# Building a discriminatively ordered subspace on the generating matrix to classify high-dimensional spectral data

Rui Zhu[a], Kazuhiro Fukui[b], Jing-Hao Xue[a,*]

[a] *Department of Statistical Science, University College London, London WC1E 6BT, UK*
[b] *Department of Computer Science, Graduate School of Systems and Information Engineering, University of Tsukuba, Ibaraki 305-8573, Japan*

**Abstract**

Soft independent modelling of class analogy (SIMCA) is a widely-used subspace method for spectral data classification. However, since the class subspaces are built independently in SIMCA, the discriminative between-class information is neglected. An appealing remedy is to first project the original data to a more discriminative subspace. For this, generalised difference subspace (GDS) that explores the information between class subspaces in the generating matrix can be a strong candidate. However, due to the difference between a class subspace (of infinite scale) and a class (of finite scale), the eigenvectors selected by GDS may not also be discriminative for classifying samples of classes. Therefore in this paper, we propose a discriminatively ordered subspace (DOS): different from GDS, our DOS selects the eigenvectors with high discriminative ability between classes rather than between class subspaces. The experiments on three real spectral datasets demonstrate

*Corresponding author. Tel.: +44-20-7679-1863; Fax: +44-20-3108-3105
*Email address:* jinghao.xue@ucl.ac.uk (Jing-Hao Xue)

that applying DOS before SIMCA outperforms its counterparts.

*Keywords:* Discriminatively ordered subspace, generalised difference subspace, generating matrix, SIMCA, spectral data classification, subspace method

## 1. Introduction

High-dimensional spectral data, such as near infrared (NIR) spectroscopic data and mass spectrometry (MS) data, are widely used in a variety of fields, for example chemometrics, bioinformatics and hyperspectral image analysis. In the analysis of spectral data, classification is an omnipresent task [4, 10, 2, 9, 7, 13], which enables us to distinguish different species, identify the geographical origins of the products, or predict molecular substructure, to name a few.

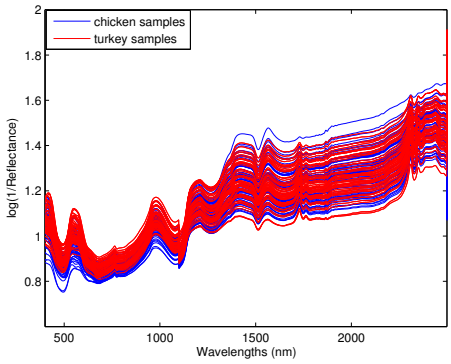


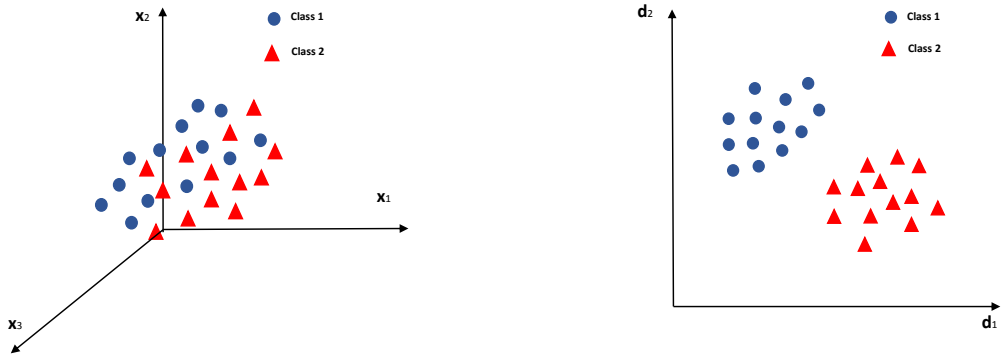

Figure 1: Spectra of meat samples from two classes: chicken and turkey.

Figure 1 shows an example for NIR spectroscopic data of two classes, the chicken meat samples and the turkey meat samples. Each curve depicts the spectrum of a sample, which is usually represented by a high-dimensional

feature vector. A classification task is to classify the spectra of new samples into the two classes based on the information provided by some labelled training spectra. In this paper, we focus on two-class classification. Based on the two-class classification results, multi-class classification can be readily obtained by using the one-vs-one or one-vs-all strategy [3].

Soft independent modelling of class analogy (SIMCA) [12] is a subspace-based classification method that is widely used in the two-class classification of high-dimensional spectral data in chemometrics [4, 10, 2]. When SIMCA is used for two-class classification, firstly two class subspaces are built for the two classes separately through using principal component analysis (PCA). Then an $F$-test, which tests whether the residual standard deviation of a new sample from the subspace of a class is statistically significantly different from the residual standard deviation of the training set of that class, is used to determine the class membership of the new sample. The PC-subspace is considered as a good class model for high-dimensional data because it extracts the most variable information in the data to few PCs and gets rid of a large amount of redundant information in the original feature dimensions. SIMCA is originally designed for both outlier detection and classification. In this paper, we treat SIMCA as a simple classification method that assign a new sample to the class with the smallest $F$-value as suggested in [8].

In spite of its wide use, SIMCA suffers from the problem that the class subspaces are built independently without considering between-class information. Therefore the $F$-value calculated independently for each class may not be discriminative enough to classify a new sample.




(a) Original feature space. (b) Discriminative subspace.

Figure 2: (a) Two classes of samples are mixed together in the original 3-dimensional feature space. (b) The same groups of samples can be well separated when they are projected to a discriminative 2-dimensional subspace.

An appealing solution to this problem is to find a more discriminative subspace than the original feature space and project the data to this subspace before applying SIMCA. The projections of the samples to this discriminative subspace are expected to be more separated and can be more easily classified than those in the original feature space, as illustrated in Figure 2. Also, as the new subspace contains more discriminative information for classification, the $F$-value calculated in this subspace is expected to be more discriminative. It is therefore the objective of our work in this paper to find such a discriminative subspace.

Recently, Fukui and Maki [6] propose the generalised difference subspace (GDS) projection as a preprocessing method to improve a popular subspace-based classifier called mutual subspace method (MSM) in image set-based object recognition. GDS aims to tackle an issue of MSM: the class subspaces are independently generated by PCA in a class-by-class manner, and thus

may not be strongly discriminative for classification. This issue is actually the same as that of SIMCA. Hence, we believe the GDS projection can also be utilised as a preprocessing method for SIMCA to improve its classification performance.

GDS is a subspace containing the information about difference between class subspaces, and thus is supposed to be more discriminative than the original feature space. GDS is generated on the basis of a generating matrix $\boldsymbol{G}_D$, which is calculated as the sum of the projection matrices of the two class subspaces and can provide between-class information. Fukui and Maki [6] show that the eigenvectors of $\boldsymbol{G}_D$ with small eigenvalues contain the information of difference between class subspaces while those with large eigenvalues contain the information about similarity between class subspaces. The GDS projection thus keeps only the last few eigenvectors with small eigenvalues and discards the first few eigenvectors with large eigenvalues, in order to make use of the difference information.

The GDS projection shows superior performance on face recognition and hand shape recognition problems. However, there is a limitation of the GDS. The GDS projection discards the eigenvectors of $\boldsymbol{G}_D$ with large eigenvalues because they contain similarity information between class subspaces and thus are assumed ineffective for classification. This assumption is, however, not always valid due to the conceptual difference between a class subspace (of infinite scale) and a class (of finite scale). For example, two separable classes may span the same subspace. More technically, this assumption defines similarity information by using the eigenvector directions only, without considering the distribution of the projected samples in these directions. If the projected

samples of different classes in the directions of similarity (i.e. the directions with large eigenvalues of $\boldsymbol{G}_D$) are still class separable, then these directions can also be discriminative in separating classes (although not discriminative in separating class subspaces), and thus discarding them can be harmful for classification of samples.

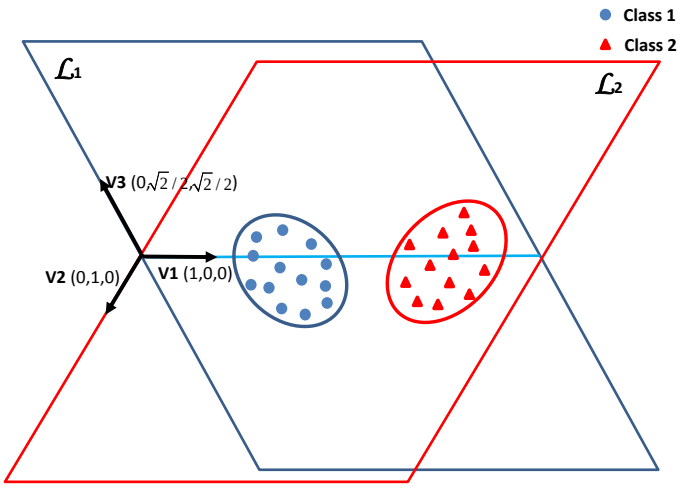
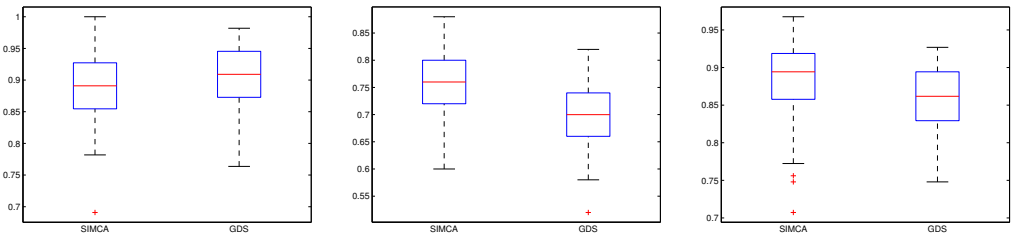



Figure 3: An illustrative example of the difference between a class subspace (of infinite scale) and a class (of finite scale).

To illustrate the difference between a class subspace and a class, we show an intuitive example in Figure 3. The infinite scale subspace of class 1, $\mathcal{L}_1$, is spanned by $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$, and the infinite scale subspace of class 2, $\mathcal{L}_2$, is spanned by $\boldsymbol{v}_1$ and $\boldsymbol{v}_3$. The samples of the two classes lie in the two ellipses with finite scales in $\mathcal{L}_1$ and $\mathcal{L}_2$, respectively. It is obvious that $\boldsymbol{v}_1$ is the intersection of $\mathcal{L}_1$ and $\mathcal{L}_2$, which represents the same direction, i.e. the similarity information, between class subspaces. The GDS projection

discards $\boldsymbol{v}_1$ because it is the eigenvector of $\boldsymbol{G}_D$ with the largest eigenvalue and contains similarity information between class subspaces. However, the samples of the two classes are class separable on the direction of $\boldsymbol{v}_1$, which suggests that $\boldsymbol{v}_1$ contains discriminative information between classes. (We shall demonstrate another motivating example for this issue in Section 2.3.1 using a real spectral dataset.)




(a) meat (b) Phenyl (c) fat

Figure 4: Classification accuracies of SIMCA and the GDS-preprocessed SIMCA on three real spectral datasets: meat, Phenyl and fat. In each panel, the left-hand boxplot is for SIMCA, and the right-hand boxplot is for the GDS-preprocessed SIMCA.

Moreover, here we illustrate that discarding the eigenvectors of $\boldsymbol{G}_D$ with large eigenvalues can be harmful for classification using three real spectral datasets: meat, Phenyl and fat. In Figure 4, we plot the classification accuracies of SIMCA and the GDS-preprocessed SIMCA on the three datasets. We can clearly observe that a preprocessing step of SIMCA by GDS does not necessarily benefit the classification performance of SIMCA; it actually has an negative effect (lowering classification accuracy) on SIMCA for the Phenyl dataset and the fat dataset. Detailed discussion on this will be provided in Section 3.

To make use of the between-class information in $\boldsymbol{G}_D$ and to overcome

the above limitation of the GDS projection, we propose a discriminatively ordered subspace (DOS): our DOS is spanned by the most discriminative eigenvectors of $\boldsymbol{G}_D$ instead of the eigenvectors with small eigenvalues and extracts the most discriminative information from the data. That is, we sort the eigenvectors in terms of their discriminative ability and select the top-ranked eigenvectors with high discriminative abilities to generate the DOS projection. This discriminatively ordering procedure during the generation of the subspace is where the term 'discriminatively ordered' was from in DOS. As our objective is to develop DOS to tackle the issue of SIMCA, the discriminative ability of an eigenvector is measured by the classification accuracy of SIMCA on the samples projected to this eigenvector. The higher the classification accuracy, the higher the discriminative ability. We choose this filter-type of eigenvector selection scheme for high-dimensional spectral data, taking into consideration its simplicity and efficiency, as well as the uncorrelatedness and orthogonality of the candidate eigenvectors. The effectiveness of the DOS-preprocessed SIMCA will be demonstrated in Section 3.

The rest of this paper is organised as follows. In Section 2, a discussion of the GDS projection and a detailed description of the DOS projection are provided. In Section 3, GDS and DOS are compared for the improvement of classification performance of SIMCA on real spectral datasets. Section 4 presents some concluding remarks.

## 2. Methodology

### *2.1. SIMCA*

In the training phase of SIMCA, suppose $\boldsymbol{X}_k \in \mathbb{R}^{N_k \times p}$ is the training set of class $k$ ($k = 1, 2$), in which there are $N_k$ training instances and each instance is represented by a $p$-dimensional data vector (i.e. in the original $p$-dimensional feature space). To build the principal component (PC) subspace for each class, we apply eigendecomposition to the covariance matrix of the $k$th class:

$$\mathrm{Cov}(\boldsymbol{X}_k) = \frac{1}{N_k - 1}(\boldsymbol{X}_k^c)^T \boldsymbol{X}_k^c = \boldsymbol{U}_k \boldsymbol{\Sigma} \boldsymbol{U}_k^{-1} \ , \tag{1}$$

where $\boldsymbol{X}_k^c$ is the column-centred $\boldsymbol{X}_k$; the columns of $\boldsymbol{U}_k \in \mathbb{R}^{p \times q_k}$ denote the normalised eigenvectors, and $\boldsymbol{\Sigma}$ is a diagonal matrix with eigenvalues $\{\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{q_k}\}$. We select the first $r_k$ ($r_k \leq q_k$) columns of $\boldsymbol{U}_k$ as the basis vectors $\boldsymbol{W}_k$ that spans the $k$th class subspace $\mathcal{P}_k$, which is $r_k$-dimensional.

It follows that the projection matrix $\boldsymbol{P}_k \in \mathbb{R}^{p \times p}$ of $\mathcal{P}_k$ can be written as

$$\boldsymbol{P}_k = \boldsymbol{W}_k \boldsymbol{W}_k^T \ . \tag{2}$$

In the test phase, a new sample $\boldsymbol{x}_{new}$ is assigned based on the following two residuals. First, the residual of the $k$th class in the training set:

$$\boldsymbol{E}_k = \boldsymbol{X}_k^c - \boldsymbol{X}_k^c \boldsymbol{P}_k \ . \tag{3}$$

Second, the residual of $\boldsymbol{x}_{new}$ when it is projected to the $k$th class subspace:

$$\boldsymbol{e}_{k,new} = \boldsymbol{x}^c_{new} - \boldsymbol{x}^c_{new}\boldsymbol{P}_k \ , \tag{4}$$

where $\boldsymbol{x}^c_{new}$ is centred by the mean vector of $\boldsymbol{X}_k$. Then $\boldsymbol{x}_{new}$ is assigned to the class with the smallest $F$-value [8], where the $F$-value is defined as

$$F = \frac{||\boldsymbol{e}_{k,new}||_2^2}{||\boldsymbol{E}_k||_2^2/(N_k - r_k - 1)} \ , \tag{5}$$

in which $||\cdot||$ denotes the Frobenius norm.

### *2.2. Generalised difference subspace*

Since the class subspaces in SIMCA are built independently, the between-class information is not considered by SIMCA and thus the classification performance is limited. To improve the performance of SIMCA, we aim to find a subspace more discriminative than the original feature space. Applying SIMCA to the projections of the samples in this discriminative subspace is expected to have better performance because the samples are expected to be more separated in this subspace. The process of seeking and projecting to such a discriminative subspace can be treated as a preprocessing step of SIMCA.

Mutual subspace method (MSM) is a commonly used subspace-based method for image set-based object classification, which has a similar problem as SIMCA: MSM builds the class subspace by using PCA for each class separately. The generated class subspace of an image set of an unknown object is compared with the known class subspaces of reference objects and

classified to the class with the smallest canonical angle.

When the image set of an unknown object contains only one image, the image is represented by a feature vector and the canonical angles are calculated between the vector and the class subspaces. In this case, MSM is reduced to the commonly-used subspace method (SM) in image classification. The only difference between SM and SIMCA is the criterion for assigning new samples: SM assigns the new sample to the class with the smallest canonical angle between the sample and the class subspace, while SIMCA assigns the new sample to the class with the smallest $F$-value calculated in (5).

MSM suffers from the problem that the class subspaces generated by PCA may not be sufficiently discriminative for classification. Hence recently Fukui and Maki [6] propose to project the data onto a generalised difference subspace (GDS) as a preprocessing step of MSM, so as to improve the classification performance of MSM. GDS contains difference information between two class subspaces and is more discriminative to separate the two class subspaces than the original feature space. Thus the projections of the samples to GDS are expected to be more separated and can be better classified. Since SIMCA and MSM suffer from similar problems, we believe the GDS projection can also be used as a preprocessing method of SIMCA to improve the classification performance of the latter.

#### *2.2.1. GDS*

The GDS projection is proposed on the basis of the properties of the difference subspace (DS) of two class subspaces. The DS, denoted by $\mathcal{D}$, is

calculated by using the sum matrix $\boldsymbol{G}_D \in \mathbb{R}^{p\times p}$, which is defined as

$$\boldsymbol{G}_D = \sum_{k=1}^{K} \boldsymbol{P}_k \ , \tag{6}$$

where $K = 2$. Applying eigendecomposition to $\boldsymbol{G}_D$, we obtain

$$\boldsymbol{G}_D = \boldsymbol{V}_D \boldsymbol{\Lambda}_D \boldsymbol{V}_D^T \ , \tag{7}$$

where the columns in $\boldsymbol{V}_D = [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_{r_D}] \in \mathbb{R}^{p\times r_D}$ are the normalised eigenvectors of $\boldsymbol{G}_D$, and $\boldsymbol{\Lambda}_D$ denotes the diagonal matrix with corresponding eigenvalues $\{\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{r_D}\}$ in descending order, where $r_D = \text{rank}(\boldsymbol{G}_D)$.

The DS is defined as the subspace spanned by the eigenvectors $\boldsymbol{v}_i$ in $\boldsymbol{V}_D$ with corresponding eigenvalues $\lambda_i$ less than one. As shown by Fukui and Maki [6], these eigenvectors are proportional to the difference between the canonical vector pairs of the two class subspaces, and hence they contain the difference information between the two class subspaces.

In addition to DS, Fukui and Maki [6] also define the principal component subspace (PCS), denoted by $\mathcal{M}$, which is spanned by the eigenvectors $\boldsymbol{v}_i$ in $\boldsymbol{V}_D$ with corresponding eigenvalues $\lambda_i$ larger than one. They point out that $\mathcal{M}$ contains the similarity information between class subspaces, because the eigenvectors are proportional to the sum of the canonical vector pairs.

Based on the properties of the DS, Fukui and Maki [6] propose the generalised DS (GDS) projection for $K$ ($K \geq 2$) classes. The GDS projection discards the first few eigenvectors of $\boldsymbol{G}_D$ with large eigenvalues and keeps only the last few eigenvectors of $\boldsymbol{G}_D$ with small eigenvalues. In this way, the

GDS spanned by the last few eigenvectors contains difference information between class subspaces. The projections of the samples onto GDS are expected to be more separated and can be better classified. The dimension of GDS is determined by maximising the mean canonical angles between class subspaces, as suggested in Fukui and Maki [6].

### *2.2.2. The generating matrix*

To further investigate the properties of the sum matrix $\boldsymbol{G}_D$ and the GDS, we introduce the generating matrix proposed in Therrien [11]. The generating matrix is defined as the linear combination of the projection matrices of the two class subspaces [11]. Therrien [11] shows that the generating matrix can be used to find the intersection of the class subspaces.

For two classes, the generating matrix $\boldsymbol{G} \in \mathbb{R}^{p\times p}$ can be written as

$$\boldsymbol{G} = \sum_{k=1}^{K} \alpha_k \boldsymbol{P}_k \ , \tag{8}$$

where $K = 2$, $\alpha_k \in (0, 1)$, and $\sum_{k=1}^{K} \alpha_k = 1$. Applying eigendecomposition to $\boldsymbol{G}$, we can obtain

$$\boldsymbol{G} = \boldsymbol{V}_G \boldsymbol{\Lambda}_G \boldsymbol{V}_G^T \ , \tag{9}$$

where the columns of $\boldsymbol{V}_G \in \mathbb{R}^{p\times r_G}$ denote the normalised eigenvectors of $\boldsymbol{G}$, and $\boldsymbol{\Lambda}_G$ denotes the diagonal matrix with eigenvalues $\{\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{r_G}\}$, where $r_G = \text{rank}(\boldsymbol{G})$.

Therrien [11] shows three important properties of $\boldsymbol{G}$. First, the eigenvalues of $\boldsymbol{G}$ are in the interval $[0, 1]$. Second, the eigenvectors with the corresponding eigenvalues of one span the intersection of the two subspaces

$\bigcap_{k=1}^{2} \mathcal{P}_k$. Third, the eigenvectors with nonzero eigenvalues span the sum subspace of the two classes, and the eigenvectors with eigenvalues of zeros span the complement of this sum subspace.

Since the vectors in $\bigcap_{k=1}^{2} \mathcal{P}_k$ are in both $\mathcal{P}_1$ and $\mathcal{P}_2$, $\bigcap_{k=1}^{2} \mathcal{P}_k$ denotes the subspace that contains the most similar directions of the two class subspaces. In other words, the most similar directions of the two class subspaces are extracted by the eigenvectors of $\boldsymbol{G}$ with eigenvalues of one. In contrast, the eigenvectors with eigenvalues of zeros are the complements of the sum subspace which contain information that is irrelevant to the two class subspaces. The larger the eigenvalue, the more similarity information the corresponding eigenvector contains.

The generation of GDS is closely related to the generating matrix: $\boldsymbol{G}_D$ and $\boldsymbol{G}$ are both linear combinations of $\boldsymbol{P}_k$ although with different coefficients. The linear coefficients of $\boldsymbol{G}_D$ are all one, i.e. $\alpha_k = 1 \; \forall \; k$, while those of $\boldsymbol{G}$ are constrained by $\alpha_k \in (0, 1)$ and $\sum_{k=1}^{K} \alpha_k = 1$. Although $\boldsymbol{G}_D$ and $\boldsymbol{G}$ are slightly different, we can derive similar properties of $\boldsymbol{G}_D$ as those of $\boldsymbol{G}$ by following the proofs in [11]. First, the eigenvalues of $\boldsymbol{G}_D$ are in the interval $[0, 2]$. Second, the eigenvectors with the corresponding eigenvalues of two span the intersection of the two subspaces $\bigcap_{k=1}^{2} \mathcal{P}_k$. Third, the eigenvectors with the corresponding eigenvalues that are nonzero span the sum subspace of the two subspace and those with zero eigenvalues span the complement of the sum subspace. Hence, with some abuse of notation, we also call the sum matrix $\boldsymbol{G}_D$ a generating matrix.

The eigenvectors of $\boldsymbol{G}_D$ with eigenvalues in $(1, 2]$ span the PCS $\mathcal{M}$ which contains similarity information between the two class subspaces. This argu-

ment seems to be consistent with the property of $\boldsymbol{G}_D$, based on the assumption that the eigenvectors closed to the intersection directions contain large amount of similarity information. Since the eigenvectors with eigenvalues of two span the intersections subspace, the eigenvectors with eigenvalues close to two could be close to the intersection directions. On the other hand, the eigenvectors with eigenvalues far from two, i.e. eigenvalues in $[0, 1)$, are far from the intersection directions. Therefore, the GDS projection aims to discard the eigenvectors that are close to the intersection directions, so as to provide a discriminative subspace.

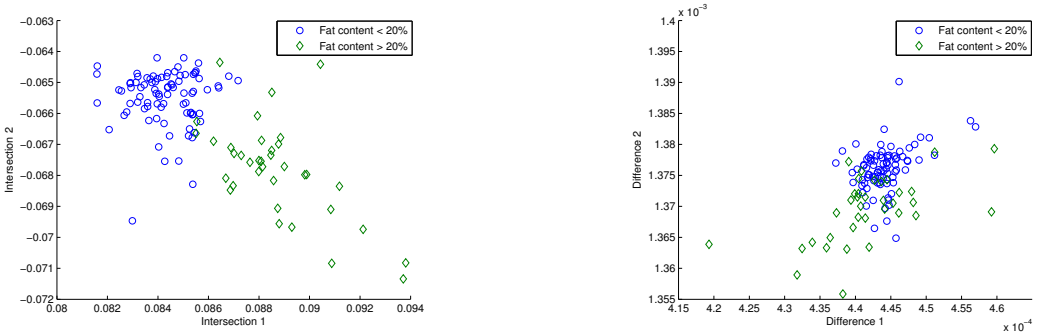
### *2.3. Discriminatively ordered subspace*

The GDS projection is based on the assumption that, because the first few eigenvectors with large eigenvalues close to the intersection directions contain similarity information between the class subspaces, they are not important for classification. However, this assumption is not always true, as a class subspace (of infinite scale) and a class (of finite scale) are different, and hence the ability to discriminate two class subspaces are not necessarily in line with the ability to discriminate samples of two classes. In the extreme case, two separable classes may span the same class subspace. More technically, the similarity information in the GDS assumption only considers the directions, while the scores or the projection values on the directions should also be considered. The eigenvectors of $\boldsymbol{G}_D$ that are close to the intersection directions between the two class subspaces can be discriminative when the scores on these eigenvectors are largely separable between classes. In the following section, we show a motivating real-data example that even the directions in the intersection subspace of the two classes can be discriminative.

### *2.3.1. Intersection and discriminative ability: a motivating example*

The fat dataset contains 193 spectra of finely chopped meat measured at 100 wavelengths, in which 122 samples contain less than 20% fat and 71 samples contain more than 20% fat. Detailed description of this dataset can be found in Section 3.1. We split the dataset into a training set and a test set: 35 samples with fat content less than 20% and 35 samples with fat content more than 20% are randomly sampled into the training set; the rest samples form the test set.

The projection matrix $\boldsymbol{P}_k$ is calculated by using all the 34 available eigenvectors of each class. There are 68 eigenvectors that can be obtained from the eigendecomposition of $\boldsymbol{G}_D$, in which the first seven eigenvectors have eigenvalues of two and the last 34 eigenvectors have eigenvalues less than one. Thus the first seven eigenvectors span the intersection of the two class subspaces and the last 34 eigenvectors span the DS.




(a) (b)

Figure 5: (a) Projections of the test samples onto two directions of the intersection. (b) Projections of the test samples onto two directions of the DS.

Figure 5 shows two scatter plots of the test samples. Figure 5a shows the projections of the test samples onto two intersection directions, and Figure 5b

shows the projections of the test samples onto the first two DS directions. It is clear that the test samples can be well separated when projected onto the two directions in the intersection subspace, whereas the projections of the test samples onto the two directions of DS show slight separation with a mixture in the central region. In other words, this indicates that the two eigenvectors in the intersection subspace are more discriminative than those in DS. Therefore, it is better to keep the two eigenvectors in the intersection subspace instead of those in the DS.

This counter-example demonstrates that the eigenvectors of $\boldsymbol{G}_D$ in the intersection directions can be discriminative and the assumption in the GDS method is not valid in this case.

### *2.3.2. Discriminatively ordered subspace*

As shown in Section 2.2.2, the eigenvectors of the generating matrix $\boldsymbol{G}_D$ contain between-class information. Thus we are able to select discriminative eigenvectors of $\boldsymbol{G}_D$ to generate a discriminative subspace for better classification. In the GDS projection, the eigenvectors of $\boldsymbol{G}_D$ are sorted by the eigenvalues in descending order, and the last few eigenvectors with small eigenvalues are selected to generate the GDS. However, as we have shown, the eigenvectors with large eigenvalues are possible to be more discriminative than those with small eigenvalues, and discarding the eigenvectors with large eigenvalues that are discriminative may be harmful for classification.

Therefore, instead of using the GDS projection, we aim to select the most discriminative eigenvectors of $\boldsymbol{G}_D$ to generate a discriminative subspace. We propose a discriminatively ordered subspace (DOS), which uses the discriminative ability (rather than eigenvalues) to sort the eigenvectors in ascending

order and select the last few eigenvectors with high discriminative ability to generate the discriminative subspace. In our case for improving SIMCA, the discriminative ability of an eigenvector is measured by the classification accuracy of SIMCA on the samples projected to this eigenvector. For each eigenvector, if the projections of the samples of the two classes are more separated, then the classification accuracy of SIMCA will be high. This simple eigenvector-by-eigenvector selection scheme is appropriate for high-dimensional spectral data, given that the candidate eigenvectors are uncorrelated. In the end we choose a set of eigenvectors with high discriminative abilities to span a subspace that can make the samples of the two classes more separated and improve the performance of SIMCA.

Specifically, given the generating matrix $\boldsymbol{G}_D$ in (6) and its eigendecomposition in (7), the eigenvectors $\boldsymbol{v}_i$ $(i = 1, \ldots, r_D)$ are sorted using their discriminative abilities $d_i$, which are calculated using leave-one-out cross-validation (LOOCV) on the training set as follows.

The training set is denoted as $\boldsymbol{X}_{train}^T = [\boldsymbol{X}_1^T, \boldsymbol{X}_2^T] = [\boldsymbol{x}_1^T, \ldots, \boldsymbol{x}_{N_1+N_2}^T] \in \mathbb{R}^{p\times(N_1+N_2)}$, where $\boldsymbol{X}_1^T = [\boldsymbol{x}_1^T, \ldots, \boldsymbol{x}_{N_1}^T] \in \mathbb{R}^{p\times N_1}$ and $\boldsymbol{X}_2^T = [\boldsymbol{x}_{N_1+1}^T, \ldots, \boldsymbol{x}_{N_1+N_2}^T] \in \mathbb{R}^{p\times N_2}$ are the training sets for the two classes and $\boldsymbol{x}_m \in \mathbb{R}^{1\times p}$ is the $m$th $(m = 1, \ldots, N_1 + N_2$ ) training sample.

Firstly, we project all the training samples in $\boldsymbol{X}_{train}$ to each eigenvector $\boldsymbol{v}_i \in \mathbb{R}^{p\times 1}$ and obtain the projections $\hat{\boldsymbol{X}}_{train,i} = \boldsymbol{X}_{train}\boldsymbol{v}_i \in \mathbb{R}^{(N_1+N_2)\times 1}$. For the $m$th validation, the $m$th projection, $\hat{\boldsymbol{x}}_{m,i} = \boldsymbol{x}_m\boldsymbol{v}_i \in \mathbb{R}^{1\times 1}$, is used as the validation sample and the rest projections are used as the training samples.

Secondly, we apply SIMCA to each validation by setting the dimensions of the two class subspaces to zeros, i.e. $r_1 = r_2 = 0$. Based on (3), (4),

and (5), we observe that the $F$-value is dependent on the distance from the projected validation sample to the projected class centre. We assign the validation sample to the class with the smallest $F$-value.

Thirdly, for each eigenvector $\boldsymbol{v}_i$, we obtain $N_1 + N_2$ predictions from LOOCV. The classification accuracy $d_i$ is calculated as

$$d_i = \frac{N_c}{N_1 + N_2} \ , \tag{10}$$

where $N_c$ is the number of correctly classified test samples.

Fourthly, after obtaining $d_i's$ for $i = 1, \ldots, r_D$, we sort the eigenvectors $\boldsymbol{v}_i's$ in ascending order of $d_i's$ and obtain the matrix of the sorted eigenvectors $\boldsymbol{V}_{sort} = [\boldsymbol{v}_{(1)}, \boldsymbol{v}_{(2)}, \ldots, \boldsymbol{v}_{(r_D)}]$, where the discriminative ability $d_{(1)} < d_{(2)} < \cdots < d_{(r_D)}$. The last few eigenvectors in $\boldsymbol{V}_{sort}$ are selected to span the discriminative subspace $\mathcal{D}_s$, which we term discriminatively sorted subspace (DOS).

Finally, we project the samples to DOS and apply SIMCA to the projections of the samples. The dimension of $\mathcal{D}_s$ and the dimensions of the two class subspaces in $\mathcal{D}_s$ can be tuned by cross-validation through minimising the classification error of the training set.

## 3. Experiments

In the following experiments, we compare the performances of the original SIMCA without preprocessing, the SIMCA preprocessed by the linear discriminative analysis (LDA) projection, the SIMCA preprocessed by the GDS projection, and the SIMCA preprocessed by the DOS projection. The LDA-preprocessed SIMCA is also compared since LDA is a commonly used

method to find a discriminative subspace. Three real datasets are used in the experiments: the fat dataset, the meat dataset, and the Phenyl dataset. In the illustrations presented in this section, the DOS-preprocessed SIMCA is denoted by 'DOS', the GDS-preprocessed SIMCA is denoted by 'GDS', the LDA-preprocessed SIMCA is denoted by 'LDA' and the original SIMCA is denoted by 'SIMCA'.

### *3.1. Datasets*

#### *3.1.1. The meat dataset*

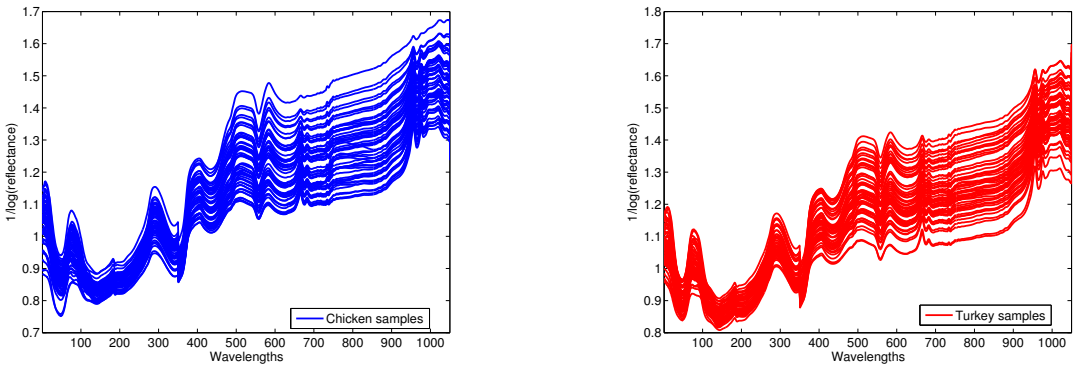



Figure 6: The spectra of the two classes in the meat dataset.

The meat dataset [1] contains beef, pork, lamb, chicken and turkey meat samples measured at 1051 wavelengths. Only the 55 chicken and 54 turkey samples in the dataset are used in our experiments since the two groups are difficult to classify. The first 350 wavelengths in the meat dataset are used because the experiments in Arnalds et al. [1] suggest that the first 350 wavelengths ranging from 400 to 1100 nm perform the best. The spectra of the meat dataset are illustrated in Figure 6.

During the training-test split, the total of 55 chicken samples and 54

turkey samples are randomly partitioned into a training set (27 chicken samples and 27 turkey samples) and a test set (28 chicken samples and 27 turkey samples).

### *3.1.2. The Phenyl dataset*

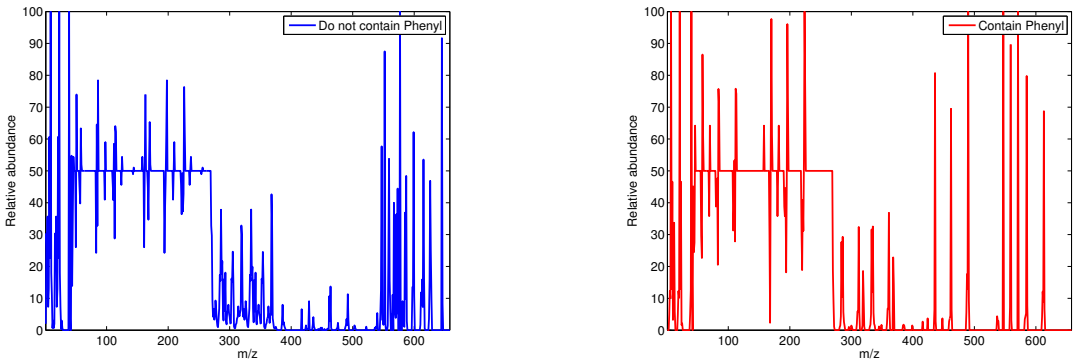



Figure 7: The spectra of the two classes in the Phenyl dataset.

The Phenyl dataset is provided in the R package, 'chemometrics'. The dataset consists of 600 mass spectra of chemical components, with 300 compounds contain the phenyl substructure and 300 compounds do not contain the substructure. Each spectrum contains 658 mass spectral features. Since a plot of the spectra of all samples is confusing, we only show the spectra of two instances in the Phenyl dataset, one for each class, in Figure 7.

We randomly select 100 samples from the Phenyl dataset for our experiments, with 50 contain the phenyl substructure and 50 do not contain the structure. These 100 instances are randomly partitioned into two equal subsets: a training set containing 50 samples (25 contain the phenyl substructure and 25 do not contain the substructure), and a test set containing 50 samples (25 contain the phenyl substructure and 25 do not contain the substructure).

### *3.1.3. The fat dataset*

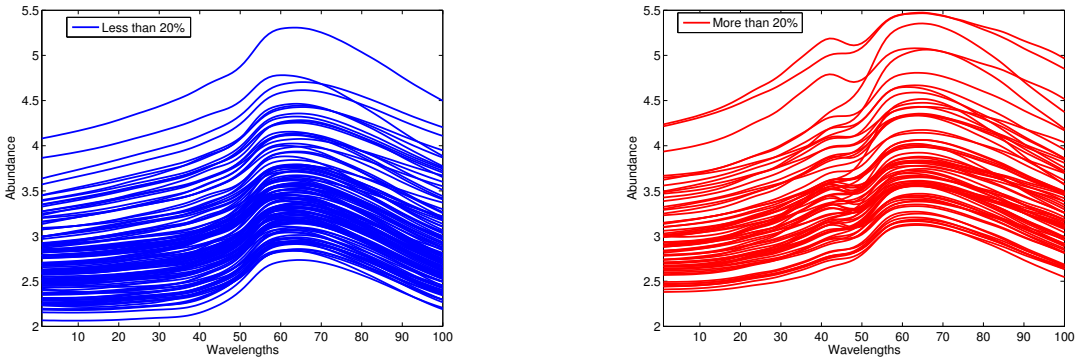



Figure 8: The spectra of two classes in the fat content dataset.

The fat content dataset [5] contains 193 spectra of finely chopped meat measured at 100 wavelengths, in which 122 meat samples contain less than 20% fat and 71 samples contain larger than 20% fat. The spectra of the data of the two classes are shown in Figure 8.

For this dataset, 100 samples are selected as a training set (50 samples with the fat content less than 20% and 50 samples with the fat content larger than 20%) and the remaining samples are selected as a test set.

### *3.2. Experiment settings*

The performances of the original SIMCA, the LDA-preprocessed SIMCA, the GDS-preprocessed SIMCA, and the DOS-preprocessed SIMCA are compared.

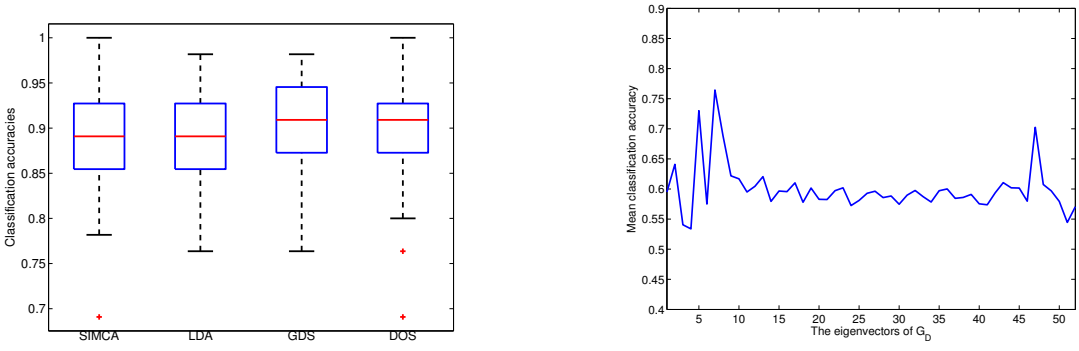
In SIMCA, the dimensions of the two class subspaces are tuned by 10-fold cross-validation. Before applying LDA, the high-dimensional spectral data are projected to the PC subspace of all available PCs. Then in LDA-preprocessed SIMCA, the dimensions of the two class subspaces are set to

zeros because only one discriminative direction can be found for two classes by LDA and this direction should be used for classification. In GDS and DOS, all the available PCs of each class subspace are used to obtain the generating matrix $\boldsymbol{G}_D$. In GDS, the dimension of GDS and the dimensions of the two class subspaces are also tuned by 10-fold cross-validation. The dimensions are chosen to minimise the classification error. In DOS, the discriminative order of the eigenvectors of $\boldsymbol{G}_D$ is determined by using the training set. Leave-one-out cross-validation (LOOCV) is used to obtain the classification accuracy of each eigenvector. The dimension of $\mathcal{D}_s$ and the dimensions of the two class subspaces are also tuned by 10-fold cross-validation. The dimensions are chosen to minimise the classification error, same as those for SIMCA and GDS.

All the experiments are repeated 100 times and the classification accuracies of all the experiments are recorded and depicted in boxplots.

### *3.3. Results*

#### *3.3.1. The meat dataset*




(a) Classification accuracies.

(b) Discriminative abilities.

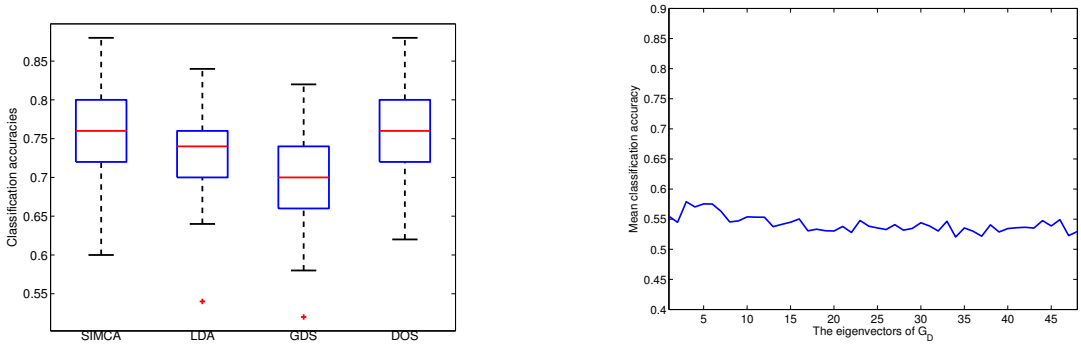
Figure 9: For the meat dataset: (a) classification accuracies of SIMCA, LDA, GDS and DOS; (b) discriminative abilities of the eigenvectors of the generating matrix $\boldsymbol{G}_D$.

Figure 9a shows the boxplots of the classification accuracies of the four methods for the meat dataset, from which we can observe that LDA performs similar to SIMCA while GDS and DOS both perform better than SIMCA.

Figure 9b shows the discriminative abilities of the eigenvectors of the generating matrix $\boldsymbol{G}_D$ versus the descending order of eigenvalues, which explains the good performance of GDS. That is, in Figure 9b, the horizontal axis shows the eigenvectors of $\boldsymbol{G}_D$ with eigenvalues in descending order and the vertical axis shows the corresponding average classification accuracies of SIMCA using the projected samples onto each of the eigenvectors. Since the first few eigenvectors of $\boldsymbol{G}_D$ do not have high discriminative abilities, discarding them, as done by GDS, can benefit classification, and thus GDS can provide good classification results.

In short, Figure 9 suggests that GDS performs well when the deletion of the first few eigenvectors (in terms of large eigenvalues) is beneficial for classification. In addition, DOS can achieve similarly good classification performance as GDS in this situation, as the first few eigenvectors are also not selected by DOS due to their low discriminative abilities.

441 *3.3.2. The Phenyl dataset*




(a) Classification accuracies. (b) Discriminative abilities.

Figure 10: For the Phenyl dataset: (a) classification accuracies of SIMCA, LDA, GDS and DOS; (b) discriminative abilities of the eigenvectors of the generating matrix $\boldsymbol{G}_D$.

442 As we have seen in Figure 4, GDS may fail to provide good classification
443 results in the cases of the Phenyl and fat datasets. Now we shall see that
444 DOS may provide good classification results even when GDS fails in these
445 cases.

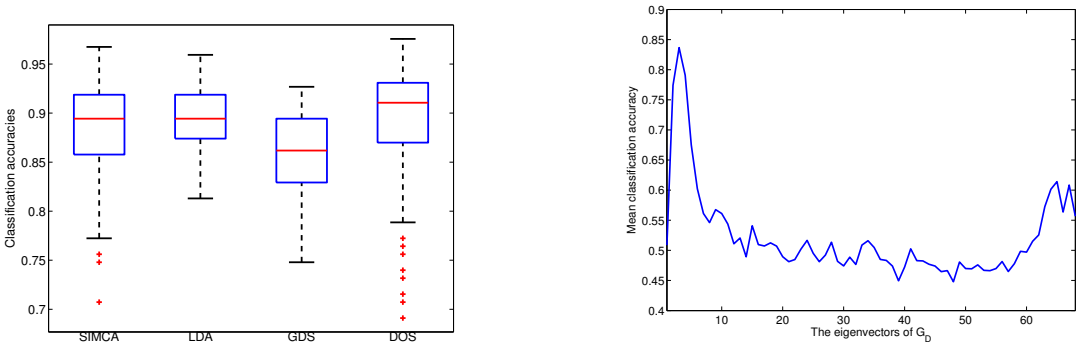
446 Figure 10a shows that GDS performs worse than SIMCA, which indicates
447 that the GDS projection is not a good preprocessing method for the Phenyl
448 dataset. LDA performs better than GDS, but worse than SIMCA. In con-
449 trast, DOS performs better than GDS and LDA, although only providing
450 similar classification accuracies as SIMCA in this case.

451 To explain this result, we can check Figure 10b, which shows the discrim-
452 inative abilities of the eigenvectors of $\boldsymbol{G}_D$ for the Phenyl dataset. On the
453 one hand, we observe that the first few eigenvectors with large eigenvalues
454 have higher discriminative abilities than the remaining ones. Thus deleting
455 the first few eigenvectors is harmful to classification. This explains why GDS
456 cannot provide good classification results. On the other hand, we also ob-

serve that the discriminative abilities of the eigenvectors are ranged from 0.52 to 0.58, which suggests that the discriminative abilities of the eigenvectors are similar to each other. Since the eigenvectors are similarly important to classification in this case, it is hard to achieve better classification by selecting from these eigenvectors. This explains why DOS performs similarly to SIMCA.

In summary, Figure 10 indicates that GDS fails to provide good classification results in the situation where the first few eigenvectors (in terms of large eigenvalues) of $\boldsymbol{G}_D$ are important for classification. DOS can provide better classification results than GDS in this situation. However, the classification results of DOS do not show noticeable improvement compared with those of SIMCA for this dataset, because the eigenvectors of $\boldsymbol{G}_D$ have similar discriminative abilities.

### *3.3.3. The fat dataset*




(a) Classification accuracies. (b) Discriminative abilities.

Figure 11: For the fat dataset: (a) classification accuracies of SIMCA, LDA, GDS and DOS; (b) discriminative abilities of the eigenvectors of the generating matrix $\boldsymbol{G}_D$.

Here we shall demonstrate that DOS can achieve better classification accuracies than SIMCA when the discriminative abilities of the eigenvectors of the generating matrix $\boldsymbol{G}_D$ have a large variation. In this situation, DOS can select the most discriminative eigenvectors to make the samples more separate and is a good preprocessing method for classification.

As shown in Figure 11a for the fat dataset, GDS performs worse than SIMCA and LDA, but DOS can achieve better performance than SIMCA and LDA.

Once again, let us use Figure 11b to explain the above results. On the one hand, because the discriminative abilities of the first few eigenvectors are higher than the remaining ones, GDS deletes the first few eigenvectors of $\boldsymbol{G}_D$ that are actually discriminative for classification, leading to a poor performance. On the other hand, Figure 11b shows that the discriminative abilities range from 0.45 to 0.85, which indicate a large difference in discriminative abilities between the eigenvectors. Hence DOS can select the most discriminative eigenvectors of $\boldsymbol{G}_D$ and provide better classification results than SIMCA.

To sum up, Figure 11 suggests that DOS performs well when there is a large difference in the discriminative abilities of the eigenvectors of the generating matrix $\boldsymbol{G}_D$. The good performance of DOS demonstrates that selecting the eigenvectors of $\boldsymbol{G}_D$ by using the discriminative ability instead of using eigenvalues can be effective, when GDS fails to provide improvement in classification.

### *3.3.4. Summary of experiments*

We would like to convey two messages through our experiments.

Firstly, from Figure 9b, Figure 10b and Figure 11b, we can observe that there is no negative correlation between eigenvalues and discriminative abilities of the eigenvectors of the generating matrix $\boldsymbol{G}_D$. The eigenvectors with large eigenvalues, although close to the intersection of two class subspaces, may have high discriminative abilities and can largely benefit classification of the samples of the two classes.

Secondly, from Figure 9a, Figure 10a and Figure 11a, we can observe that DOS can provide superior or at least comparable classification performance to SIMCA, LDA and GDS. The classification results suggest that it is appropriate to use high discriminative ability, instead of using low eigenvalues (or being away from the intersection of class subspaces), to select the eigenvectors of $\boldsymbol{G}_D$ to span a discriminative subspace for classification.

### *3.4. Discussion*

#### *3.4.1. Intersection of two class subspaces and its discriminative ability*

In Section 2.3.1, we have shown a motivating example that the intersection of two class subspaces can be discriminative for the fat dataset. In this section, we further investigate the relationship between the intersection and its discriminative ability for all the three datasets.

To check whether an eigenvector $\boldsymbol{v}_i$ is the intersection between class subspaces, we define $||\boldsymbol{e}_1||_2^2$ and $||\boldsymbol{e}_2||_2^2$ to measure the Euclidean distances from $\boldsymbol{v}_i$ to its projections in the two class subspaces, respectively. When $\boldsymbol{v}_i$ is in both class subspaces, it is the intersection of the two class subspaces. To be more specific, the Euclidean distances from $\boldsymbol{v}_i$ to its projections in the two class subspaces are zeros when $\boldsymbol{v}_i$ is the intersection. The larger the Euclidean distances, the farther $\boldsymbol{v}_i$ away from the two class subspaces.

Suppose the two class subspaces, $S(\boldsymbol{P}_1)$ and $S(\boldsymbol{P}_2)$, are defined by two projection matrices $\boldsymbol{P}_1 \in \mathbb{R}^{p\times p}$ and $\boldsymbol{P}_2 \in \mathbb{R}^{p\times p}$, respectively. The Euclidean distances from $\boldsymbol{v}_i$ to its projections in the two subspaces can be calculated as

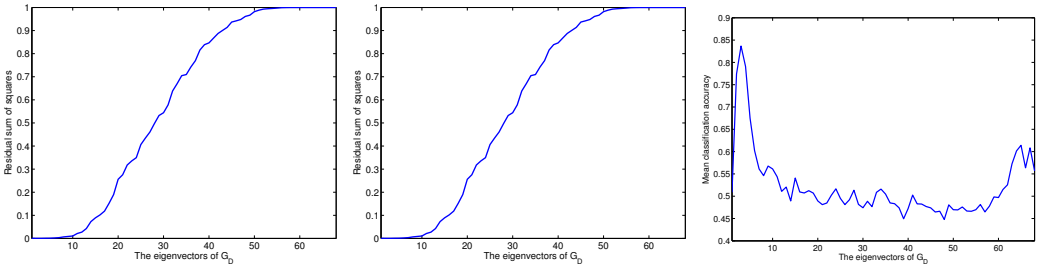
$$||\boldsymbol{e}_1||_2^2 = ||\boldsymbol{P}_1\boldsymbol{v}_i - \boldsymbol{v}_i||_2^2 \tag{11}$$

and

$$||\boldsymbol{e}_2||_2^2 = ||\boldsymbol{P}_2\boldsymbol{v}_i - \boldsymbol{v}_i||_2^2, \tag{12}$$

respectively. As $||\boldsymbol{e}_1||_2^2$ and $||\boldsymbol{e}_2||_2^2$ decrease, $\boldsymbol{v}_i$ goes closer to the two class subspaces and to the intersection. If $||\boldsymbol{e}_1||_2^2 = 0$ and $||\boldsymbol{e}_2||_2^2 = 0$, then $\boldsymbol{v}_i$ is the intersection of the two class subspaces, because $\boldsymbol{v}_i$ is in both subspaces, i.e. $\boldsymbol{P}_1\boldsymbol{v}_i = \boldsymbol{v}_i$ and $\boldsymbol{P}_2\boldsymbol{v}_i = \boldsymbol{v}_i$.

In the following part of this section, we discuss the relationship between the subspace intersection and its discriminative ability based on the values of $||\boldsymbol{e}_1||_2^2$, $||\boldsymbol{e}_2||_2^2$, and the corresponding discriminative abilities of the eigenvectors of $\boldsymbol{G}_D$.




(a) $||\boldsymbol{e}_1||_2^2$ (b) $||\boldsymbol{e}_2||_2^2$ (c) Discriminative ability

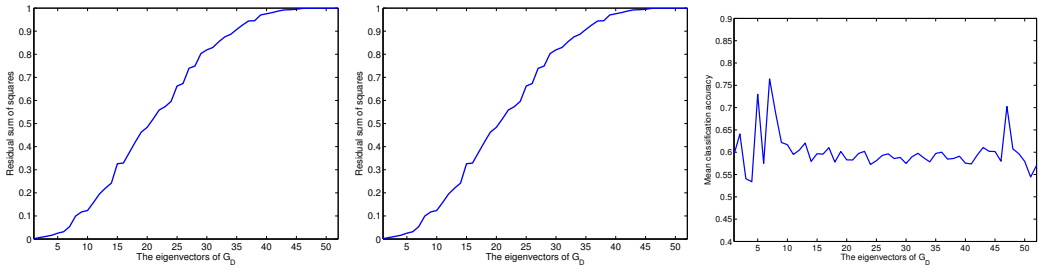
Figure 12: For the eigenvectors of $\boldsymbol{G}_D$ of the fat dataset: their distances ($||\boldsymbol{e}_1||_2^2$ and $||\boldsymbol{e}_2||_2^2$) to the two class subspaces, and their discriminative abilities.

As an extension of the motivating example in Section 2.3.1 for the fat

dataset, we present three plots in Figure 12 illustrating the relationship between the intersection of the two class subspaces and its discriminative ability.

Figure 12a and Figure 12b plot $||\boldsymbol{e}_1||_2^2$ and $||\boldsymbol{e}_2||_2^2$ against the descending order of eigenvalues, respectively. More specifically, in Figure 12a and Figure 12b, the horizontal axis lists the eigenvectors of $\boldsymbol{G}_D$ in the order of descending eigenvalues, and the vertical axis shows their values of $||\boldsymbol{e}_1||_2^2$ and $||\boldsymbol{e}_2||_2^2$. Figure 12c depicts the discriminative abilities of the eigenvectors, which is the same as Figure 11b.

We can clearly observe that the first few eigenvectors with the largest eigenvalues span the intersection of the two class subspaces of the fat dataset, because $||\boldsymbol{e}_1||_2^2$ and $||\boldsymbol{e}_2||_2^2$ of these eigenvectors are all zeros. However, we can also find that the corresponding discriminative abilities of these eigenvectors are higher compared with other eigenvectors, as shown in Figure 12c. That is, for the fat dataset, the intersection between the two class subspaces has high discriminative ability.

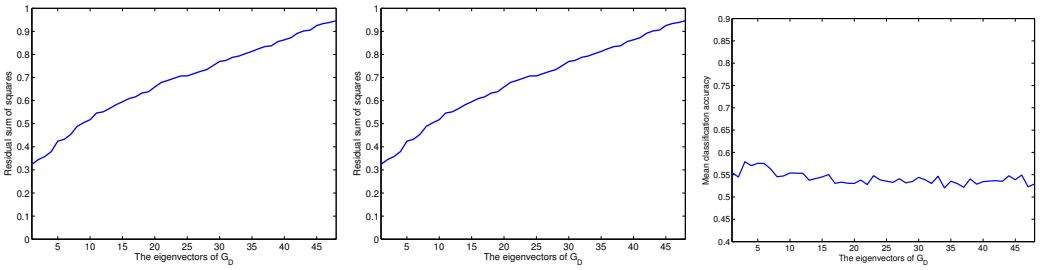



(a) $||\boldsymbol{e}_1||_2^2$ (b) $||\boldsymbol{e}_2||_2^2$ (c) Discriminative ability

Figure 13: For the eigenvectors of $\boldsymbol{G}_D$ of the meat dataset: their distances ($||\boldsymbol{e}_1||_2^2$ and $||\boldsymbol{e}_2||_2^2$) to the two class subspaces, and their discriminative abilities.

In contrast to the relationship observed in the fat dataset, here we shall see that the intersection can also have low discriminative ability.

The first eigenvector of the meat dataset is the intersection between the two class subspaces, as shown in Figure 13a and Figure 13b. The discriminative ability of this eigenvector is 0.6, which is low compared with many other eigenvectors. In other words, for the meat dataset, the intersection of the two class subspaces has low discriminative ability.




(a) $||\boldsymbol{e}_1||_2^2$ (b) $||\boldsymbol{e}_2||_2^2$ (c) Discriminative ability

Figure 14: For the eigenvectors of $\boldsymbol{G}_D$ of the Phenyl dataset: their distances ($||\boldsymbol{e}_1||_2^2$ and $||\boldsymbol{e}_2||_2^2$) to the two class subspaces, and their discriminative abilities.

Despite the two datasets discussed above that there exists intersection between class subspaces, now we show another dataset, the Phenyl dataset, that it is also possible that there is no intersection between two class subspaces.

We can observe from Figure 14a and Figure 14b that $||\boldsymbol{e}_1||_2^2$ and $||\boldsymbol{e}_2||_2^2$ of the first eigenvector are far from zeros. Thus there seems to be no intersection between the two class subspaces for the Phenyl dataset.

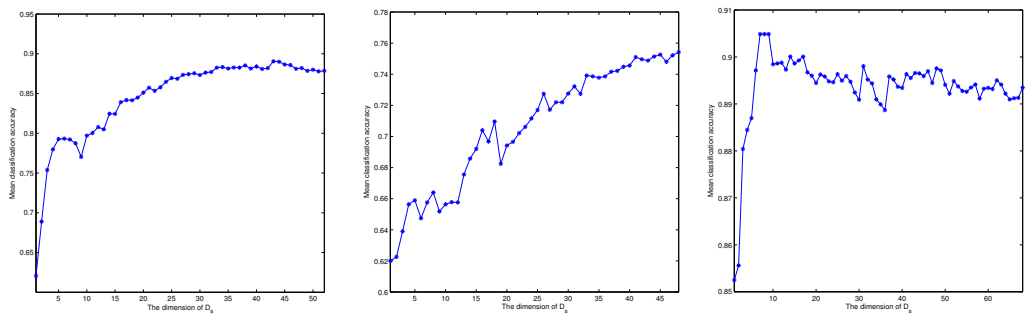
Therefore, we can draw two conclusions based on the observations from Figure 12, Figure 13, and Figure 14. First, the intersection between class subspaces does not always exist in all datasets. Second, even when the intersection exists, there is no definitely negative correlation between the intersection and its discriminative ability; that is, the discriminative ability of the

intersection of two class subspaces is data-dependent, not necessarily low.

The second conclusion above supports our argument that there is difference between a class subspace and a class. The intersection represents the same directions that two class subspaces can take, which can be discarded if we aim to classify two class subspaces. However, the intersection can be discriminative, and thus is important and cannot be simply discarded when we aim to classify the samples of two classes, which is actually the task of classification in practice.

### *3.4.2. Cross-validation of the dimension of the discriminatively ordered subspace*

In the DOS projection, the dimension of DOS $\mathcal{D}_s$ is an important parameter we need to tune. In this section, we discuss the effectiveness of using cross-validation to determine it.




(a) meat (b) Phenyl (c) fat

Figure 15: Effect of the dimension of $\mathcal{D}_s$.

Figure 15 plots the effect of the dimension of $\mathcal{D}_s$ on the classification accuracy on the test sets of the three real datasets, where the dimension changes

from one to the total number of eigenvectors in $\boldsymbol{V}_{sort}$. One hundred experiments of DOS are repeated for each dimension and the mean classification accuracies are plotted.

For the meat dataset, the dimension of $\mathcal{D}_s$ determined by 10-fold cross-validation in Section 3.3, which uses the training set only, ranges from 41 to 47 in the repeated experiments. Figure 15a shows a small peak of the mean classification accuracy of the test set around the dimension of 43, which is in line with the dimension determined by the training set-based 10-fold cross-validation.

For the fat dataset, the same effectiveness can be observed: the peak of the mean classification accuracy of the test set is around seven, as shown in Figure 15c, which is roughly consistent with the dimension (which is from two to seven) determined by using 10-fold cross-validation on the training set.

For the Phenyl dataset, Figure 15b does not show an obvious peak, and the mean classification accuracy of the test set seems to increase with the dimension and become stable when the dimension is larger than 41. The dimension determined by 10-fold cross-validation using the training set ranges from 38 to 43, which also conforms with the dimension of 41 in the test set.

In short, Figure 15 implies that the dimension of $\mathcal{D}_s$ determined by cross-validation using the training set is roughly consistent with the dimension with the largest mean classification accuracy of the test set. Thus cross-validation is an effective way to determine the dimension of $\mathcal{D}_s$ for the DOS projection.

## 4. Conclusion

SIMCA is a widely-used subspace method for classifying two-class high-dimensional spectral datasets. It suffers from the problem that the class subspaces are built independently without considering between-class information. This problem can be tackled by projecting the data to a subspace more discriminative than the original feature space before applying SIMCA. We have proposed a new method, the DOS projection, to generate such a discriminative subspace, by considering the between-class information and the discriminative ability of each basis vector of the subspace. The experiments on three real-world spectral datasets have demonstrated the effectiveness of the DOS projection.

Recently, subspace-based classification methods have been generalised to multi-view or tensor versions [14, 15, 16]. Inspired by these research, we aim to extend the DOS projection to multi-view or tensor versions in the future.

## References

[1] T. Arnalds, J. McElhinney, T. Fearn, G. Downey, A hierarchical discriminant analysis for species identification in raw meat by visible and near infrared spectroscopy, Journal of Near Infrared Spectroscopy 12 (3) (2004) 183–188.

[2] L. A. Berrueta, R. M. Alonso-Salces, K. Héberger, Supervised pattern recognition in food analysis, Journal of Chromatography A 1158 (1–2) (2007) 196–214.

[3] C. M. Bishop, Pattern Recognition and Machine Learning, Springer, 2006.

[4] G. Downey, Tutorial review. Qualitative analysis in the near-infrared region, Analyst 119 (11) (1994) 2367–2375.

[5] F. Ferraty, P. Vieu, Nonparametric Functional Data Analysis: Theory and Practice, Springer Science & Business Media, 2006.

[6] K. Fukui, A. Maki, Difference subspace and its generalization for subspace-based methods, IEEE Transactions on Pattern Analysis and Machine Intelligence 37 (11) (2015) 2164–2177.

[7] J. Holloway, T. Priya, A. Veeraraghavan, S. Prasad, Image classification in natural scenes: Are a few selective spectral channels sufficient?, in: 2014 IEEE International Conference on Image Processing (ICIP), IEEE, 2014, pp. 655–659.

[8] B. Mertens, M. Thompson, T. Fearn, Principal component outlier detection and SIMCA: a synthesis, Analyst 119 (1994) 2777–2784.

[9] Z. Pan, G. Healey, M. Prasad, B. Tromberg, Face recognition in hyperspectral images, IEEE Transactions on Pattern Analysis and Machine Intelligence 25 (12) (2003) 1552–1560.

[10] Y. Roggo, P. Chalus, L. Maurer, C. Lema-Martinez, A. Edmond, N. Jent, A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies, Journal of Pharmaceutical and Biomedical Analysis 44 (3) (2007) 683–700.

[11] C. W. Therrien, Eigenvalue properties of projection operators and their application to the subspace method of feature extraction, Computers, IEEE Transactions on 100 (9) (1975) 944–948.

[12] S. Wold, Pattern recognition by means of disjoint principal components models, Pattern Recognition 8 (3) (1976) 127–139.

[13] L. Zhang, L. Zhang, D. Tao, X. Huang, On combining multiple features for hyperspectral remote sensing image classification, IEEE Transactions on Geoscience and Remote Sensing 50 (3) (2012) 879–893.

[14] L. Zhang, L. Zhang, D. Tao, X. Huang, Tensor discriminative locality alignment for hyperspectral image spectral–spatial feature extraction, IEEE Transactions on Geoscience and Remote Sensing 51 (1) (2013) 242–256.

[15] L. Zhang, Q. Zhang, L. Zhang, D. Tao, X. Huang, B. Du, Ensemble manifold regularized sparse low-rank approximation for multiview feature embedding, Pattern Recognition 48 (10) (2015) 3102–3112.

[16] L. Zhang, X. Zhu, L. Zhang, B. Du, Multidomain subspace classification for hyperspectral images, IEEE Transactions on Geoscience and Remote Sensing 54 (10) (2016) 6138–6150.