

EXTENSION DE LA NMF SUPERVISÉE POUR L'INTÉGRATION DE DONNÉES OMIQUES

Aurélie Mercadié^{1,2}, Éléonore Gravier², Gwendal Josse², Nathalie Vialaneix¹ & Céline Brouard¹

¹ *Université de Toulouse, INRAE, UR MIAT, F-31320, Castanet-Tolosan, France, {aurelie.mercadie, nathalie.vialaneix, celine.brouard}@inrae.fr*

² *Pierre Fabre Dermo-Cosmétique, F-31300, Toulouse, France, {eleonore.gravier, gwendal.josse}@pierre-fabre.com*

Résumé. Le but de cette communication est de présenter une nouvelle méthode d'intégration découvrant des relations entre données « omiques » qui caractérisent des profils typiques de groupes distincts d'individus. Ici, la méthode de Factorisation Matricielle Nonnegative (NMF) est étendue à la problématique d'intégration de données dans un cadre supervisé. Pour cela, nous nous basons sur une variante du problème d'optimisation FR-lda, présenté par [Leuschner et al., 2019], les liens entre omiques étant assurés par l'appariement des individus dans la décomposition. En outre, [Leuschner et al., 2019] utilisent l'algorithme Majoration-Minimisation (MM) menant à des termes de mise à jour multiplicatifs pour résoudre leur problème. Afin d'assurer une parcimonie exacte dans la décomposition, nous proposons de combiner cette approche avec une mise à jour basée sur une approche proximale.

Mots-clés. intégration multi-omiques, apprentissage supervisé, NMF, proximal optimization.

Abstract. The aim of this proposal is to develop an integrative method to find relationships between “omics” characterizing typical profiles that correspond to distinct groups of individuals. Here, the Nonnegative Matrix Factorization (NMF) is extended to this integration problem in a supervised framework. We start from a variant of the FR-lda optimization problem proposed by [Leuschner et al., 2019], using individuals to link omics. In addition, in [Leuschner et al., 2019], a surrogate approach, the Majorize-Minimization (MM) algorithm is used to solve the optimization problem and it leads to multiplicative updates that ensure positivity of the solution. To ensure exact sparsity in the decomposition, we propose to combine this method with another update based on a proximal approach.

Keywords. multi-omics integration, supervised learning, NMF, proximal optimization.

1 Contexte

Le développement des approches haut débit en biologie permet dorénavant la production massive de données dites « omiques », et ce pour des contextes applicatifs variés.

Généralement acquis sur les mêmes échantillons, chacun de ces tableaux de données omiques illustre une partie seulement d'un système biologique complexe. L'intégration de ces données permet donc d'étudier ce système en globalité, et de mettre en lumière les relations existantes entre les divers acteurs moléculaires. Globalement, les méthodes statistiques d'intégration peuvent être regroupées en deux grandes familles : celles explorant les interactions éventuelles entre omiques, appelées *méthodes non supervisées* [Meng et al., 2016, Eicher et al., 2020] ; et celles qui visent à obtenir des interactions prédisant ou expliquant une donnée extérieure, tel un phénotype d'intérêt chez les échantillons étudiés, appelées *méthodes supervisées* [Ritchie et al., 2015, Eicher et al., 2020].

Ici, nous abordons la question sous un angle mixte, celui de l'analyse exploratoire d'omiques multiples dans laquelle une information complémentaire caractérisant ces individus (attribut clinique ou expérimental par exemple) est d'intérêt pour la compréhension du phénomène biologique. Il s'agit alors de mettre en lumière comment ces divers omiques interagissent *selon la typologie clinique des individus*. Dans notre cas, nous nous restreindrons au cas où l'attribut clinique est catégoriel, c'est-à-dire, au cas où les individus sont partitionnés en groupes distincts correspondant aux niveaux de cet attribut catégoriel. Ce problème est déjà abordé par la méthode DIABLO [Singh et al., 2019], dérivée de l'analyse généralisée des corrélations canoniques [Tenenhaus et al., 2014] qui recherche des projections maximisant un critère de covariance entre paires d'omiques et avec l'attribut catégoriel.

Dans cette communication, nous proposons une extension de la Factorisation Matricielle Non-négative (NMF) [Lee and Seung, 2001] et plus particulièrement de sa version supervisée [Lee and Choi, 2009]. En effet, cette méthode de réduction de dimension offre un cadre bien adapté au problème d'intégration de données omiques pour des individus structurés en groupe. D'une part, elle a été mise en œuvre pour l'analyse de données positives, ce qui est le cadre naturel de nombreuses données omiques (données de comptages comme le RNA-seq ou les données métagénomiques, données compositionnelles comme en métabolomique ou protéomique, ...). D'autre part, relativement aux approches factorielles ou PLS, son interprétation est elle-même facilitée par la contrainte de positivité de la solution, la décomposition retenue s'expliquant aisément en termes de profils types et d'appartenance à ces profils de chacun des individus. De plus, la version supervisée intègre un critère imposant spécifiquement à la décomposition d'être prédictive de la variable catégorielle expliquant les groupes, ce qui rend son interprétation plus simple que des approches maximisant un critère global de covariance entre paires d'omiques. Enfin, son cadre d'optimisation flexible rend possible l'intégration de contraintes variées, permettant une bonne adaptation au problème pratique posé et utilisable même dans un cadre de grande dimension où les tailles d'échantillon sont souvent très petites devant le nombre de variables décrivant chaque type d'omiques.

Diverses variantes de la NMF avaient déjà été proposées pour l'analyse et l'intégration de données omiques : dans un cadre complètement non supervisé, [Zhang et al., 2011, Yang and Michailidis, 2016] l'ont utilisée pour l'intégration d'omiques. Une extension supervisée de la NMF, bien adaptée au cas où les individus sont partitionnés en divers groupes, a, par ailleurs, été proposée par [?, Leuschner et al., 2019, Li et al., 2022] qui l'ont principalement utilisée pour des problèmes de classification de textes et d'images mais cette version a aussi trouvé des applications en microbiologie cellulaire [Shreeves et al., 2021].

Nous proposons ici une variante intégrative à la méthode FR-lda de [Leuschner et al., 2019]. Cette variante étend la méthode initiale selon deux angles : d’une part, en permettant l’analyse simultanée (l’intégration) de plusieurs omiques et, d’autre part, en améliorant l’approche de résolution du problème d’optimisation pour qu’elle conduise à une parcimonie exacte de la solution. Dans la suite, on notera $\mathbf{X}^{(j)} \in \mathbb{R}_+^{n \times p_j}$ ($j \in \{1, 2\}$), deux matrices à entrées positives contenant les mesures de deux types d’omiques sur les mêmes n individus (p_j étant le nombre de variables mesurées dans chaque omique). On notera également $\mathbf{y} \in \{0, 1\}^n$ le vecteur d’appartenance des individus aux groupes. Dans le cadre de cette communication, pour des questions de simplicité, la présentation est restreinte à deux groupes mais la méthode s’étend sans difficulté à un nombre plus important de groupes.

Dans la suite, nous introduisons le cadre général de la NMF ainsi que l’approche proposée par [Leuschner et al., 2019] dans la section 2. Puis, dans la section 3, nous présenterons une nouvelle version intégrative et supervisée de la NMF et nous détaillerons la résolution du problème d’optimisation associé à cette dernière. Enfin, la conclusion aborde les questions ouvertes concernant la méthode proposée et les développements à venir.

2 NMF et NMF supervisée

2.1 NMF

Le but de la NMF est d’approximer une matrice $\mathbf{X} \in \mathbb{R}_+^{n \times p}$ tout en réduisant sa dimension, initialement très grande ($n \ll p$). Pour se faire, \mathbf{X} est décomposée en deux matrices non-négatives tel que $\mathbf{X} \simeq \mathbf{WH}$ avec $\mathbf{W} \in \mathbb{R}_+^{n \times K}$ et $\mathbf{H} \in \mathbb{R}_+^{K \times p}$, où K est le nombre de composantes latentes construites par la méthode, choisi par l’utilisateur. [Lee and Seung, 2001] décrivent deux variantes de la NMF qui se fondent sur deux fonctions de coût distinctes : la divergence de Kullback-Leibler et la norme de Fröbenius. Le choix de la fonction de coût dépend de la distribution statistique que l’on attribue au terme d’erreur. En supposant que ce dernier suit une loi normale, on préfère utiliser la norme de Fröbenius (problème des moindres carrés) et la NMF conduit alors à résoudre le problème suivant :

$$\min_{\mathbf{W}, \mathbf{H} \geq 0} \frac{1}{2} \|\mathbf{X} - \mathbf{WH}\|_F^2.$$

La NMF est initialement une méthode non supervisée, conçue pour les analyses exploratoires.

2.2 NMF supervisée

[Leuschner et al., 2019] ont développé plusieurs variantes de la NMF adaptées à la classification d’images MALDI¹, dont la NMF « FR-lda ». Les images correspondent ici à des

1. Matrix Assisted Laser Desorption/Ionization

tissus de poumon atteints ou non d'un cancer. Cette méthode cherche donc à extraire des composantes latentes parcimonieuses qui caractérisent correctement les deux types de tissus. Soit $\mathbf{X} \in \mathbb{R}_+^{n \times K}$ la matrice décrivant les images et $\mathbf{y} \in \{0, 1\}^n$ le vecteur encodant les 2 groupes, on cherche à résoudre le problème suivant :

$$\min_{\mathbf{W}, \mathbf{H}, \boldsymbol{\beta} \geq 0} \mathcal{F}_0(\mathbf{W}, \mathbf{H}, \boldsymbol{\beta})$$

où

$$\mathcal{F}_0(\mathbf{W}, \mathbf{H}, \boldsymbol{\beta}) = \frac{1}{2} \|\mathbf{X} - \mathbf{W}\mathbf{H}\|_F^2 + \frac{\mu}{2} \|\mathbf{W}\|_F^2 + \lambda \|\mathbf{H}\|_1 + \frac{\nu}{2} \|\mathbf{H}\|_F^2 + \frac{\gamma}{2} \|\mathbf{y} - \mathbf{X}\mathbf{H}^\top \boldsymbol{\beta}\|_2^2 \quad (1)$$

avec :

- $\mathbf{W} \in \mathbb{R}_+^{n \times K}$, les contributions des individus aux composantes latentes ;
- $\mathbf{H} \in \mathbb{R}_+^{K \times p}$, les composantes latentes ;
- $\boldsymbol{\beta} \in \mathbb{R}_+^K$, les coefficients de régression ;
- $\lambda, \gamma, \nu, \mu > 0$, les paramètres de régularisation, fixés.

La fonction objective de l'équation 1 est décomposée en plusieurs termes : le premier terme est un terme d'adéquation de la reconstruction aux données initiales, classique en NMF. Le terme faisant intervenir la norme ℓ_1 est utilisé pour obtenir des composantes latentes parcimonieuses pour une meilleure interprétabilité. Les termes faisant intervenir la norme ℓ_2 sont présents pour régulariser la solution mais également, dans le cas de \mathbf{W} , pour éviter un problème de non identifiabilité trivial posé par l'équivalence $\mathbf{W}\mathbf{H} = c\mathbf{W} \times \frac{1}{c}\mathbf{H}$ pour tout c . Enfin, le dernier terme permet d'obtenir des composantes latentes qui discriminent bien les groupes représentés par \mathbf{y} .

2.3 Résolution

Les problèmes d'optimisation qui apparaissent dans la NMF sont des problèmes non-convexes et non-linéaires [Fernsel and Maass, 2018]. En effet, \mathcal{F}_0 n'est pas convexe en \mathbf{W} , \mathbf{H} , et $\boldsymbol{\beta}$ simultanément, mais des problèmes d'optimisation convexes distincts peuvent être écrits pour chacune des variables, l'un de ces problèmes incluant une contrainte non lisse (*non-smooth*). Une approche pour résoudre ce type de problème est d'utiliser les algorithmes itératifs où chacune des variables d'intérêt est mise à jour successivement à chaque itération. Nous définissons une marginale issue de \mathcal{F}_0 pour chacune de ces variables.

Afin d'obtenir les différentes règles de mise à jour, les auteurs utilisent une approche Majoration-Minimisation (MM). Cette dernière consiste à minimiser la fonction objective \mathcal{F} en minimisant une seconde fonction, appelée surrogée et notée $\mathcal{Q}_{\mathcal{F}}$. $\mathcal{Q}_{\mathcal{F}}(\cdot, \cdot)$ est une fonction $\mathbb{R}^D \times \mathbb{R}^D \rightarrow \mathbb{R}$ facile à minimiser en sa première variable pour toutes les valeurs de la seconde variable et qui satisfait $\mathcal{Q}_{\mathcal{F}}(\mathbf{a}, \mathbf{a}) = \mathcal{F}(\mathbf{a})$, $\forall \mathbf{a} \in \mathbb{R}^D$ et $\mathcal{F}(\mathbf{a}) \leq \mathcal{Q}_{\mathcal{F}}(\mathbf{a}, \mathbf{b})$, $\forall \mathbf{a}, \mathbf{b} \in \mathbb{R}^D$. Comme, $\mathcal{Q}_{\mathcal{F}}(\mathbf{a}, \mathbf{a}^{(t)})$ est une borne supérieure de $\mathcal{F}(\mathbf{a})$ pour tout $\mathbf{a} \in \mathbb{R}^D$, en définissant une mise à jour pour \mathbf{a} de la forme

$$\mathbf{a}^{(t+1)} = \operatorname{argmin}_{\mathbf{a}} \mathcal{Q}_{\mathcal{F}}(\mathbf{a}, \mathbf{a}^{(t)}), \quad (2)$$

on s'assure qu'à chaque itération de l'algorithme, on obtient une diminution monotone de la valeur de \mathcal{F} . Dans le cas particulier de la NMF, ce principe mène à des mises à jour

multiplicatives (*c.à.d* exprimées uniquement via des additions, divisions et multiplications), ce qui assure la positivité des quantités optimisées. Deux méthodes sont utilisées pour définir la surrogée : le principe de contrainte supérieure quadratique (ou UQBP, pour \mathbf{W} et \mathbf{H}) et l'inégalité de Jensen (pour β) [Fernsel and Maass, 2018].

2.3.1 Principe UQBP

Pour la mise à jour de \mathbf{W} et \mathbf{H} , [Fernsel and Maass, 2018] utilisent le développement de Taylor à l'ordre 2 de la marginale de \mathcal{F} en la variable d'intérêt et obtiennent la proposition suivante :

Proposition 1 ([Fernsel and Maass, 2018]) *Soit $\mathcal{F} : \mathbb{R}^D \rightarrow \mathbb{R}$ une fonction deux fois continûment dérivable et $\mathbf{M}(\mathbf{b}) \in \mathbb{R}^{D \times D}$ une matrice telle que, $\forall \mathbf{b} \in \mathbb{R}^D$, $\mathbf{M}(\mathbf{b}) - \Delta^2 \mathcal{F}(\mathbf{b})$ est semi-définie positive (avec $\Delta^2 \mathcal{F}(\mathbf{b})$ la Hessienne de \mathcal{F} en \mathbf{b}). Alors, quand \mathcal{F} est un polynôme du second degré,*

$$\mathcal{Q}_{\mathcal{F}}(\mathbf{a}, \mathbf{b}) = \mathcal{F}(\mathbf{b}) + (\mathbf{a} - \mathbf{b})^\top \nabla \mathcal{F}(\mathbf{b}) + \frac{1}{2}(\mathbf{a} - \mathbf{b})^\top \mathbf{M}(\mathbf{b})(\mathbf{a} - \mathbf{b})$$

est une surrogée de \mathcal{F} . De plus, quand $\mathbf{M}(\mathbf{b})$ est symétrique et définie positive, la surrogée correspondante est strictement convexe et sa minimisation mène à la règle de mise à jour suivante :

$$\mathbf{a}^{(t+1)} = \mathbf{a}^{(t)} - \mathbf{M}^{-1}(\mathbf{a}^{(t)}) \nabla \mathcal{F}(\mathbf{a}^{(t)}).$$

Pour la mise à jour de \mathbf{W} , [Fernsel and Maass, 2018] proposent de définir $\mathbf{M}(\mathbf{b})$ comme

$$\mathbf{M}(\mathbf{b}) = \left[\delta_{i,j} \frac{(\Delta^2 \mathcal{F}(\mathbf{b}) \mathbf{b})_i}{\mathbf{b}_i} \right]_{i,j=1,\dots,D}.$$

Dans le cadre de la version de la NMF de [Leuschner et al., 2019, Fernsel and Maass, 2018], ce choix de \mathbf{M} conduit à une mise à jour multiplicative de \mathbf{M} . Une approche similaire est mise en œuvre pour la mise à jour de \mathbf{H} , intégrant le terme additionnel $\lambda \|\mathbf{H}\|_1$ dans la méthode de résolution.

2.3.2 MM basée sur l'inégalité de Jensen

Dans le second cas, la marginale de \mathcal{F} en β est majorée par l'inégalité de Jensen pour obtenir la proposition suivante :

Proposition 2 ([Fernsel and Maass, 2018]) *Soit \mathcal{F} une fonction objective dérivée d'une fonction convexe, dérivable en continu $f : \mathbb{R}^D \rightarrow \mathbb{R}$ où $\forall \mathbf{a} \in \mathbb{R}^D$, $\mathcal{F}(\mathbf{a}) = f(\mathbf{c}^\top \mathbf{a})$ avec $\mathbf{c} \in \mathbb{R}^D$ une variable auxiliaire. Alors,*

$$\mathcal{Q}_{\mathcal{F}}(\mathbf{a}, \mathbf{b}) = \sum_{k=1}^D \frac{\mathbf{c}_k \mathbf{b}_k}{\mathbf{c}^\top \mathbf{b}} f \left(\frac{\mathbf{c}^\top \mathbf{b}}{\mathbf{b}_k} \mathbf{a}_k \right)$$

est une surrogée de \mathcal{F} .

2.3.3 Résolution globale du problème de NMF

Finalement, les auteurs obtiennent les règles de mise à jour suivantes :

$$\begin{aligned} & \text{— } \mathbf{W}^{(t+1)} \leftarrow \mathbf{W}^{(t)} \odot (\mathbf{X}\mathbf{H}^\top (. /) (\mathbf{W}^{(t)}\mathbf{H}\mathbf{H}^\top + \mu\mathbf{W}^{(t)})) \\ & \text{— } \mathbf{H}^{(t+1)} \leftarrow \mathbf{H}^{(t)} \odot ((\mathbf{W}^\top\mathbf{X} + \gamma\boldsymbol{\beta}\mathbf{y}^\top\mathbf{X}) (. /) (\mathbf{W}^\top\mathbf{W}\mathbf{H}^{(t)} + \nu\mathbf{H}^{(t)} + \lambda\boldsymbol{\beta}\boldsymbol{\beta}^\top\mathbf{H}^{(t)}\mathbf{X}^\top\mathbf{X} + \lambda)) \\ & \text{— } \boldsymbol{\beta}^{(t+1)} \leftarrow \boldsymbol{\beta}^{(t)} \odot ((\mathbf{H}\mathbf{X}^\top\mathbf{y}) (. /) (\mathbf{H}\mathbf{X}^\top\mathbf{X}\mathbf{H}^\top\boldsymbol{\beta}^{(t)})) \end{aligned}$$

avec \odot le produit termes à termes et $(. /)$ la division termes à termes.

Pour la mise à jour de \mathbf{H} , la méthode requiert la stricte positivité des entrées de la matrice à toutes les étapes de l'algorithme (pour assurer la définition correcte de la matrice auxiliaire \mathbf{M} définie dans la section 2.3.1). Ceci est en contradiction avec la contrainte de parcimonie imposée sur \mathbf{H} par l'introduction de la norme ℓ_1 dans la fonction objective. Pour résoudre cette difficulté, [Fernsel and Maass, 2018] ajoutent une petite valeur positive à toutes les entrées de \mathbf{H} et ne récupèrent donc qu'une parcimonie approchée.

3 Description de l'approche proposée

3.1 Fonction objective

Nous proposons d'étendre l'approche de [Fernsel and Maass, 2018] à l'intégration de données sous l'angle d'une NMF supervisée correspondant au problème de minimisation suivant :

$$\min_{\mathbf{W}, \mathbf{H}^{(1)}, \mathbf{H}^{(2)}, \boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)}} \mathcal{F}(\mathbf{W}, \mathbf{H}^{(1)}, \mathbf{H}^{(2)}, \boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)}) \quad (3)$$

où

$$\begin{aligned} \mathcal{F}(\mathbf{W}, \mathbf{H}^{(1)}, \mathbf{H}^{(2)}, \boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)}) &= \frac{1}{2} \left(\sum_{j=1}^2 \|\mathbf{X}^{(j)} - \mathbf{W}\mathbf{H}^{(j)}\|_F^2 \right) + \frac{\gamma}{2} \left(\sum_{j=1}^2 \|\mathbf{y} - \mathbf{X}^{(j)}\mathbf{H}^{(j)\top}\boldsymbol{\beta}^{(j)}\|_2^2 \right) \\ &+ \sum_{j=1}^2 \lambda \|\mathbf{H}^{(j)}\|_1 + \frac{\mu}{2} \|\mathbf{W}\|_F^2 \end{aligned} \quad (4)$$

avec :

- $\mathbf{W} \in \mathbb{R}_+^{n \times K}$, les contributions communes des individus aux composantes latentes ;
- $\forall j \in \{1, 2\}$, $\mathbf{H}^{(j)} \in \mathbb{R}_+^{K \times p_j}$, les composantes latentes ;
- $\forall j \in \{1, 2\}$, $\boldsymbol{\beta}^{(j)} \in \mathbb{R}_+^K$, les coefficients de régression ;
- $\lambda, \gamma, \nu, \mu > 0$, les paramètres de régularisation, fixés.

De manière similaire à l'équation 1, la fonction objective inclut un terme d'adéquation de la décomposition aux données initiales (ici, en deux parties, chacune correspondant à un omique), un terme de régularisation pour assurer la parcimonie des composantes latentes (aussi en deux parties), un terme de régularisation ℓ_2 pour \mathbf{W} dont le rôle est similaire à celui de l'équation 1 et un terme assurant que les composantes latentes discriminent bien les groupes représentés par \mathbf{y} (aussi en deux parties).

3.2 Algorithme général de résolution

Nous proposons la résolution du problème de l'équation (3) par l'approche générale décrite dans l'algorithme 1. Dans cet algorithme, les termes de mise à jour pour \mathbf{W} , $\mathbf{H}^{(j)}$, et $\boldsymbol{\beta}^{(j)}$ (étapes 3-5 de l'algorithme) utilisent tous une approche par surrogée, telle que définie dans la section 2.3. Toutefois, contrairement à [Leuschner et al., 2019], le terme de mise à jour pour $\mathbf{H}^{(j)}$ est obtenu via une approche proximale afin d'obtenir une parcimonie exacte.

Algorithm 1 Vue d'ensemble de l'algorithme utilisé pour la résolution de l'équation (3)

- 1: Initialiser les matrices $\mathbf{W}^{(0)}$, $\mathbf{H}^{(j,0)}$ et vecteurs $\boldsymbol{\beta}^{(j,0)}$ avec des valeurs strictement positives ($\forall j \in \{1, 2\}$).
- 2: **for all** $t = 1, \dots, T$ **do**
- 3: $\mathbf{W}^{(t+1)} \leftarrow \mathbf{W}^{(t)} \odot \mathbf{A}(\mathbf{W}^{(t)})$, où \odot est la multiplication terme à terme
- 4: pour $j = 1, 2$, $\mathbf{H}^{(j,t+1)} \leftarrow \text{prox}_{g_j} \left(\tilde{\mathbf{H}}^{(j)} \right)$, avec $\tilde{\mathbf{H}}^{(j)} = \mathbf{H}^{(j,t)} - \frac{1}{\eta} \nabla f_j(\mathbf{H}^{(j,t)})$
- 5: pour $j = 1, 2$, $\boldsymbol{\beta}^{(j,t+1)} \leftarrow \boldsymbol{\beta}^{(j,t)} \odot \mathbf{u}(\boldsymbol{\beta}^{(j,t)})$
- 6: **end for**
- 7: **return** $\mathbf{W} := \mathbf{W}^{(T+1)}$, $\mathbf{H}^{(j)} := \mathbf{H}^{(j,T+1)}$ et $\boldsymbol{\beta}^{(j)} := \boldsymbol{\beta}^{(j,T+1)}$ ($j = 1, 2$)

où $\mathbf{A}(\mathbf{W}^{(t)})$ et $\mathbf{u}(\boldsymbol{\beta}^{(j,t)})$ sont, respectivement, une matrice et un vecteur avec des entrées positives, dont l'expression exacte est donnée en section 3.3, prox est l'opérateur proximal et f_j et g_j sont deux fonctions, définies également en section 3.3.

3.3 Optimisation

Afin d'alléger les notations, nous omettons le terme t pour les matrices qui ne sont pas mises à jour. Pour la même raison, étant donné que les fonctions objectives pour $\mathbf{H}^{(j)}$ et $\boldsymbol{\beta}^{(j)}$ sont séparables en j , les indices j (référant aux jeux d'omiques) sont également omis.

3.3.1 Mise à jour de \mathbf{W}

La règle de mise à jour pour la matrice des contributions des individus \mathbf{W} est obtenue grâce à l'approche MM comme expliqué en Section 2.3.

$$\mathbf{W}^{(t+1)} = \mathbf{W}^{(t)} \odot \mathbf{A}(\mathbf{W}^{(t)}) = \mathbf{W}^{(t)} \odot \mathbf{C}(\cdot/)(\mathbf{W}^{(t)}\mathbf{B}),$$

avec $\mathbf{C} = \mathbf{X}^{(1)}\mathbf{H}^{(1)\top} + \mathbf{X}^{(2)}\mathbf{H}^{(2)\top}$ et $\mathbf{B} = \mathbf{C} + \mu\mathbb{I}_K$.

3.3.2 Mise à jour de $\boldsymbol{\beta}^{(j)}$

La règle de mise à jour pour $\boldsymbol{\beta}$ est obtenue grâce à l'approche MM comme expliqué en section 2.3 :

$$\boldsymbol{\beta}^{(t+1)} = \boldsymbol{\beta}^{(t)} \odot \mathbf{u}(\boldsymbol{\beta}^{(t)}) = \boldsymbol{\beta}^{(t)} \odot (\mathbf{H}\mathbf{X}^\top \mathbf{y}) (\cdot/)(\mathbf{H}\mathbf{X}^\top \mathbf{X}\mathbf{H}^\top \boldsymbol{\beta}^{(t)}).$$

3.3.3 Mise à jour de $\mathbf{H}^{(j)}$

Dans leur approche, comme indiqué dans la section 2.3, [Leuschner et al., 2019] ne récupèrent qu'une parcimonie approchée pour \mathbf{H} . Ici, nous proposons une méthode plus directe, qui assure une parcimonie exacte pour cette matrice, au travers de l'approche proximale.

L'approche est basée sur la résolution générale de problèmes $\min_{\mathbf{a} \in \mathbb{R}_+^D} \mathcal{F}(\mathbf{a})$ dans lesquels $\mathcal{F}(\mathbf{a})$ prend la forme $f(\mathbf{a}) + \lambda g(\mathbf{a})$ avec $\lambda > 0$, f une fonction dont le gradient est Lipschitz (avec une constante de Lipschitz η) et g une fonction décrivant une contrainte non lisse. Dans ce cas, [Parikh and Boyd, 2014, Bauschke and Combettes, 2017] (avec l'algorithme « Forward Backward Splitting » (FBS)) montrent que la fonction

$$\mathcal{Q}_{\mathcal{F}}(\mathbf{a}, \mathbf{b}) = f(\mathbf{b}) + (\mathbf{a} - \mathbf{b})^\top \nabla f(\mathbf{b}) + \frac{\eta}{2} \|\mathbf{a} - \mathbf{b}\|^2 + \lambda g(\mathbf{a})$$

est une surrogée de \mathcal{F} .

En notant que $\mathcal{Q}_{\mathcal{F}}(\mathbf{a}, \mathbf{a}^{(t)})$ a le même minimum que $\frac{1}{2} \|\mathbf{a} - \tilde{\mathbf{a}}\|_2^2 + \frac{\lambda}{\eta} g(\mathbf{a})$, avec $\tilde{\mathbf{a}} = \mathbf{a}^{(t)} - \frac{1}{\eta} \nabla f(\mathbf{a}^{(t)})$, il est possible de démontrer que la règle de mise à jour basée sur cette surrogée, *c.à.d.*, la solution de l'équation (2), est obtenue par :

$$\mathbf{a}^{(t+1)} = \text{prox}_g(\tilde{\mathbf{a}}^t) \quad \text{avec} \quad \tilde{\mathbf{a}}^t = \mathbf{a}^{(t)} - \frac{1}{\eta} \nabla f(\mathbf{a}^{(t)}), \quad (5)$$

où prox_g est l'opérateur proximal de g . Dans le cas particulier de la pénalité ℓ_1 réduite à \mathbb{R}_+^D , g peut s'écrire comme $g = \frac{\lambda}{\eta} \|\cdot\|_1 + \delta_+$ avec, $\forall \mathbf{a} \in \mathbb{R}^D$,

$$\delta_+(\mathbf{a}) = \begin{cases} 0 & \text{si } \mathbf{a} \geq 0 \\ +\infty & \text{sinon.} \end{cases}$$

Dans ce cas, l'opérateur proximal a une forme explicite et la mise à jour s'écrit, $\forall i \in \{1, \dots, D\}$:

$$\mathbf{a}_i^{(t+1)} = \text{prox}_{\frac{\lambda}{\eta} \|\cdot\|_1 + \delta_+}(\tilde{\mathbf{a}}_i) = \left[\tilde{\mathbf{a}}_i - \frac{\lambda}{\eta} \right]_+ \quad (6)$$

où $(z)_+ = \max(0, z)$ est la partie positive de $z \in \mathbb{R}$.

Ceci nous permet finalement d'obtenir la règle de mise à jour suivante pour \mathbf{H} :

$$\mathbf{H}^{(t+1)} = \text{prox}_g \left(\tilde{\mathbf{H}}^{(t)} \right) = \left[\tilde{\mathbf{H}}^{(t)} - \frac{\lambda}{\eta} \right]_+ \quad \text{où} \quad g = \frac{\lambda}{\eta} \|\cdot\|_1 + \delta_+$$

et

$$\tilde{\mathbf{H}}^{(t)} = \mathbf{H}^{(t)} - \frac{1}{\eta} [\mathbf{W}^\top \mathbf{W} \mathbf{H}^{(t)} + \gamma \boldsymbol{\beta} \boldsymbol{\beta}^\top \mathbf{H}^{(t)} \mathbf{X}^\top \mathbf{X}] + \frac{1}{\eta} [\gamma \boldsymbol{\beta} \mathbf{y}^\top \mathbf{X} + \mathbf{W}^\top \mathbf{X}].$$

4 Conclusion

La mise en œuvre de la méthode présentée dans cette communication est actuellement à l'étude sur un projet de recherche sur l'acné. Au sein de ce dernier, deux types d'omiques ont été prélevés sur un échantillon de petite taille : des données métabolomiques, acquises par spectrométrie de masse, et des données méta-protéomiques, soit des protéines d'origines humaine, bactérienne ou fongique, acquises en spectrométrie de masse également. L'échantillon est constitué de deux groupes de patients parfaitement équilibrés : les patients souffrant d'acné légère à modérée d'un côté, et les patients ayant une peau qualifiée de « saine » de l'autre. L'objectif est d'extraire des profils types caractérisant l'acné au travers de signatures moléculaires composées de métabolites et de protéines d'origines multiples.

Par ailleurs, plusieurs points de la méthode nécessitent d'être approfondis. En particulier, nous avons formulé le problème sous l'angle de la minimisation d'un critère de moindres carrés mais ce critère est à comparer avec un critère basé sur la divergence de Kullback-Leibler. Également, [Leuschner et al., 2019] expriment leur terme supervisé par rapport à la projection des individus sur les signatures. Une alternative, utilisée par exemple, dans [Lee et al., 2010], consiste à utiliser directement la matrice de poids dans le terme supervisé. Enfin, le choix de contraindre le terme supervisé à un seul coefficient de régression (au lieu de deux, un par jeu de données) sera également étudié.

Bibliographie

- [Bauschke and Combettes, 2017] Bauschke, H. H. and Combettes, P. L. (2017). *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. CMS Books in Mathematics. Springer International Publishing, Cham, Switzerland.
- [Eicher et al., 2020] Eicher, T., Kinnebrew, G., Patt, A., Spencer, K., Ying, K., Ma, Q., Machiraju, R., and Mathé, E. A. (2020). Metabolomics and multi-omics integration: a survey of computational methods and resources. *Metabolites*, 10(5):202.
- [Fernsel and Maass, 2018] Fernsel, P. and Maass, P. (2018). A survey on surrogate approaches to non-negative matrix factorization. *Vietnam Journal of Mathematics*, 46:987–1021.
- [Lee and Seung, 2001] Lee, D. and Seung, H. S. (2001). Algorithms for non-negative matrix factorization. *Advances in Neural Information Processing Systems (NIPS 2000)*, 13:556–562.
- [Lee and Choi, 2009] Lee, H. and Choi, S. (2009). Group nonnegative matrix factorization for EEG classification. In van Dyk, D. and Welling, M., editors, *Proceedings of the 12th International Conference on Artificial Intelligence and Statistics*, volume 5 of *Proceedings of Machine Learning Research*, pages 320–327, Clearwater Beach, Florida, USA. PMLR.
- [Lee et al., 2010] Lee, H., Yoo, J., and Choi, S. (2010). Semi-supervised nonnegative matrix factorization. *IEEE Signal Processing Letters*, 17(1):4–7.

- [Leuschner et al., 2019] Leuschner, J., Schmidt, M., Fernsel, P., Lachmund, D., Boskamp, T., and Maass, P. (2019). Supervised non-negative matrix factorization methods for MALDI imaging applications. *Bioinformatics*, 35:1940–1947.
- [Li et al., 2022] Li, P., Tseng, C., Zheng, Y., Chew, J. A., Huang, L., Jarman, B., and Needell, D. (2022). Guided semi-supervised non-negative matrix factorization on legal documents. *Algorithms*, 15(5):136.
- [Meng et al., 2016] Meng, C., Zeleznik, O. A., Thallinger, G. G., Kuster, B., Gholami, A. M., and Culhane, A. C. (2016). Dimension reduction techniques for the integrative analysis of multi-omics data. *Briefings in Bioinformatics*, 17(4):628–641.
- [Parikh and Boyd, 2014] Parikh, N. and Boyd, S. (2014). Proximal algorithms. *Foundations and Trends[®] in Optimization*, 1(3):127–239.
- [Ritchie et al., 2015] Ritchie, M. D., Holzinger, E. R., Li, R., Pendergrass, S. A., and Kim, D. (2015). Methods of integrating data to uncover genotype-phenotype interactions. *Nature Reviews Genetics*, 16:85–97.
- [Shreeves et al., 2021] Shreeves, P., Andrews, J. L., Deng, X., Ali-Adeeb, R., and Jirasek, A. (2021). Nonnegative matrix factorization with group and basis restrictions. Preprint arXiv:2107.00744v1.
- [Singh et al., 2019] Singh, A., Shannon, C. P., Gautier, B., Rohart, F., Vacher, M., Tebbutt, S. J., and Lê Cao, K.-A. (2019). DIABLO: an integrative approach for identifying key molecular drivers from multi-omics assays. *Bioinformatics (Oxford, England)*, 35:3055–3062.
- [Tenenhaus et al., 2014] Tenenhaus, A., Philippe, C., Guillemot, V., Le Cao, K.-A., Grill, J., and Frouin, V. (2014). Variable selection for generalized canonical correlation analysis. *Biostatistics*, 15(3):569–583.
- [Yang and Michailidis, 2016] Yang, Z. and Michailidis, G. (2016). A non-negative matrix factorization method for detecting modules in heterogeneous omics multi-modal data. *Bioinformatics*, 32(1):1–8.
- [Zhang et al., 2011] Zhang, S., Li, Q., Liu, J., and Zhou, X. J. (2011). A novel computational framework for simultaneous integration of multiple types of genomic data to identify microRNA-gene regulatory modules. *Bioinformatics*, 27(13):i401–409.