

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

Rui Zhu<sup>a</sup>, Kazuhiro Fukui<sup>b</sup>, Jing-Hao Xue<sup>a,\*</sup>

<sup>a</sup>*Department of Statistical Science, University College London, London WC1E 6BT, UK*

<sup>b</sup>*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](mailto: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.

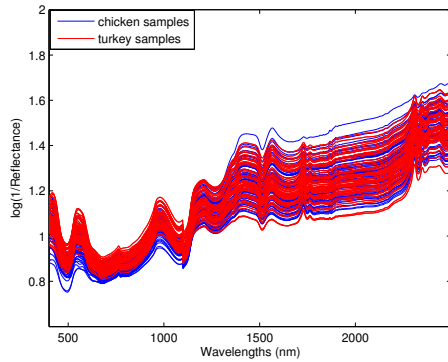


Figure 1: Spectra of meat samples from two classes: chicken and turkey.

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

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

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

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

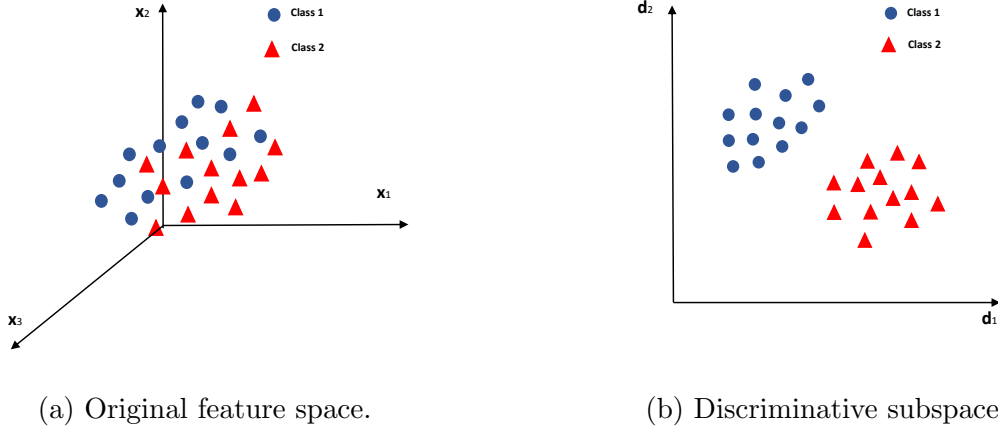


Figure 2: (a) Two classes of samples are mixed together in the original 3-dimensional feature space. (b) The same groups of samples can be well separated when they are projected to a discriminative 2-dimensional subspace.

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

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

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

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

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

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

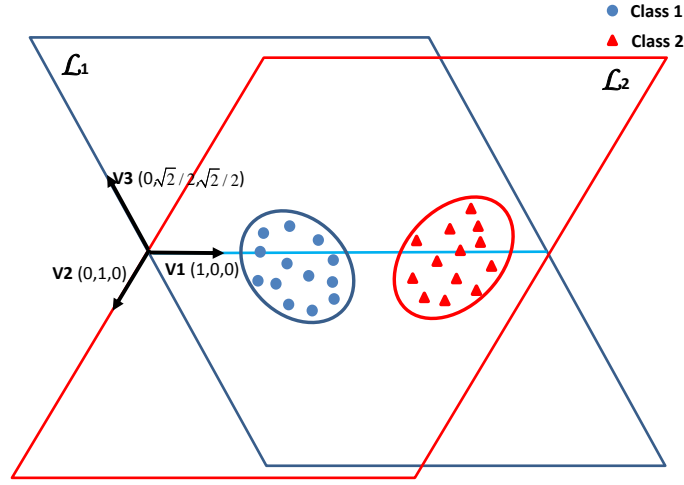


Figure 3: An illustrative example of the difference between a class subspace (of infinite scale) and a class (of finite scale).

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

discards  $\mathbf{v}_1$  because it is the eigenvector of  $\mathbf{G}_D$  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  $\mathbf{v}_1$ , which suggests that  $\mathbf{v}_1$  contains discriminative information between classes. (We shall demonstrate another motivating example for this issue in Section 2.3.1 using a real spectral dataset.)

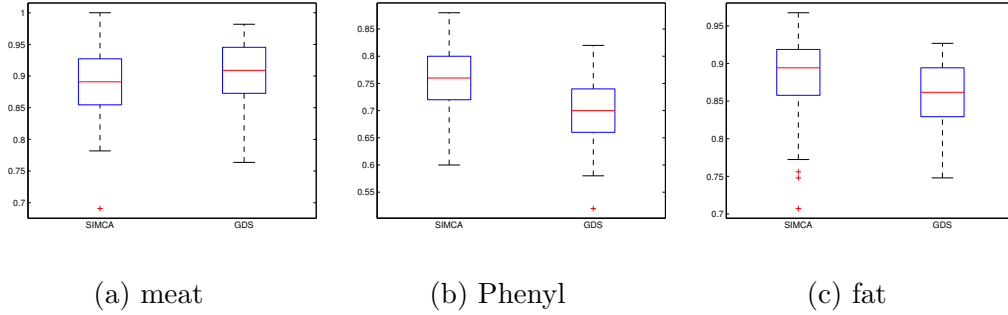


Figure 4: Classification accuracies of SIMCA and the GDS-preprocessed SIMCA on three real spectral datasets: meat, Phenyl and fat. In each panel, the left-hand boxplot is for SIMCA, and the right-hand boxplot is for the GDS-preprocessed SIMCA.

Moreover, here we illustrate that discarding the eigenvectors of  $\mathbf{G}_D$  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  $\mathbf{G}_D$  and to overcome

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

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



## 124 2. Methodology

### 125 2.1. SIMCA

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

$$\text{Cov}(\mathbf{X}_k) = \frac{1}{N_k - 1} (\mathbf{X}_k^c)^T \mathbf{X}_k^c = \mathbf{U}_k \mathbf{\Sigma} \mathbf{U}_k^{-1}, \quad (1)$$

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

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

$$\mathbf{P}_k = \mathbf{W}_k \mathbf{W}_k^T. \quad (2)$$

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

$$\mathbf{E}_k = \mathbf{X}_k^c - \mathbf{X}_k^c \mathbf{P}_k. \quad (3)$$

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

$$\mathbf{e}_{k,new} = \mathbf{x}_{new}^c - \mathbf{x}_{new}^c \mathbf{P}_k, \quad (4)$$

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

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

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

## 144 2.2. Generalised difference subspace

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

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

159 classified to the class with the smallest canonical angle.

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

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

### 179 2.2.1. GDS

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

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

$$\mathbf{G}_D = \sum_{k=1}^K \mathbf{P}_k, \quad (6)$$

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

$$\mathbf{G}_D = \mathbf{V}_D \mathbf{\Lambda}_D \mathbf{V}_D^T, \quad (7)$$

184 where the columns in  $\mathbf{V}_D = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{r_D}] \in \mathbb{R}^{p \times r_D}$  are the normalised  
 185 eigenvectors of  $\mathbf{G}_D$ , and  $\mathbf{\Lambda}_D$  denotes the diagonal matrix with correspond-  
 186 ing eigenvalues  $\{\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{r_D}\}$  in descending order, where  $r_D =$   
 187  $\text{rank}(\mathbf{G}_D)$ .

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

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

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

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

### 207 2.2.2. The generating matrix

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

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

$$\mathbf{G} = \sum_{k=1}^K \alpha_k \mathbf{P}_k, \quad (8)$$

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

$$\mathbf{G} = \mathbf{V}_G \mathbf{\Lambda}_G \mathbf{V}_G^T, \quad (9)$$

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

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

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

225 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 sub-  
 226 space that contains the most similar directions of the two class subspaces.  
 227 In other words, the most similar directions of the two class subspaces are  
 228 extracted by the eigenvectors of  $\mathbf{G}$  with eigenvalues of one. In contrast, the  
 229 eigenvectors with eigenvalues of zeros are the complements of the sum sub-  
 230 space which contain information that is irrelevant to the two class subspaces.  
 231 The larger the eigenvalue, the more similarity information the corresponding  
 232 eigenvector contains.

233 The generation of GDS is closely related to the generating matrix:  $\mathbf{G}_D$   
 234 and  $\mathbf{G}$  are both linear combinations of  $\mathbf{P}_k$  although with different coefficients.  
 235 The linear coefficients of  $\mathbf{G}_D$  are all one, i.e.  $\alpha_k = 1 \ \forall \ k$ , while those of  $\mathbf{G}$   
 236 are constrained by  $\alpha_k \in (0, 1)$  and  $\sum_{k=1}^K \alpha_k = 1$ . Although  $\mathbf{G}_D$  and  $\mathbf{G}$  are  
 237 slightly different, we can derive similar properties of  $\mathbf{G}_D$  as those of  $\mathbf{G}$  by  
 238 following the proofs in [11]. First, the eigenvalues of  $\mathbf{G}_D$  are in the interval  
 239  $[0, 2]$ . Second, the eigenvectors with the corresponding eigenvalues of two  
 240 span the intersection of the two subspaces  $\bigcap_{k=1}^2 \mathcal{P}_k$ . Third, the eigenvectors  
 241 with the corresponding eigenvalues that are nonzero span the sum subspace  
 242 of the two subspace and those with zero eigenvalues span the complement of  
 243 the sum subspace. Hence, with some abuse of notation, we also call the sum  
 244 matrix  $\mathbf{G}_D$  a generating matrix.

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

247 ment seems to be consistent with the property of  $\mathbf{G}_D$ , based on the assump-  
 248 tion that the eigenvectors closed to the intersection directions contain large  
 249 amount of similarity information. Since the eigenvectors with eigenvalues of  
 250 two span the intersections subspace, the eigenvectors with eigenvalues close  
 251 to two could be close to the intersection directions. On the other hand, the  
 252 eigenvectors with eigenvalues far from two, i.e. eigenvalues in  $[0, 1)$ , are far  
 253 from the intersection directions. Therefore, the GDS projection aims to dis-  
 254 card the eigenvectors that are close to the intersection directions, so as to  
 255 provide a discriminative subspace.

### 256 2.3. Discriminatively ordered subspace

257 The GDS projection is based on the assumption that, because the first  
 258 few eigenvectors with large eigenvalues close to the intersection directions  
 259 contain similarity information between the class subspaces, they are not im-  
 260 portant for classification. However, this assumption is not always true, as a  
 261 class subspace (of infinite scale) and a class (of finite scale) are different, and  
 262 hence the ability to discriminate two class subspaces are not necessarily in  
 263 line with the ability to discriminate samples of two classes. In the extreme  
 264 case, two separable classes may span the same class subspace. More techni-  
 265 cally, the similarity information in the GDS assumption only considers the  
 266 directions, while the scores or the projection values on the directions should  
 267 also be considered. The eigenvectors of  $\mathbf{G}_D$  that are close to the intersection  
 268 directions between the two class subspaces can be discriminative when the  
 269 scores on these eigenvectors are largely separable between classes. In the  
 270 following section, we show a motivating real-data example that even the di-  
 271 rections in the intersection subspace of the two classes can be discriminative.

272 *2.3.1. Intersection and discriminative ability: a motivating example*

273 The fat dataset contains 193 spectra of finely chopped meat measured  
 274 at 100 wavelengths, in which 122 samples contain less than 20% fat and 71  
 275 samples contain more than 20% fat. Detailed description of this dataset can  
 276 be found in Section 3.1. We split the dataset into a training set and a test set:  
 277 35 samples with fat content less than 20% and 35 samples with fat content  
 278 more than 20% are randomly sampled into the training set; the rest samples  
 279 form the test set.

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

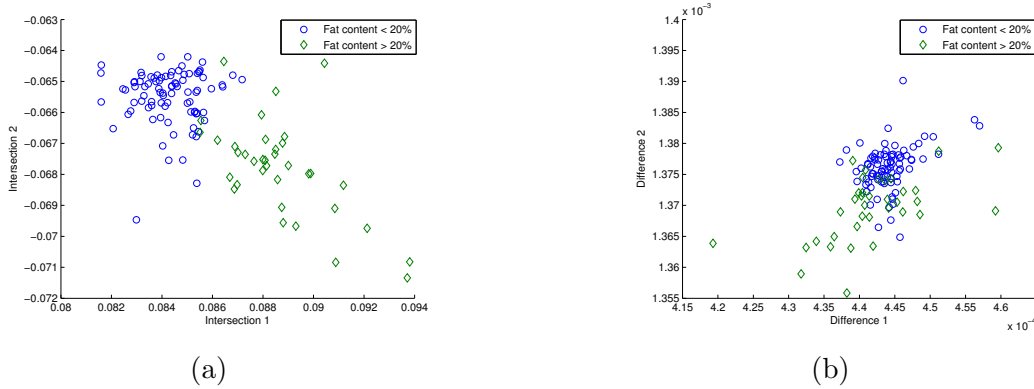


Figure 5: (a) Projections of the test samples onto two directions of the intersection. (b) Projections of the test samples onto two directions of the DS.

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



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

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

### 299 *2.3.2. Discriminatively ordered subspace*

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

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

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

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

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

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

336 Secondly, we apply SIMCA to each validation by setting the dimensions  
 337 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  $\mathbf{v}_i$ , we obtain  $N_1 + N_2$  predictions from LOOCV. The classification accuracy  $d_i$  is calculated as

$$d_i = \frac{N_c}{N_1 + N_2} , \quad (10)$$

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

Fourthly, after obtaining  $d'_i$ s for  $i = 1, \dots, r_D$ , we sort the eigenvectors  $\mathbf{v}'_i$ s in ascending order of  $d'_i$ s and obtain the matrix of the sorted eigenvectors  $\mathbf{V}_{sort} = [\mathbf{v}_{(1)}, \mathbf{v}_{(2)}, \dots, \mathbf{v}_{(r_D)}]$ , where the discriminative ability  $d_{(1)} < d_{(2)} < \dots < d_{(r_D)}$ . The last few eigenvectors in  $\mathbf{V}_{sort}$  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

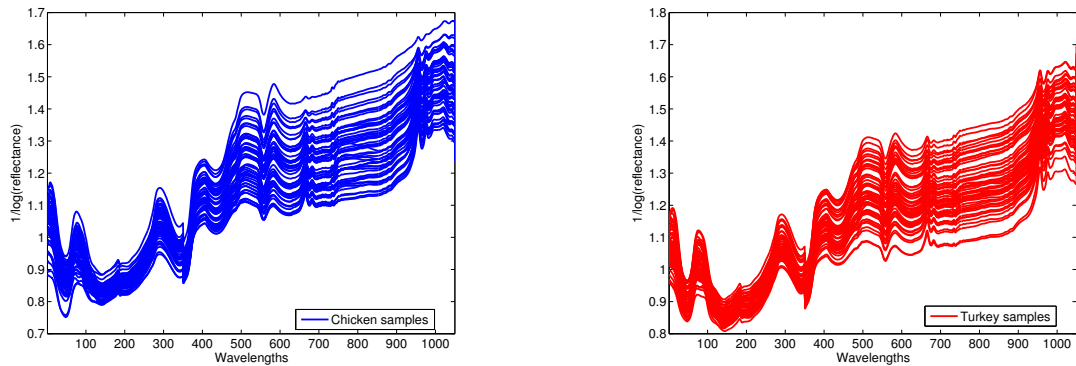


Figure 6: The spectra of the two classes in the meat dataset.

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

376 turkey samples are randomly partitioned into a training set (27 chicken sam-  
 377 ples and 27 turkey samples) and a test set (28 chicken samples and 27 turkey  
 378 samples).

### 379 3.1.2. The Phenyl dataset

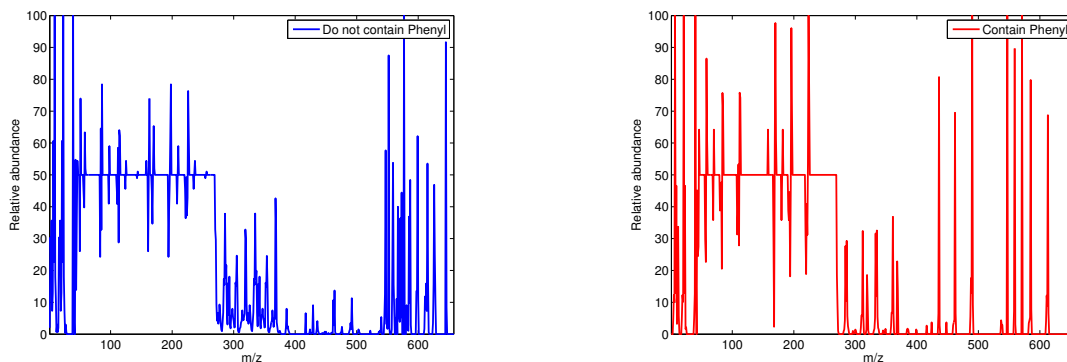


Figure 7: The spectra of the two classes in the Phenyl dataset.

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

386 We randomly select 100 samples from the Phenyl dataset for our exper-  
 387 iments, with 50 contain the phenyl substructure and 50 do not contain the  
 388 structure. These 100 instances are randomly partitioned into two equal sub-  
 389 sets: a training set containing 50 samples (25 contain the phenyl substructure  
 390 and 25 do not contain the substructure), and a test set containing 50 samples  
 391 (25 contain the phenyl substructure and 25 do not contain the substructure).

### 3.1.3. The fat dataset

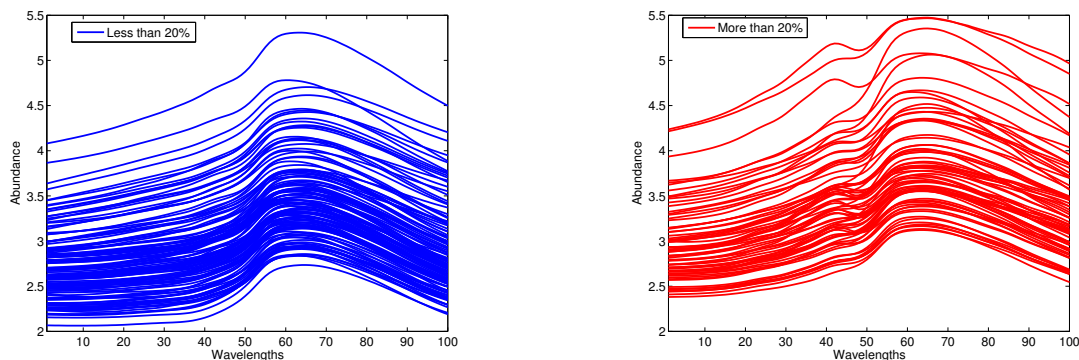


Figure 8: The spectra of two classes in the fat content dataset.

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.

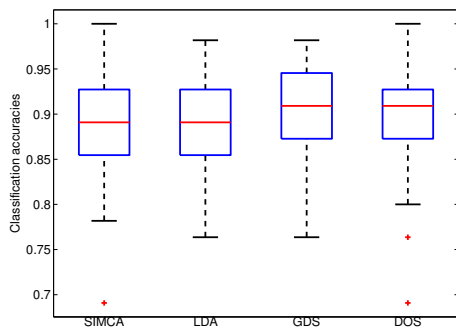
In SIMCA, the dimensions of the two class subspaces are tuned by 10-fold cross-validation. Before applying LDA, the high-dimensional spectral data are projected to the PC subspace of all available PCs. Then in LDA-preprocessed SIMCA, the dimensions of the two class subspaces are set to

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

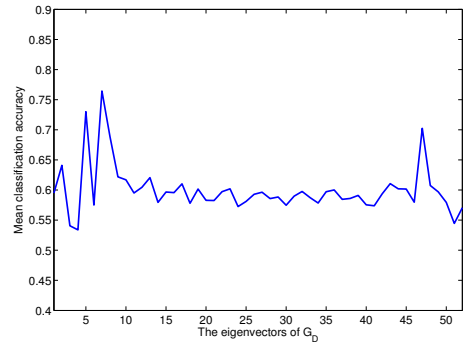
420 All the experiments are repeated 100 times and the classification accura-  
 421 cies of all the experiments are recorded and depicted in boxplots.

### 422 3.3. Results

#### 423 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  $\mathbf{G}_D$ .

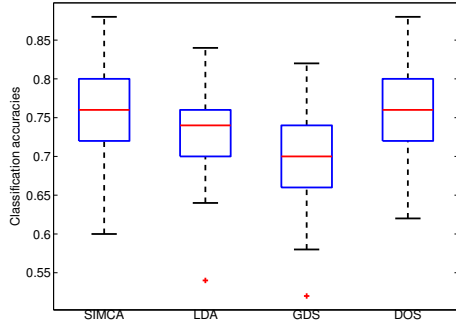
424 Figure 9a shows the boxplots of the classification accuracies of the four  
425 methods for the meat dataset, from which we can observe that LDA performs  
426 similar to SIMCA while GDS and DOS both perform better than SIMCA.

427 Figure 9b shows the discriminative abilities of the eigenvectors of the  
428 generating matrix  $\mathbf{G}_D$  versus the descending order of eigenvalues, which ex-  
429 plains the good performance of GDS. That is, in Figure 9b, the horizontal  
430 axis shows the eigenvectors of  $\mathbf{G}_D$  with eigenvalues in descending order and  
431 the vertical axis shows the corresponding average classification accuracies of  
432 SIMCA using the projected samples onto each of the eigenvectors. Since the  
433 first few eigenvectors of  $\mathbf{G}_D$  do not have high discriminative abilities, dis-  
434 carding them, as done by GDS, can benefit classification, and thus GDS can  
435 provide good classification results.

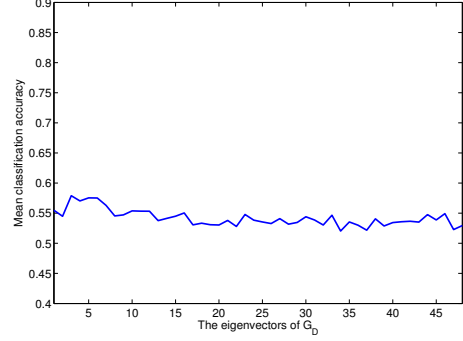
436 In short, Figure 9 suggests that GDS performs well when the deletion  
437 of the first few eigenvectors (in terms of large eigenvalues) is beneficial for  
438 classification. In addition, DOS can achieve similarly good classification  
439 performance as GDS in this situation, as the first few eigenvectors are also  
440 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  $\mathbf{G}_D$ .

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.

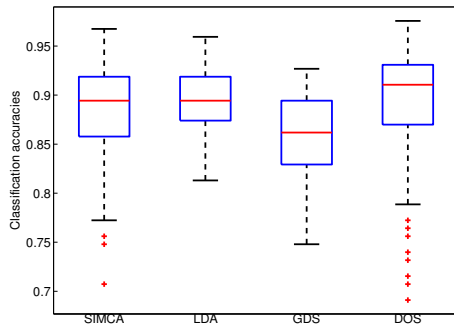
446 Figure 10a shows that GDS performs worse than SIMCA, which indicates  
 447 that the GDS projection is not a good preprocessing method for the Phenyl  
 448 dataset. LDA performs better than GDS, but worse than SIMCA. In con-  
 449 trast, DOS performs better than GDS and LDA, although only providing  
 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  $\mathbf{G}_D$  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-

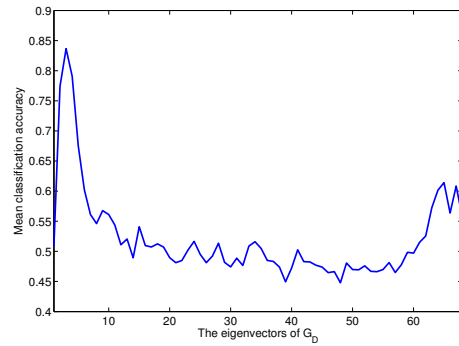
serve that the discriminative abilities of the eigenvectors are ranged from 0.52 to 0.58, which suggests that the discriminative abilities of the eigenvectors are similar to each other. Since the eigenvectors are similarly important to classification in this case, it is hard to achieve better classification 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  $\mathbf{G}_D$  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  $\mathbf{G}_D$  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  $\mathbf{G}_D$ .

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

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

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

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

#### 494 3.3.4. *Summary of experiments*

495 We would like to convey two messages through our experiments.

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

502 Secondly, from Figure 9a, Figure 10a and Figure 11a, we can observe  
 503 that DOS can provide superior or at least comparable classification perfor-  
 504 mance to SIMCA, LDA and GDS. The classification results suggest that it  
 505 is appropriate to use high discriminative ability, instead of using low eigen-  
 506 values (or being away from the intersection of class subspaces), to select the  
 507 eigenvectors of  $\mathbf{G}_D$  to span a discriminative subspace for classification.

### 508 3.4. Discussion

#### 509 3.4.1. Intersection of two class subspaces and its discriminative ability

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

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

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

$$\|\mathbf{e}_1\|_2^2 = \|\mathbf{P}_1 \mathbf{v}_i - \mathbf{v}_i\|_2^2 \quad (11)$$

and

$$\|\mathbf{e}_2\|_2^2 = \|\mathbf{P}_2 \mathbf{v}_i - \mathbf{v}_i\|_2^2, \quad (12)$$

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

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

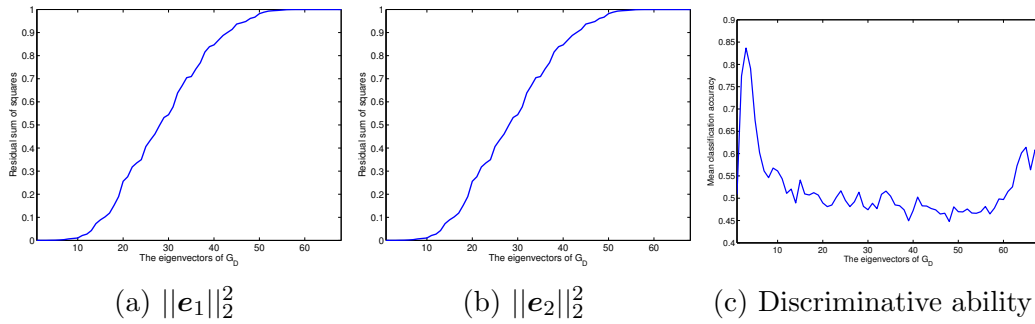


Figure 12: For the eigenvectors of  $\mathbf{G}_D$  of the fat dataset: their distances ( $\|\mathbf{e}_1\|_2^2$  and  $\|\mathbf{e}_2\|_2^2$ ) to the two class subspaces, and their discriminative abilities.

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  $\|\mathbf{e}_1\|_2^2$  and  $\|\mathbf{e}_2\|_2^2$  against the descending order of eigenvalues, respectively. More specifically, in Figure 12a and Figure 12b, the horizontal axis lists the eigenvectors of  $\mathbf{G}_D$  in the order of descending eigenvalues, and the vertical axis shows their values of  $\|\mathbf{e}_1\|_2^2$  and  $\|\mathbf{e}_2\|_2^2$ . 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  $\|\mathbf{e}_1\|_2^2$  and  $\|\mathbf{e}_2\|_2^2$  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.

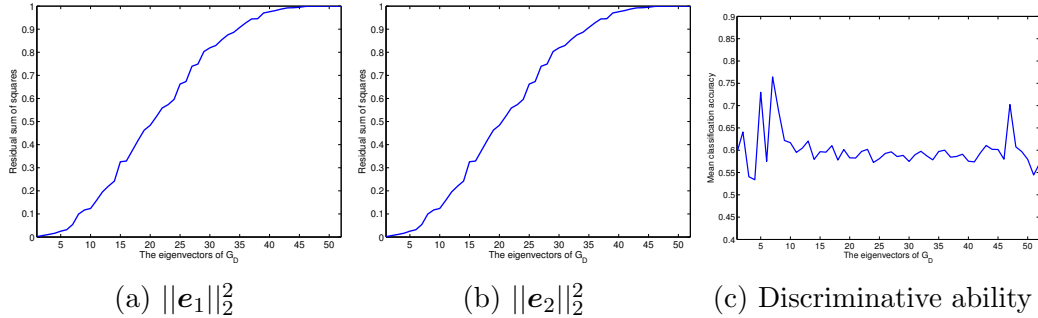


Figure 13: For the eigenvectors of  $\mathbf{G}_D$  of the meat dataset: their distances ( $\|\mathbf{e}_1\|_2^2$  and  $\|\mathbf{e}_2\|_2^2$ ) to the two class subspaces, and their discriminative abilities.

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

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

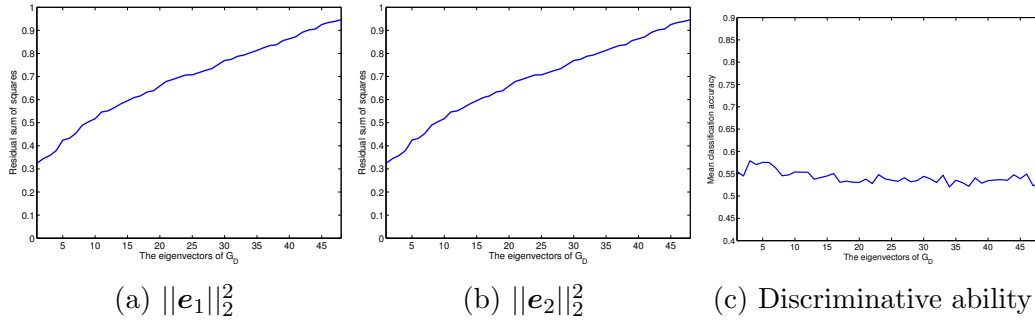


Figure 14: For the eigenvectors of  $G_D$  of the Phenyl dataset: their distances ( $\|e_1\|_2^2$  and  $\|e_2\|_2^2$ ) to the two class subspaces, and their discriminative abilities.

557 Despite the two datasets discussed above that there exists intersection  
 558 between class subspaces, now we show another dataset, the Phenyl dataset,  
 559 that it is also possible that there is no intersection between two class sub-  
 560 spaces.

561 We can observe from Figure 14a and Figure 14b that  $\|e_1\|_2^2$  and  $\|e_2\|_2^2$  of  
 562 the first eigenvector are far from zeros. Thus there seems to be no intersection  
 563 between the two class subspaces for the Phenyl dataset.

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

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

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

### 577 3.4.2. Cross-validation of the dimension of the discriminatively ordered sub- 578 space

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

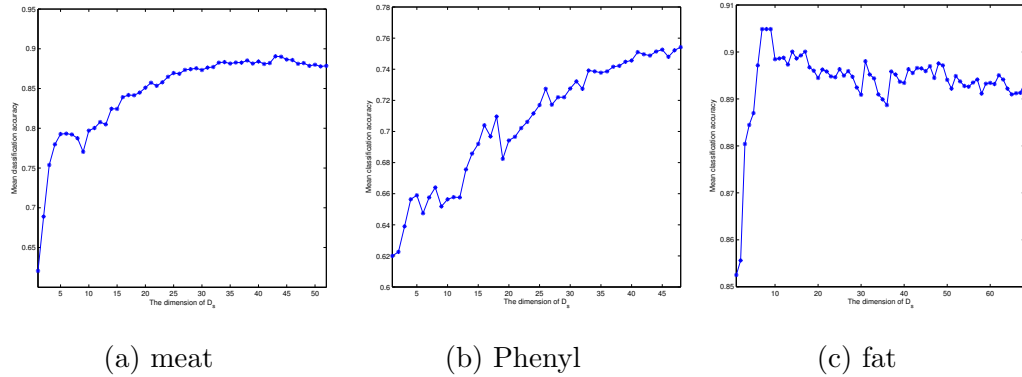


Figure 15: Effect of the dimension of  $\mathcal{D}_s$ .

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



584 from one to the total number of eigenvectors in  $\mathbf{V}_{sort}$ . One hundred exper-  
 585 iments of DOS are repeated for each dimension and the mean classification  
 586 accuracies are plotted.

587 For the meat dataset, the dimension of  $\mathcal{D}_s$  determined by 10-fold cross-  
 588 validation in Section 3.3, which uses the training set only, ranges from 41 to  
 589 47 in the repeated experiments. Figure 15a shows a small peak of the mean  
 590 classification accuracy of the test set around the dimension of 43, which is in  
 591 line with the dimension determined by the training set-based 10-fold cross-  
 592 validation.

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

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

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

## 608 4. Conclusion

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

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

## 622 References

- 623 [1] T. Arnalds, J. McElhinney, T. Fearn, G. Downey, A hierarchical dis-  
624 criminant analysis for species identification in raw meat by visible and  
625 near infrared spectroscopy, *Journal of Near Infrared Spectroscopy* 12 (3)  
626 (2004) 183–188.
- 627 [2] L. A. Berrueta, R. M. Alonso-Salces, K. Héberger, Supervised pattern  
628 recognition in food analysis, *Journal of Chromatography A* 1158 (1–2)  
629 (2007) 196–214.

- 630 [3] C. M. Bishop, Pattern Recognition and Machine Learning, Springer,  
631 2006.
- 632 [4] G. Downey, Tutorial review. Qualitative analysis in the near-infrared  
633 region, *Analyst* 119 (11) (1994) 2367–2375.
- 634 [5] F. Ferraty, P. Vieu, Nonparametric Functional Data Analysis: Theory  
635 and Practice, Springer Science & Business Media, 2006.
- 636 [6] K. Fukui, A. Maki, Difference subspace and its generalization for  
637 subspace-based methods, *IEEE Transactions on Pattern Analysis and*  
638 *Machine Intelligence* 37 (11) (2015) 2164–2177.
- 639 [7] J. Holloway, T. Priya, A. Veeraraghavan, S. Prasad, Image classification  
640 in natural scenes: Are a few selective spectral channels sufficient?, in:  
641 2014 IEEE International Conference on Image Processing (ICIP), IEEE,  
642 2014, pp. 655–659.
- 643 [8] B. Mertens, M. Thompson, T. Fearn, Principal component outlier de-  
644 tection and SIMCA: a synthesis, *Analyst* 119 (1994) 2777–2784.
- 645 [9] Z. Pan, G. Healey, M. Prasad, B. Tromberg, Face recognition in hyper-  
646 spectral images, *IEEE Transactions on Pattern Analysis and Machine*  
647 *Intelligence* 25 (12) (2003) 1552–1560.
- 648 [10] Y. Roggo, P. Chalus, L. Maurer, C. Lema-Martinez, A. Edmond,  
649 N. Jent, A review of near infrared spectroscopy and chemometrics in  
650 pharmaceutical technologies, *Journal of Pharmaceutical and Biomed-*  
651 *ical Analysis* 44 (3) (2007) 683–700.

- 652 [11] C. W. Therrien, Eigenvalue properties of projection operators and their  
 653 application to the subspace method of feature extraction, *Computers,*  
 654 *IEEE Transactions on* 100 (9) (1975) 944–948.
  
- 655 [12] S. Wold, Pattern recognition by means of disjoint principal components  
 656 models, *Pattern Recognition* 8 (3) (1976) 127–139.
  
- 657 [13] L. Zhang, L. Zhang, D. Tao, X. Huang, On combining multiple features  
 658 for hyperspectral remote sensing image classification, *IEEE Transactions*  
 659 *on Geoscience and Remote Sensing* 50 (3) (2012) 879–893.
  
- 660 [14] L. Zhang, L. Zhang, D. Tao, X. Huang, Tensor discriminative locality  
 661 alignment for hyperspectral image spectral–spatial feature extraction,  
 662 *IEEE Transactions on Geoscience and Remote Sensing* 51 (1) (2013)  
 663 242–256.
  
- 664 [15] L. Zhang, Q. Zhang, L. Zhang, D. Tao, X. Huang, B. Du, Ensemble  
 665 manifold regularized sparse low-rank approximation for multiview fea-  
 666 ture embedding, *Pattern Recognition* 48 (10) (2015) 3102–3112.
  
- 667 [16] L. Zhang, X. Zhu, L. Zhang, B. Du, Multidomain subspace classification  
 668 for hyperspectral images, *IEEE Transactions on Geoscience and Remote*  
 669 *Sensing* 54 (10) (2016) 6138–6150.