbf{r}_i(k)^T\mathbf{z}_{j(k)} / \mathbf{r}_l(k)^T\mathbf{z}_{j(k)})\ \mathbf{r}_l(k) \tag{20}$$

The basis exchange algorithms are based on the Gauss-Jordan transformation (19), (20). The index l of the vector $\mathbf{b}_l(k)$ which is removed from the basis $\mathbf{B}_k$ (15) during the k-th stage is determined by the *exit criterion* of a particular basis exchange algorithm. The *entry criterion* determines which vector $\mathbf{z}_{j(k)}$ from the set $\boldsymbol{Z}$ (18) enters the new basis $\mathbf{B}_{k+1}$ (15). The *stop criterion* allows to determine the final stage K of the basis exchange algorithm.

The basis exchange algorithms can be controlled by various criterion functions belonging to the family of the *convex and piecewise linear* (*CPL*) criterion functions. The *CPL* criterion functions allow defining different goals for the basis exchange algorithms. The *CPL* criterion functions define the *exit criterion*, the *entry criterion* and the *stop criterion* of the basis exchange algorithms. These criterions are chosen in a way that ensures decreasing of the criterion function during each stage k of the algorithm.

The basis exchange algorithms allow to generate sequences of square, non-singular matrices (*bases*) $\mathbf{B}_k$ ($k = 1,\dots, K$) in accordance with the Gauss-Jordan transformation [10].

Remark 1: The vector $\mathbf{z}_{j(k)}$ cannot enter the new basis $\mathbf{B}_{k+1}$ (15) if the below condition is met:

$$\mathbf{r}_l(k)^T\mathbf{z}_{j(k)} = 0 \tag{21}$$

The above statement results directly from the Gauss-Jordan transformation. There should be no division by the zero in the equation (19). The condition (21) has also an interesting geometric interpretation as the move in the parameter space along the parallel hyperplane $h_{j(k)} = \{\mathbf{w}: \mathbf{z}_{j(k)}^T\mathbf{w} = 1\}$ [9].

5 Matrices Inversion through Basis Exchange

The Gauss-Jordan transformation (19), (20) can be used in the multistage procedure aimed at the inversion of the squared data

$$\boldsymbol{Z} = [\mathbf{z}_1,\dots, \mathbf{z}_n] \tag{22}$$

The squared data matrix $\boldsymbol{Z}$ is composed of the n vectors $\mathbf{z}_j$ of the dimensionality n.

The vectors $\mathbf{z}_i$ (22) are equal to the columns of the pooled covariance matrix $\boldsymbol{\Sigma}_p$ (9), if this matrix is expected to be inversed for the Fisher's solution $\mathbf{w}_F$ (8) used in the discriminant analysis. In the classical regression model (12), the matrix $\boldsymbol{Y}^T\boldsymbol{Y}$ (12) should be inversed. In this case, the matrix $\boldsymbol{Z}$ (22) is composed of the $n + 1$ vectors $\mathbf{z}_j$ of the dimensionality $n + 1$ and the vectors $\mathbf{z}_i$ (22) are equal to the columns of the matrix $\boldsymbol{Y}^T\boldsymbol{Y}$ (12).

The proposed multistage procedure of the matrix $\boldsymbol{Z}$ (22) inversion begins ($k = 0$) with the matrices $\mathbf{B}_0$ (15) and $\mathbf{B}_0^{-1}$ (16) which are equal to the unit matrix $\boldsymbol{I} = [\mathbf{e}_1,\dots, \mathbf{e}_n]$:

$$\mathbf{B}_0 = \mathbf{B}_0^{-1} = \boldsymbol{I} = [\mathbf{e}_1,\dots, \mathbf{e}_n] \tag{23}$$

In this case, the vectors $\mathbf{r}_i(0)$ (16) are equal to the unit vectors $\mathbf{e}_i$ ($(\forall i \in \{1,...,n\})$ $\mathbf{r}_i(0) = \mathbf{e}_i$). The vectors $\mathbf{r}_i(k)$ (16) are transformed in accordance with the dependencies (19) and (20) in successive stages k ($k = 0, 1,..., K$) on the basis of the indices $l(k)$ and $j(k)$:

$$(l(0), j(0)), (l(1), j(1)),\ldots, (l(K), j(K)) \tag{24}$$

The index $l(k)$ of the vector $\mathbf{b}_{l(k)}$ leaving the basis $\mathbf{B}_k$ (15) during the k-th stage should be determined by the exit criterion. The index $j(k)$ of the vector $\mathbf{z}_{j(k)}$ (21) entering the basis $\mathbf{B}_{k+1}$ (16) should be determined by the entry criterion. The vector $\mathbf{z}_{j(k)}$ (21) which enters the basis replaces the vector $\mathbf{b}_{l(k)}$ and constitutes the $l(k)$-th row of the matrix $\mathbf{B}_{k+1}$ (16). The stop criterion determines the final stage K.

During the multistage procedure of the matrix $\boldsymbol{Z}$ (21) inversion all the unit vectors $\mathbf{e}_i$ in the matrix $\mathbf{B}_0$ (22) are expected to be replaced by the vectors $\mathbf{z}_j$ (21).

Remark 2: The multistage procedure of the data matrix $\boldsymbol{Z}$ (22) inversion succeeds if and only if each unit vectors $\mathbf{e}_i$ in the matrix $\mathbf{B}_0$ (23) is replaced by some vector $\mathbf{z}_{j'}$ (22). In this case, the matrix $\mathbf{B}_K^{-1}$ (16) obtained after the K stages of the basis exchange is equal to the inverse matrix $\boldsymbol{Z}^{-1}$ (22):

$$\boldsymbol{Z}^{-1} = \mathbf{B}_K^{-1} \tag{25}$$

During the multistage procedure of the basis $\mathbf{B}_k$ (15) transformations not every exchange of the vector $\mathbf{e}_i$ (22) on the vector $\mathbf{z}_{j'}$ (21) is feasible (*Remark* 1). Generally, the vector $\mathbf{z}_{i'}$ (22) cannot be entered into the basis $\mathbf{B}_k$ (16), if $\mathbf{z}_{j'}$ is a linear combination of such vectors $\mathbf{z}_j$, which were introduced earlier into the basis $\mathbf{B}_k$ [9].

6 The Inversion Criterion Function

The convex and piecewise linear (*CPL*) collinearity criterion functions have been defined recently and used for the purpose of extraction of collinear patterns from a high dimensional data set [10]. Similar *CPL* criterion function could be useful also in the task of large matrices inversion. Let us define for this purpose the below *CPL* penalty functions $\varphi_j(\mathbf{w})$ on the basis of the n vectors $\mathbf{z}_j$ (21) of the dimensionality n [12]:

$$(\forall \mathbf{z}_j \in \boldsymbol{Z}\ (22)) \tag{26}$$

$$\varphi_j(\mathbf{w}) = |1 - \mathbf{z}_j^T\mathbf{w}| = \begin{cases} 1 - \mathbf{z}_j^T\mathbf{w} & if \quad \mathbf{z}_j^T\mathbf{w} \le 1 \\ \mathbf{z}_j^T\mathbf{w} - 1 & if \quad \mathbf{z}_j^T\mathbf{w} > 1 \end{cases}$$

where $\mathbf{w} = [w_1,...,w_n]^T$ is the weight vector ($\mathbf{w} \in R^n$).

The inversion criterion function $\Phi_{inv}(\mathbf{w})$ is defined here as the sum of the *CPL* penalty functions $\varphi_j(\mathbf{w})$ (26) determined by the n vectors $\mathbf{z}_j$ from the squared matrix $\boldsymbol{Z}$ (22):

$$\Phi_{inv}(\mathbf{w}) = \sum_{j = 1,\ldots,n} \varphi_j(\mathbf{w}) \tag{27}$$

The inversion criterion function $\Phi_{inv}(\mathbf{w})$ is convex and piecewise linear (*CPL*) as the sum of the *CPL* penalty functions $\varphi_i(\mathbf{w})$ (26). The minimal value $\Phi_{inv}(\mathbf{w}^*)$ of the criterion function $\Phi_{inv}(\mathbf{w})$ (27) can be found efficiently by using the basis exchange algorithm [8]:

$$(\forall \mathbf{w}) \quad \Phi_{inv}(\mathbf{w}) \ge \Phi_{inv}(\mathbf{w}^*) = \Phi_{inv}^* \ge 0 \tag{28}$$

The minimal value $\Phi_{inv}(\mathbf{w}^*)$ of the criterion function $\Phi_{inv}(\mathbf{w})$ (27) is useful in the process of the matrix $\boldsymbol{Z}$ (22) inversion.

In order to analyse properties of the inversion criterion function $\Phi_{inv}(\mathbf{w})$ (27) minimization the two types of the dual hyperplanes h_i^1 and h_i^0 in the *n*-dimensional parameter space R^n are introduced [10]. Each vector $\mathbf{z}_j$ (22) allows to define the below dual hyperplane h_j^1:

$$(\forall j \in \{1,\ldots,n\}) \quad h_i^1 = \{\mathbf{w}: \mathbf{z}_i^T\mathbf{w} = 1\} \tag{29}$$

Similarly, each of n unit vectors $\mathbf{e}_i = [0,\ldots,1,\ldots,0]^T$ defines the below hyperplane h_i^0:

$$(\forall i \in \{1,\ldots,n\}) \quad h_i^0 = \{\mathbf{w}: \mathbf{e}_i^T\mathbf{w} = 0\} = \{\mathbf{w}: w_i = 0\} \tag{30}$$

where $\mathbf{w} = [w_1,...,w_n]^T \in R^n$.

Let us consider the *k*-th subset $\boldsymbol{S}_k$ of n linearly independent vectors $\mathbf{z}_j$ (22) and $\mathbf{e}_i$ (30):

$$\boldsymbol{S}_k = \{\mathbf{z}_j: j \in J_k\} \cup \{\mathbf{e}_i: i \in I_k\} \tag{31}$$

The set $\boldsymbol{S}_k$ is composed of r_k feature vectors $\mathbf{z}_j$ ($j \in J_k$) and $n - r_k$ unit vectors $\mathbf{e}_i$ ($i \in I_k$).

The intersection point of the r_k hyperplanes h_i^1 (29) defined by the vectors $\mathbf{z}_i$ ($j \in J_k$) and the $n - r_k$ hyperplanes h_i^0 (30) defined by the unit vectors $\mathbf{e}_i$ ($i \in I_k$) from the subset S_k (31) is called the *k*-th *vertex* $\mathbf{w}_k$ in the parameter space R^n. The below linear equations can be linked to the vertex $\mathbf{w}_k$:

$$(\forall j \in J_k) \quad \mathbf{w}_k^T\mathbf{z}_j = 1 \tag{32}$$

and

$$(\forall i \in I_k) \quad \mathbf{w}_k^T\mathbf{e}_i = 0 \tag{33}$$

The equations (32) and (33) can be represented in the matrix form:

$$\mathbf{B}_k \mathbf{w}_k = \mathbf{1}' = [1,\ldots,1,0,\ldots,0]^T \tag{34}$$

where the square, nonsingular matrix $\mathbf{B}_k$ is the *k*-th *basis* linked to the vertex $\mathbf{w}_k$:

$$\mathbf{B}_k = [\mathbf{z}_{j(1)},\ldots,\mathbf{z}_{j(rk)},\mathbf{e}_{i(rk+1)},\ldots,\mathbf{e}_{i(n)}]^T \tag{35}$$

and (16)

$$\mathbf{w}_k = \mathbf{B}_k^{-1}\,\mathbf{1}' = \mathbf{r}_1(k) + \ldots + \mathbf{r}_{rk}(k) \tag{36}$$

It can be proved that the minimal value Φ_{inv}^* (28) of the convex and piecewise linear criterion (*CPL*) function $\Phi_{inv}(\mathbf{w})$ (27) can be found in one of the vertices $\mathbf{w}_k$ (36) [9,10]:

$$(\exists \mathbf{w}_k^*)\ (\forall \mathbf{w})\ \Phi_{inv}(\mathbf{w}) \geq \Phi_{inv}(\mathbf{w}_k^*) = \Phi_{inv}^* \geq 0 \tag{37}$$

The basis exchange algorithms allow to find efficiently the minimal value $\Phi_{inv}(\mathbf{w}_k^*)$ of the *CPL* criterion functions $\Phi_{inv}(\mathbf{w})$ (27) even in the case of high dimensional matrices $\boldsymbol{Z}$ (22).

The inversion criterion function $\Phi_{inv}(\mathbf{w})$ (27) allows to precise the exit criterion, the entry criterion and the stop criterion according to the selected strategy of the this function minimization. These criteria should be chosen in a such manner that the function $\Phi_{inv}(\mathbf{w})$ (27) decreases maximally during each learning step k. The *steepest descent strategy* is often used in the tasks of the *CPL* criterion functions minimization [9].

7 The Minimal Value of the Inversion Criterion Function

Let us assume that the squared matrix $\boldsymbol{Z}$ (22) is reversible, so the matrix $\boldsymbol{Z}^{-1} = [\mathbf{r}_1,\ldots, \mathbf{r}_n]$ (16) exists. In this case, the below linear equation (34) has well defined solution $\mathbf{w}_Z$:

$$\boldsymbol{Z}\,\mathbf{w}_Z = \mathbf{1} = [1,\ldots,1]^T \tag{38}$$

and

$$\mathbf{w}_Z = \boldsymbol{Z}^{-1}\,\mathbf{1} = \mathbf{r}_1 + \ldots + \mathbf{r}_n \tag{39}$$

Lemma 1: If the data matrix $\boldsymbol{Z}$ (21) is reversible, then each of the n dual hyperplanes h_j^1 (29) passes through the point $\mathbf{w}_Z$ (39).

This Lemma results directly from the set of the linear equations (38). The i-th equation in the set (38) has the form $\mathbf{z}_i^T\mathbf{w}_Z = 1$. This means that the i-th hyperplane h_i^1 (28) passes through the point $\mathbf{w}_Z$ (39). Therefore the solution $\mathbf{w}_Z$ (39) is the point of intersection of all the hyperplanes h_j^1 (28) defined by the n vectors $\mathbf{z}_j$ (22). It also means that the point $\mathbf{w}_Z$ (39) is one of the vertices $\mathbf{w}_k$ (36).

Lemma 2: If the squared matrix $\boldsymbol{Z}$ (22) is reversible ($\boldsymbol{Z}^{-1}$ exists), then the value $\Phi_{inv}(\mathbf{w}_Z)$ of the criterion function $\Phi_{inv}(\mathbf{w})$ (27) in the vertex $\mathbf{w}_Z$ (39) is equal to zero:

$$\Phi_{inv}(\mathbf{w}_Z) = 0 \tag{40}$$

Proof: If the j-th hyperplane h_i^1 (29) passes through the point $\mathbf{w}_Z$ (39), then $\mathbf{z}_i^T\mathbf{w}_Z = 1$. This means that the value $\varphi_i(\mathbf{w}_Z)$ of the j-th penalty functions $\varphi_i(\mathbf{w})$ (26) is eqal to zero in this point ($\varphi_i(\mathbf{w}_Z) = 0$). In accordance with the *Lemma* 1, each hyperplane h_i^1 (29) passes through the vertex $\mathbf{w}_Z$ (39). So, each penalty functions $\varphi_j(\mathbf{w})$ (26) and the inversion criterion function $\Phi_{inv}(\mathbf{w})$ (27) are equal to zero in the point $\mathbf{w}_Z$ (39). □

Theorem 1: The minimal value $\Phi_{inv}(\mathbf{w}_k^*)$ (37) of the inversion criterion function $\Phi_{inv}(\mathbf{w})$ (27) defined on elements $\mathbf{z}_j$ of the data matrix $\boldsymbol{Z}$ (22) is equal to zero if and only if the matrix $\boldsymbol{Z}$ is reversible ($\boldsymbol{Z}^{-1}$ exists).

Proof: As follows from the *Lemma* 2, the minimal value $\Phi_{inv}(\mathbf{w}_k^*)$ (28) is equal to zero if the data matrix $\boldsymbol{Z}$ (21) is reversible. In this case there exists such optimal vertex $\mathbf{w}_k^* = \mathbf{w}_Z$ (39) that the minimal value $\Phi_{inv}(\mathbf{w}_k^*)$ (37) is equal to zero. We can also remark that the value $\Phi_{inv}(\mathbf{w}_{k'})$ of the criterion function $\Phi_{inv}(\mathbf{w})$ (27) is greater than zero in any other vertex $\mathbf{w}_{k'}$ (36):

$$(\forall \mathbf{w}_{k'} \neq \mathbf{w}_k^*)\;\; \Phi_{inv}(\mathbf{w}_{k'}) > 0 \tag{41}$$

If the data matrix $\boldsymbol{Z}$ (22) is not reversible, then there does not exist a vertex $\mathbf{w}_{k'}$ (36) through which all the hyperplanes h_i^1 (29) pass. If the j'-th hyperplane $h_{i'}^1$ (29) does not pass through the vertex $\mathbf{w}_{k'}$ (36), then the penalty function $\varphi_{j'}(\mathbf{w})$ (26) is greater than zero in the point $\mathbf{w}_{k'}$ ($\varphi_{j'}(\mathbf{w}_{k'}) > 0$). As a result $\Phi_{inv}(\mathbf{w}_{k'}) > 0$. □

The minimal value $\Phi_{inv}(\mathbf{w}_k^*)$ (28) of the inversion criterion function $\Phi_{inv}(\mathbf{w})$ (27) can be found by using the basis exchange algorithm based on the Gauss-Jordan transformations (19), (20) of the inverse matrices $\mathbf{B}_k^{-1} = [\mathbf{r}_1(k),\ldots, \mathbf{r}_n(k)]$ (16) during successive steps k ($k = 1,\ldots, K$). If the data matrix $\boldsymbol{Z}$ (22) is reversible, then after a finite number K ($K = n$) of the steps k the optimal vertex $\mathbf{w}_K^*$ constituting the minimal value $\Phi_{inv}(\mathbf{w}_k^*) = 0$ (28) is reached. In this case, the matrix $\mathbf{B}_K^{-1}$ resulting from the Gauss-Jordan transformations (19), (20) of the matrices $\mathbf{B}_k^{-1}$ is equal to the inverse data matrix $\boldsymbol{Z}$ (22) ($\boldsymbol{Z}^{-1} = \mathbf{B}_K^{-1}$).

One of the useful properties of the minimal value $\Phi_{inv}(\mathbf{w}_k^*)$ (37) of the inversion criterion function $\Phi_{inv}(\mathbf{w})$ (27) is their *invariance* in respect to the reversible, linear transformation of the *n* vectors $\mathbf{z}_j$ constituting the *n* columns of the data matrix $\boldsymbol{Z}$ (22):

$$\Phi_{inv}'(\mathbf{w}_k') = \Phi_{inv}(\mathbf{w}_k^*) \tag{42}$$

Where the symbol $\Phi_{inv}'(\mathbf{w}_k')$ stands for the minimal value (37) of the criterion function $\Phi_{inv}'(\mathbf{w})$ (27) defined on the transformed vectors $\mathbf{z}_j'$:

$$(\forall j \in \{1,\ldots,n\}) \quad \mathbf{z}_j' = \boldsymbol{A}\,\mathbf{z}_j, \text{ where the matrix } \boldsymbol{A}^{-1} \text{ exists} \tag{43}$$

The invariance property (42) results from the below equalities

$$(\forall j \in \{1,\ldots,n\}) \quad (\mathbf{w}')^T\mathbf{z}_j' = \mathbf{w}^T\mathbf{z}_j, \text{ where } \mathbf{w}' = \boldsymbol{A}^{-1}\,\mathbf{w} \tag{44}$$

So the penalty functions $\varphi_i'(\mathbf{w}')$ (26) defined on the transformed vectors $\mathbf{z}_j'$ (43) have the same values as the functions $\varphi_j(\mathbf{w})$ in the point $\mathbf{w}$:

$$(\forall j \in \{1,\ldots,n\}) \quad \varphi_i'(\mathbf{w}') = \varphi_i(\mathbf{w}) \tag{45}$$

The linear transformation (43) of the *n* vectors $\mathbf{z}_j$ includes their scaling:

$$(\forall j \in \{1,\ldots,n\}) \quad \mathbf{z}_i' = s_i\,\mathbf{z}_j, \text{ where } s_i \neq 0 \; (s_i \in R^1) \tag{46}$$

We can also remark that the minimal value $\Phi_{inv}(\mathbf{w}_k^*)$ (37) of the criterion function $\Phi_{inv}(\mathbf{w})$ (27) is invariant to the scaling of the thresholds s_j in the hyperplanes h_j^{sj} (29).

$$(\forall j \in \{1,\ldots,n\}) \quad h_i^{sj} = \{\mathbf{w}: \mathbf{z}_j^T\mathbf{w} = s_j, \text{ where } s_j \neq 0 \; (s_j \in R^1)\} \tag{47}$$

The above invariance properties (42) of the minimal value $\Phi(\mathbf{w}_k^*)$ (37) of the *CPL* criterion function $\Phi(\mathbf{w})$ (27) encourage the use of this type of functions and the basis exchange algorithms also for the efficient solution of large, high dimensional systems of linear equations [7]:

$$\boldsymbol{A}\,\mathbf{w} = \mathbf{b} \tag{48}$$

where $\boldsymbol{A}$ is the matrix of dimension $m * n$ and $\mathbf{b} = [b_1,...,b_m]^T \in R^m$ is the *m*-dimensional vector.

8 Concluding Remarks

The convex and piecewise linear (*CPL*) inversion criterion function $\Phi_{inv}(\mathbf{w})$ (27) has been defined here on the *m* vectors $\mathbf{z}_i$ constituting the matrix $\boldsymbol{Z}$ (22). The *CPL* criterion function $\Phi_{inv}(\mathbf{w})$ (27) allows to select the exit criterion, the entry criterion and the stop criterion of the basis exchange algorithm aimed at the reversing of the matrix $\boldsymbol{Z}$. The reverse $\boldsymbol{Z}^{-1}$ matrix can be computed efficiently by using the basis exchange algorithm. In this approach, the number *K* (24) of the basis $\mathbf{B}_k$ (35) exchanges is not greater than the dimensionality *n* of the feature space $F[n]$ ($K \leq n$).

The convex and piecewise linear (*CPL*) criterion function $\Phi_{inv}(\mathbf{w})$ (27) can serve to solve also other problems related to the inversion of high dimensional matrices $\boldsymbol{Z}$ (22). In accordance with the *Theorem* 1, the minimal value $\Phi_{inv}(\mathbf{w}_k^*)$ (37) of the function $\Phi_{inv}(\mathbf{w})$ (27) is equal to zero if and only if the inverse matrix $\boldsymbol{Z}^{-1}$ exists. If the matrix $\boldsymbol{Z}$ is singular, the minimal value $\Phi_{inv}(\mathbf{w}_k^*)$ (37) is greater than zero ($\Phi_{inv}(\mathbf{w}_k^*) > 0$). The minimal value $\Phi_{inv}(\mathbf{w}_k^*)$ (37) can be used as the detector of the matrix $\boldsymbol{Z}$ (22) singularity. Such measure of singularity degree of the matrix $\boldsymbol{Z}$ (22) has useful invariance property (42).

The relaxed linear separability method (*RLS*) of feature subset selection has been developed recently on the basis of the *CPL* perceptron criterion function [13]. The perceptron criterion function is used in the *RLS* method for efficient designing of linear classifiers and to evaluate the linear seprabilty of learning sets in different feature subspaces $F_k[n_k]$ ($F_k[n_k] \subset F[n]$). One of the uses of the *RLS* method was to extract optimal subsets of genes from the *Breast cancer* data set [14]. The *Breast Cancer* data set contains descriptions of 46 cancer and 51 non-cancer women. Each woman in this set was characterized by n = 24481 genes. The *RLS* method allowed to select the optimal subset of n_1 = 12 genes and such linear combination of these selected genes, which correctly (100%) distinguish cancer from non-cancer women in this set. In this example, the dimension $n * n$ of the inverted matrices almost reached the number $6*10^8$ [14]. It's possible to include the inversion criterion function $\Phi_{inv}(\mathbf{w})$ (27) into *RLS* method. Such inclusion could increase the range of applications of the *RLS* method.

The presented here method of large matrices inversion is based on the basis exchange algortihm linked to minimization of the *CPL* inverse criterion function $\Phi_{inv}(\mathbf{w})$ (27). The basis exchange algortihms controlled by other types of the *CPL* criterion functions have been used in many tasks of data mining and machine learning [9]. For example, the optimal gene subset selection problem has been solved efficiently in accordance the mentioned above *RLS* method [14]. It has been shown experimentally that the basis exchange algortihms can supply effective tools for the exploration of large, multivariate data sets.

The proposed method of matrices inversion could also be useful in other chalenging computational tasks. For example, computations of pseudo inverse covariance matrices in undersampled data sets could be performed this way [15].

Acknowledgements

This work was supported by the project St/2017 grant from the Nalecz Institute of Biocybernetics and Biomedical Engineering, Polish Academy of Sciences

Competing Interests

Author has declared that no competing interests exist.

References

[1] Hand D, Smyth P, Mannila H. Principles of data mining. MIT Press, Cambridge; 2001.

[2] Johnson RA, Wichern DW. Applied multivariate statistical analysis. Prentice-Hall, Inc., Englewood Cliffs, New York; 1991.

[3] Duda OR, Hart PE, Stork DG. Pattern Classification. J. Wiley, New York; 2001.

[4] Cai T, Liu W, Luo X. A constrained ℓ1 minimization approach to sparse precision matrix estimation. Journal of the American Statistical Association. 2011;106:594–607.

[5] Sharma G, Agarwala A, Bhattacharya B. A fast parallel Gauss-Jordan algorithm for matrix inversion using cuda. Computers & Structures. 2013;128:31–37.

[6] Sun T, Zhang C. Sparse matrix inversion with scaled lasso. Journal of Machine Learning Research. 2013;14.

[7] Bobrowski L, Niemiro W. A method of synthesis of linear discriminant function in the case of nonseparabilty. Pattern Recognition. 1984;17:205-210.

[8] Bobrowski L. Design of piecewise linear classifiers from formal neurons by some basis exchange technique. Pattern Recognition. 1991;24(9):863-870.

[9] Bobrowski L. Data mining based on convex and piecewise linear criterion functions (*in Polish*). Technical University Białystok; 2005.

[10] Simonnard M. Linear programming, Prentice-Hall; 1966.

[11] Rosenblatt F. Principles of neurodynamics: Perceptrons and the theory of brain mechanisms. Spartan Books, Washington; 1962.

[12] Bobrowski L. Discovering main vertexical planes in a multivariate data space by using *CPL* functions. ICDM. 2014, Ed. Perner P., Springer Verlag, Berlin; 2014.

[13] Bobrowski L, Łukaszuk T. Relaxed linear separability (*RLS*) approach to feature (Gene) Subset Selection. In: Selected Works in Bioinformatics, Edited by: Xuhua Xia. INTECH. 2011;103–118.

[14] Van't Veer LJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature. 2002;41-5(6871):530–536.

[15] Mahmood R, Aamir KM, Milojević-Jevrić M, Radenović, Zia T. On designing invertible pseudo covariance matrix for undersampled cases in classification, British Journal of Mathematics and Computer Science. 2016;17(5):1-9. (Article no. BJMCS. 27435).

Peer-review history:
The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)
http://sciencedomain.org/review-history/18203