PMI RESEARCH & DEVELOPMENT

Discrete wavelet preconditioning of Krylov spaces and PLS regression

Athanassios Kondylis [1] and Joe Whittaker [2]

CompStat 2010, Paris

[1]Philip Morris International, R&D, Computational Plant Biology, Switzerland
[2]Lancaster University, Department of Mathematics and Statistics, UK

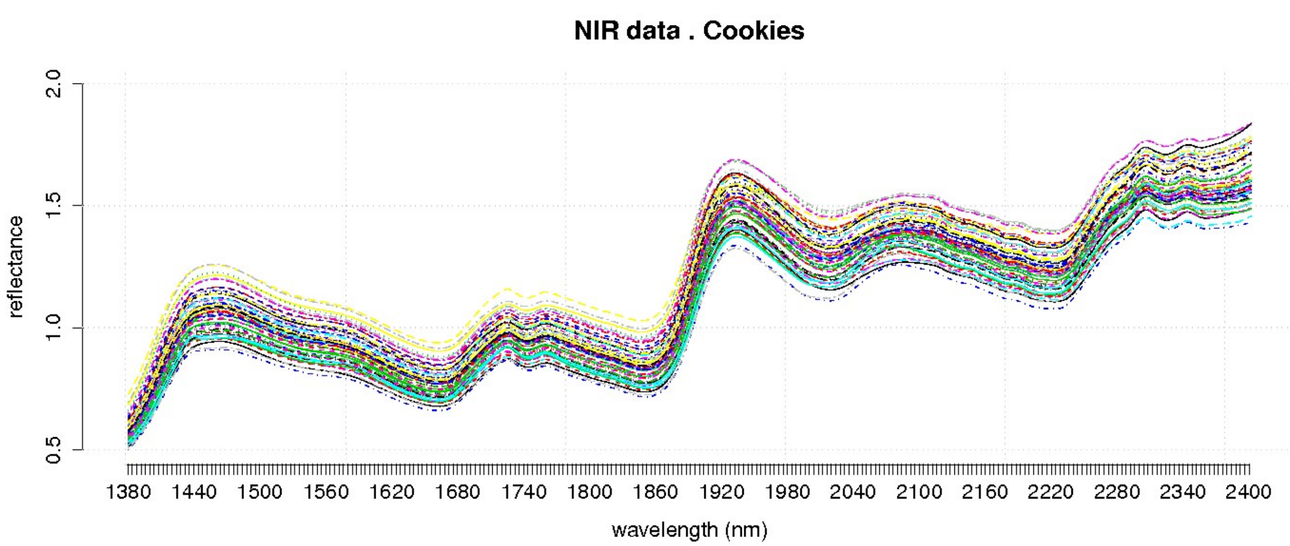
PMI RESEARCH & DEVELOPMENT

the regression problem

use high throughput spectral data (NMR, GC-MS, NIR) :

$$\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p), \quad \boldsymbol{x}_j \in R^n, \quad j = 1, \ldots, p < n$$

to predict the response(s) of interest : $\boldsymbol{Y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_q), \quad q < p$

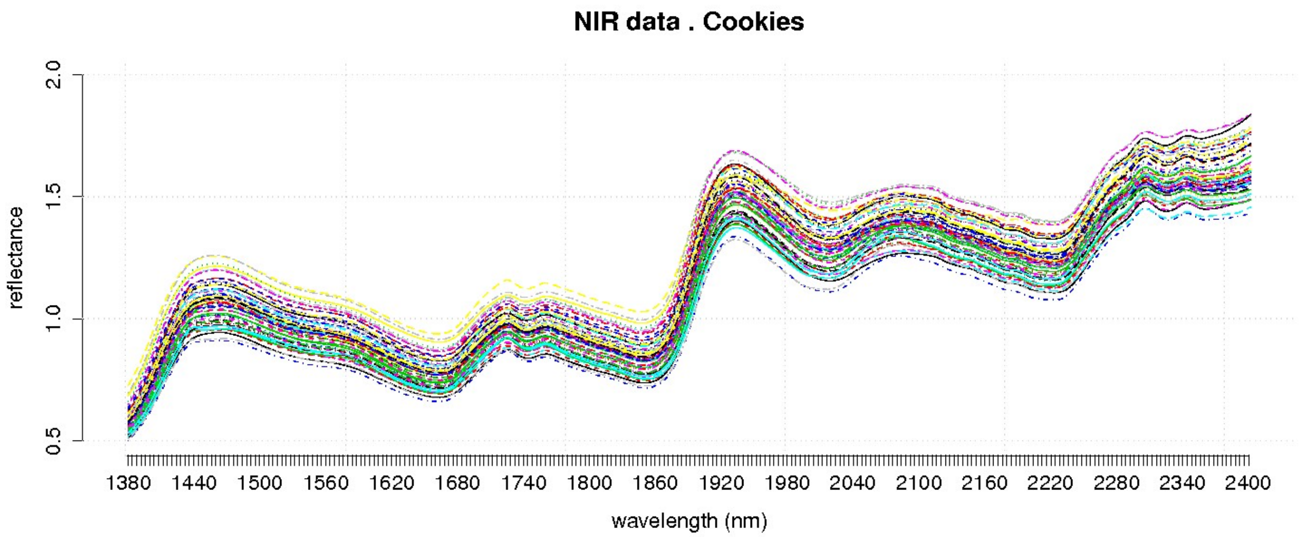


the regression problem

focus on a single response $q = 1$

deal with high dimensionality of the data

take into account the spectral form of the data



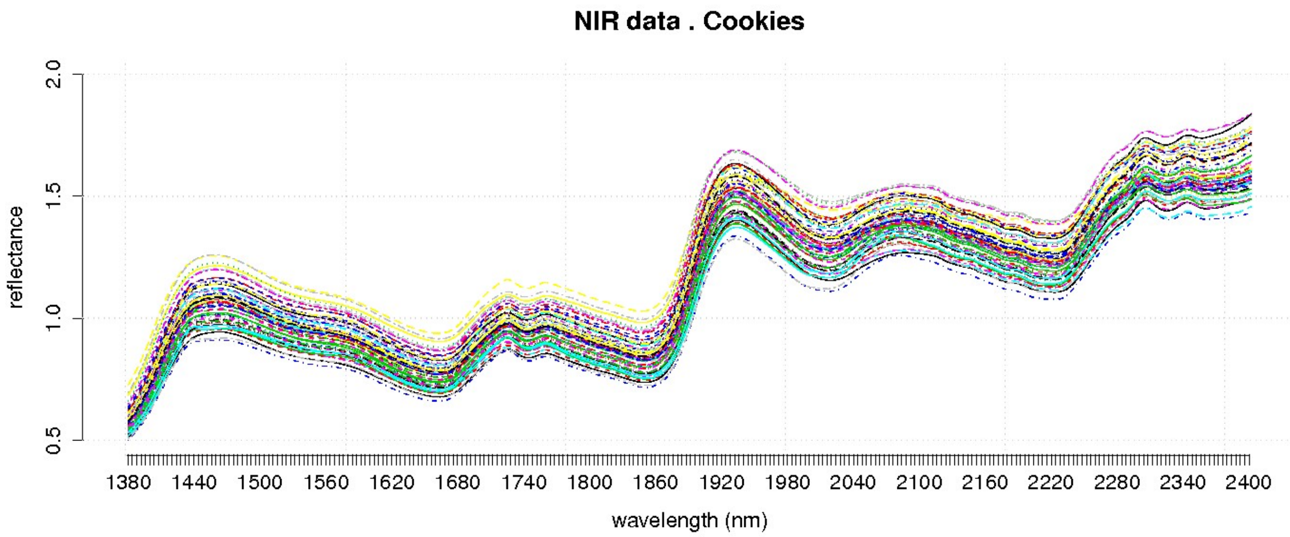
the regression problem

focus on a single response $q = 1$

deal with high dimensionality of the data

take into account the spectral form of the data

find spectral regions relevant for prediction



PLS regression

Solve the normal equations :

$$\frac{1}{n}\,\mathsf{A}\,\boldsymbol{\beta} = \frac{1}{n}\,\mathsf{b}, \ \ \text{for} \ \ \mathsf{A} = \boldsymbol{X}'\boldsymbol{X}, \ \mathsf{b} = \boldsymbol{X}'\boldsymbol{y}$$

The PLS regression coefficient $\widehat{\boldsymbol{\beta}}_m^{pls}$ is a Krylov solution :

$$\widehat{\boldsymbol{\beta}}_m^{pls} = \operatorname{argmin}_{\boldsymbol{\beta}} \left\{(\boldsymbol{y}-\widehat{\boldsymbol{y}})'(\boldsymbol{y}-\widehat{\boldsymbol{y}})\right\}, \ \widehat{\boldsymbol{y}} = \boldsymbol{X}\boldsymbol{\beta}, \ \boldsymbol{\beta} \in \mathcal{K}_m(\mathsf{b},\ \mathsf{A})$$

for

$$\mathcal{K}_m(\mathsf{b},\ \mathsf{A}) = \ \text{span}(\mathsf{b},\ \mathsf{A}^1\,\mathsf{b}, \ldots, \ \mathsf{A}^{m-1}\,\mathsf{b}).$$

PLS regression

Solve the normal equations :

$$\frac{1}{n}\,\mathsf{A}\,\boldsymbol{\beta} = \frac{1}{n}\,\mathsf{b}, \ \text{for} \ \ \mathsf{A} = \boldsymbol{X}'\boldsymbol{X}, \ \mathsf{b} = \boldsymbol{X}'\boldsymbol{y}$$

The PLS regression coefficient $\widehat{\boldsymbol{\beta}}^{pls}_m$ is a Krylov solution :

$$\widehat{\boldsymbol{\beta}}^{pls}_m = \operatorname{argmin}_{\boldsymbol{\beta}} \left\{ (\boldsymbol{y} - \widehat{\boldsymbol{y}})'(\boldsymbol{y} - \widehat{\boldsymbol{y}}) \right\}, \ \widehat{\boldsymbol{y}} = \boldsymbol{X}\boldsymbol{\beta}, \ \boldsymbol{\beta} \in \mathcal{K}_m(\mathsf{b}, \mathsf{A})$$

for

$$\mathcal{K}_m(\mathsf{b}, \mathsf{A}) = \text{span}(\mathsf{b}, \mathsf{A}^1\,\mathsf{b}, \ldots, \mathsf{A}^{m-1}\,\mathsf{b}).$$

truncate $\widehat{\boldsymbol{\beta}}^{ls}$ on the first m conjugate gradient directions

PMI RESEARCH & DEVELOPMENT

PLS regression

Solve the normal equations :

$$\frac{1}{n}\,\mathsf{A}\,\boldsymbol{\beta} = \frac{1}{n}\,\mathsf{b}, \quad \text{for} \quad \mathsf{A} = \boldsymbol{X}'\boldsymbol{X}, \; \mathsf{b} = \boldsymbol{X}'\boldsymbol{y}$$

The PLS regression coefficient $\widehat{\boldsymbol{\beta}}_m^{pls}$ is a Krylov solution :

$$\widehat{\boldsymbol{\beta}}_m^{pls} = \operatorname{argmin}_{\boldsymbol{\beta}} \left\{ (\boldsymbol{y} - \widehat{\boldsymbol{y}})'(\boldsymbol{y} - \widehat{\boldsymbol{y}}) \right\}, \; \widehat{\boldsymbol{y}} = \boldsymbol{X}\boldsymbol{\beta}, \; \boldsymbol{\beta} \in \mathcal{K}_m(\mathsf{b},\, \mathsf{A})$$

for

$$\mathcal{K}_m(\mathsf{b},\, \mathsf{A}) = \text{span}(\mathsf{b},\, \mathsf{A}^1\,\mathsf{b}, \ldots, \mathsf{A}^{m-1}\,\mathsf{b}).$$

truncate $\widehat{\boldsymbol{\beta}}^{ls}$ on the first m conjugate gradient directions

efficient dimension reduction & excellent prediction performance

PLS regression

Solve the normal equations :

$$\frac{1}{n}\,\mathsf{A}\,\boldsymbol{\beta} = \frac{1}{n}\,\mathsf{b},\ \ \text{for}\ \ \mathsf{A} = \boldsymbol{X}'\boldsymbol{X},\ \ \mathsf{b} = \boldsymbol{X}'\boldsymbol{y}$$

The PLS regression coefficient $\widehat{\boldsymbol{\beta}}_m^{pls}$ is a Krylov solution :

$$\widehat{\boldsymbol{\beta}}_m^{pls} = \operatorname{argmin}_{\boldsymbol{\beta}} \left\{ (\boldsymbol{y} - \widehat{\boldsymbol{y}})'(\boldsymbol{y} - \widehat{\boldsymbol{y}}) \right\},\ \widehat{\boldsymbol{y}} = \boldsymbol{X}\boldsymbol{\beta},\ \boldsymbol{\beta} \in \mathcal{K}_m(\,\mathsf{b},\ \mathsf{A})$$

for

$$\mathcal{K}_m(\,\mathsf{b},\ \mathsf{A}) = \ \mathsf{span}(\,\mathsf{b},\ \mathsf{A}^1\,\mathsf{b}, \ldots, \mathsf{A}^{m-1}\,\mathsf{b}).$$

truncate $\widehat{\boldsymbol{\beta}}^{ls}$ on the first m conjugate gradient directions

efficient dimension reduction & excellent prediction performance

PLS solution not easy to interpret, nonlinear function of response

Wavelets and DWT

orthonormal basis functions that allow to locally decompose a function f

$$f(x) = \sum_{r,k \in Z} d_{r,k} \, \psi_{r,k}(x),$$

$\psi_{r,k}$: the mother wavelet,
$d_{r,k}$: the wavelet coefficients,
r, k : integers that control *translations* and *dilations*

Wavelets and DWT

orthonormal basis functions that allow to locally decompose a function f

$$f(x) = \sum_{r,k \in Z} d_{r,k} \, \psi_{r,k}(x),$$

$\psi_{r,k}$: the mother wavelet,
$d_{r,k}$: the wavelet coefficients,
r, k : integers that control *translations* and *dilations*

Discrete Wavelet Transform (DWT):
orthogonal matrix $\mathcal{W}'\mathcal{W} = \mathcal{W}\mathcal{W}' = I$
extremely fast to compute (pyramid algorithm)

Spectral regions relevant for prediction

out-of-scope : denoise and reconstruct spectra

our goal : flag the spectral regions that are relevant for prediction

Spectral regions relevant for prediction

out-of-scope : denoise and reconstruct spectra

our goal : flag the spectral regions that are relevant for prediction

rationale :

rescale the PLS regression coefficient vector

rescaling takes place in the wavelet domain. It takes into account:

1. local features of the spectra captured in the wavelet coefficients
2. information on the response inherent to PLS regression

select a few non zero wavelet coefficients $d_{r,k}$ based on their relevance for prediction

DW preconditioning Krylov subspaces

Use the discrete wavelet matrix $\mathcal{W}$ to precondition the normal equations:

$$\frac{1}{n}\mathcal{W}\ \mathsf{A}\,\boldsymbol{\beta} = \frac{1}{n}\mathcal{W}\ \mathsf{b}, \tag{1}$$

solve on the transformed coordinates :

$$\frac{1}{n}\mathcal{W}\ \mathsf{A}\,\mathcal{W}'\,\widetilde{\boldsymbol{\beta}} = \frac{1}{n}\mathcal{W}\ \mathsf{b},\ \ \boldsymbol{\beta}\ \in\ \mathcal{K}_m(\widetilde{\mathsf{b}},\widetilde{\mathsf{A}})\,,\ \ \widetilde{\mathsf{A}} = \mathcal{W}\ \mathsf{A}\,\mathcal{W}',\ \widetilde{\mathsf{b}} = \mathcal{W}\ \mathsf{b}$$

recover the original solution in original coordinates by applying the inverse wavelet transform, that is :

$$\boldsymbol{\beta} = \mathcal{W}'\,\widetilde{\boldsymbol{\beta}}.$$

DW preconditioning Krylov subspaces

Use the discrete wavelet matrix $\mathcal{W}$ to precondition the normal equations:

$$\frac{1}{n}\mathcal{W}\ \mathsf{A}\,\boldsymbol{\beta} = \frac{1}{n}\mathcal{W}\ \mathsf{b}, \tag{2}$$

solve on the transformed coordinates :

$$\frac{1}{n}\mathcal{W}\ \mathsf{A}\,\mathcal{W}'\,\widetilde{\boldsymbol{\beta}} = \frac{1}{n}\mathcal{W}\ \mathsf{b},\ \ \boldsymbol{\beta}\ \in\ \mathcal{K}_m(\widetilde{\mathsf{b}},\ \widetilde{\mathsf{A}})\ ,\ \ \widetilde{\mathsf{A}} = \mathcal{W}\ \mathsf{A}\,\mathcal{W}',\ \ \widetilde{\mathsf{b}} = \mathcal{W}\ \mathsf{b}$$

recover the original solution in original coordinates by applying the inverse wavelet transform, that is :

$$\boldsymbol{\beta} = \mathcal{W}'\,\widetilde{\boldsymbol{\beta}}.$$

it is often the case in biochemical applications that interpretation in transformed coordinates is more interesting than in the original coordinates

DW preconditioning Krylov subspaces

precondition Krylov using $\mathcal{W}$ to work on the wavelet domain

run PLS on the wavelet domain (Trygg and Wold (1998))

rescale the PLS solution (Kondylis and Whittaker (2007))

1. **Initialize** $(s = 0)$ with a PLS to define importance factors $\boldsymbol{\mu}_m^0 = \boldsymbol{\mu}_m^{pls}$, as:

$$\mu_j^s = \lambda \sqrt{\frac{(\widehat{\widetilde{\beta}}_{m,j}^{\,s})^2}{\sum_j (\widehat{\widetilde{\beta}}_{m,j}^{\,s})^2}} \tag{3}$$

2. define **relevant subset** A^s from $\boldsymbol{\mu}_m^{s-1}$ using a multiple testing procedure
3. **Stop** if this subset has not changed. **Output:** a set of coefficients

$$\{\widehat{\widetilde{\beta}}_{m,j}^{\,s*};\ j \in A^{\,s*}\} \cup \{\widehat{\widetilde{\beta}}_{m,j'}^{\,s*};\ j' \in B^{\,s*}\}.$$

recover the Krylov solution in the original coordinates system

PMI RESEARCH & DEVELOPMENT

Illustration : cookies data

well known data set in statistical literature

- introduced : B.G. Osborne, T. Fearn, A.R. Miller, and S. Douglas (1984)
- PLS regression on smooth factors (K. Goutis and T. Fearn (1996))
- robust PLS methods (M. Hubert, P.J. Rousseeuw, S. Van Aelst (2008))
- bayesian variable selection (P.J. Brown, T. Fearn, M. Vannucci (2001))

Illustration : cookies data

well known data set in statistical literature

- introduced : B.G. Osborne, T. Fearn, A.R. Miller, and S. Douglas (1984)
- PLS regression on smooth factors (K. Goutis and T. Fearn (1996))
- robust PLS methods (M. Hubert, P.J. Rousseeuw, S. Van Aelst (2008))
- bayesian variable selection (P.J. Brown, T. Fearn, M. Vannucci (2001))

responses : fat, sucrose, dry flour, and water

predictors : 700 points measuring NIR reflectance from 1100 to 2498 nm in steps of 2

we study fat concentration

we keep reflectance for wavelengths ranging from 1380 to 2400 nm

Training set : 1 to 40 - Test set : 41 to 72

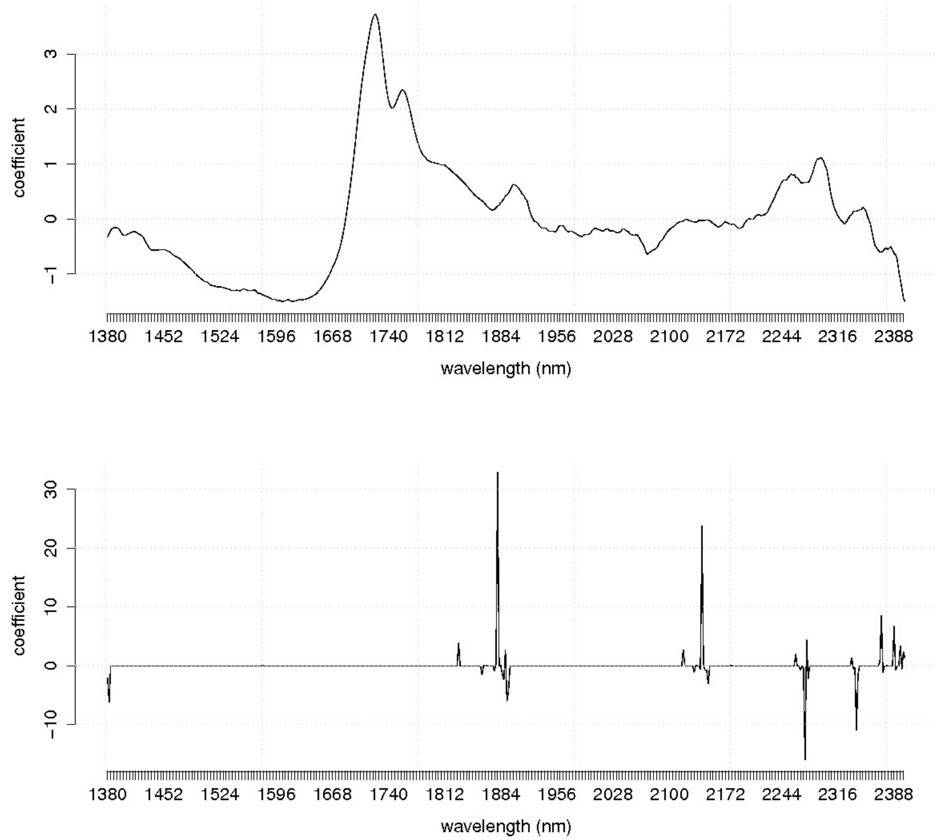


Figure 1: Cookies data: regression coefficients for PLS (upper panel), and DW-PLS (lower panel). The response variable is fat. The number of components has been settled to 5 according to literature knowledge. The Haar wavelet has been used for DW-PLS.