Sparse models project

Louca Malerba

January 2026

1 Prise de note sur le papier

1.1 Abstract

Lier un ensemble de variables X à une réponse y est crutial en chimie et en data science plus généralement. Une prédiction peut-être permise par de l'interprétation des données, apr exemple en localisant les caractéristiques les plus importantes. Pour cela, on utilise de la réduction de dimensionnalité soit par de la projection (PLS, PSA, ...) soit de la sélection de variable (lasso, ...). sPLS (sparce Partial Least Squares) combine les deux stratégies, en intégrant le processus de selection de variable à PLS.

Le papier sur lequel on travaille, Dual-sPLS généraliser l'algorithme classique PLS, fournissant un méthode équilibrant précision de prédiction et interpretation efficace.

Cette méthode est basée sur la pénalisation inspirée par les méthodes de régression classiques (lasso, group lasso, least squares, ridge) et utilise la notion de norme dual.

The resulting sparsity is enforced by an intuitive shrinking ratio parameter. Dual-sPLS favorably compares to similar regression methods, on simulated and real chemical data.

1.2 Introduction

Le papier commence par expliquer le problème réel : en chimie, on a des spectres (matrice X) et on veut prédire une propriété, comme la densité (vecteur y).

Problème: surraprentissage possible car on a P longueurs d'ondes $>>$ N échantillons

Le papier explique ensuite que les méthodes classiques présentent des limites:

La PLS classique (à définir*) réduit la dimension mais utilise toutes les variables, ce qui est ininterprétable pour un humain.

Lasso sélectionne réellement les variables mais ignore la structure de corrélation entre-elles

Le Dual-sPLS est une famille de méthodes dont l'avantage est d'être flexible. Il permet de faire de la parcimonie: on force le modèle à ne choisir que quelques

variables clés, ce qui rend le résultat interprétable (on sait quelles bandes du spectre comptent).

En clair: "Le problème traité est celui de la régression en haute dimension ($P >> N$). L'objectif du Dual-sPLS est de combiner la réduction de dimension de la PLS avec la sélection de variables du Lasso, tout en offrant une formulation 'duale' qui permet de mieux grouper les variables spectrales importantes."

1.3 Background

Ici, on va regarder en détail les bases mathématiques sur lesquelles se base l'algorithme :

1.3.1 PLS classique

L'idée est de construire des "composantes latentes" **t** qui sont des combinaisons linéaires des variables originales **X**.

L'objectif:

$max_w(y^T Xw)$ s.t. $||w||_2 = 1$

$\Rightarrow$ On cherche un vecteur de poids w qui maximise la covariance entre la projection Xw et la cible y , sous contrainte que w soit unitaire ($||w||_2 = 1$).

Or: Par les multiplicateurs de Lagrange, on montre que le poids optimal est simplement $w \propto X^T y$ (la covariance brute). (équation 9)

Le mécanisme itératif (Algorithme 1) :

- On calcule le poids w_m.
- On calcule la composante $t_m = X_m w_m$.
- La Déflation (Éq. 10) : C'est l'étape la plus mathématique. On retire de X l'information déjà captée par t_m en projetant X sur l'orthogonal de t_m. On recommence avec ce "résidu" pour trouver la composante suivante.

1.3.2 Least absolute shrinkage and selection operator (LASSO)

Le papier rappelle pourquoi on ne fait pas juste une régression classique (Moindres Carrés):

$argmin_\beta(||y - X\beta||_2^2)$ s.t. $||\beta|| \leq \lambda$

Le paramètre threshold λ contrôle la taille des coefficients du modèle: C'est l'opérateur qui "tue" les petits coefficients. Si la valeur est en dessous d'un seuil, elle devient 0. Sinon, on la réduit vers zéro. C'est ce qui crée la parcimonie.

1.3.3 sPLS: le mariage des deux

Le Sparse PLS est une tentative de mélanger les deux mondes : garder la structure itérative de la PLS, mais forcer les poids w à être parcimonieux (comme dans le Lasso).

On peut reprendre l'équation:

$max_w(y^T Xw)$ s.t. $||w||_2 = 1$

et en remarquant que:

$\hat{Cov}(Xw, y) = \frac{1}{N} W^T z$ où $z = N\hat{Cov}(X, y)$

et en ajoutant le paramètre de couplage $\lambda_s > 0$ on obtient:

$\hat{\mathbf{w}} = \arg\min\{-\widehat{\text{Cov}}(\mathbf{Xw}, \mathbf{y}) + \lambda_s|\mathbf{w}|_1\}$, avec $w^T w = 1$

$\Rightarrow$ On ne veut plus seulement maximiser la covariance (le terme Cov), on veut aussi que la norme L1 de w soit petite.

1.4 Compréhension haut niveau de Dual-sPLS

Plutôt que de s'embêter avec des contraintes compliquées, les auteurs disent : maximiser la covariance, c'est comme minimiser une "fonction de coût" basée sur une norme $\Omega(w)$.

En clair : On définit une "forme" (la norme Ω) et on cherche le vecteur w qui rentre le mieux dedans tout en pointant vers les données.

Le papier propose plusieurs variantes, mais la plus importante pour toi est la Dual-sPLS$_L$(Pseudo-Lasso).

Sa formule (Éq. 23) est :

$\Omega(w) = \lambda||w||_1 + ||w||_2$

C'est le mélange parfait : le $||w||_2$ garde la structure de la PLS classique, et le $\lambda||w]]_1$ force les petits coefficients à devenir zéro.

Seuillage doux: C'est le passage le plus important du papier (Éq. 30). Pour trouver w, on ne fait pas d'optimisation complexe, on utilise une fonction de seuillage δ_ν . La logique est simple :

Tu calcules la corrélation brute entre tes variables et ta cible : $z = XTy$. Tu appliques un **seuil ν ** :

Si la corrélation d'une variable est plus petite que $\nu \rightarrow$ Hop, elle devient 0.

Si elle est plus grande $\rightarrow$ On la garde, mais on la réduit un peu.

Comment choisir le seuil ν ?

C'est là que le papier est très fort. Au lieu de te demander de deviner une valeur mathématique abstraite pour λ , il propose un Shrinking Ratio (σ).

Exemple : Tu décides que tu veux garder seulement 20% des variables (donc 80% de zéros). L'algorithme regarde tes données, calcule le seuil ν automatiquement pour que 80% des variables passent à la trappe.

C'est l'objet de la Figure 1 : on trie les variables et on coupe là où on a atteint le pourcentage de zéros voulu.

1.5 Résultats données simulées

Le papier utilise des "mélanges de Gaussiennes" (Figure 2).

L'intérêt : Dans la vraie vie, on ne sait jamais avec certitude quelles variables sont vraiment importantes. Ici, comme on crée les données nous-mêmes, on connaît la "Ground Truth" (la vérité terrain). À dire à l'oral : "On a testé le modèle sur des données simulées pour vérifier s'il était capable de retrouver exactement les variables que nous avions nous-mêmes cachées dans le bruit."

1.6 Résultats données réelles

C'est ton cas d'usage. Le papier mentionne le traitement Savitzky-Golay.

Ce que c'est : C'est un filtre de lissage + dérivée. En spectroscopie, les spectres se ressemblent tous (de grandes bosses). Dériver permet d'accentuer les pics et de supprimer les dérives de la ligne de base.

Haut niveau : C'est l'étape de Feature Engineering indispensable avant d'envoyer les données dans la sPLS.