Vers une nouvelle décomposition de matrice

Jean-Jacques Fuchs
IRISA/Université de Rennes I
Campus de Beaulieu - 35042 Rennes Cedex
fuchs@irisa.fr

Résumé – Pour approximer une matrice par une matrice de rang plus faible on utilise en général la décomposition en valeurs et vecteurs singuliers, ce qui est optimal pour toutes les normes unitairement invariantes comme la norme de Frobenius par exemple. Cela se justifie d'un point de vue statistique, si on suppose que chaque composante a été perturbée par un bruit additif gaussien et que l'on recherche une estimée de la matrice exacte sous-jacente au maximum de vraisemblance. Sous d'autres hypothèses il faudrait minimiser une autre norme. Nous considérons le cas où le bruit additif est supposé suivre une loi de Laplace.

Abstract – Low rank approximations of a matrix are mostly performed using the singular value decomposition which is optimal for all unitarily invariant matrix norms, such as the Frobenius norm. From a statistical point of view this approach is justified when the components are perturbed by independent and identically distributed zero mean gaussian noise. If this assumption is not valid other norms and thus approximations should be considered. The norm we consider below corresponds to noise samples that would be assumed to follow the Laplace or double-exponential distribution.

1 Introduction

En traitement du signal ou de l'image on a souvent besoin d'approximer une matrice par une matrice de rang inférieur. Si les composantes de la matrice réelle A de dimension (m,n) et de rang r sont perturbées par des variables aléatoires elle va devenir de rang plein et estimer la matrice initiale est un problème difficile qui peut nécessiter une estimation préalable du rang. Si les perturbations aléatoires sont indépendantes, centrées et gaussiennes il faut utiliser la norme de Frobénius que nous allons noter $|||A\,|||_2$:

$$|||A\,|||_2 = (\sum_{i,j} a_{i,j}^2)^{1/2} = \mathrm{trace}(A^T A)^{1/2} \tag{1}$$

et si le rang r est connu, il faut résoudre

$$\min_{\Delta} \quad |||\Delta\,|||_2 \quad \text{s.t.} \quad \mathrm{rang}(A-\Delta) = r \tag{2}$$

La solution est bien connue et se déduit de la décomposition en valeurs et vecteurs singuliers (SVD) de la matrice A. Si on suppose que les perturbations suivent une loi de Laplace (exponentielle double) $f_X(x) = \frac{\beta}{2}\, e^{-\beta|x|}$ la norme à utiliser à la place de (1) est

$$|||A\,|||_1 = \sum_{i,j} |a_{i,j}| \tag{3}$$

En effet comme les perturbations sont indépendantes, la loi jointe est égale au produit des lois et maximiser la log-vraisemblance conduit dans le cas gaussien à minimiser la norme de Frobenius et dans le cas Laplace à minimiser cette norme 'ℓ_1' (3).

Si on considère les matrices de dimension (m,n) comme des vecteurs dans un espace vectoriel de dimension $n \times m$, la norme vectorielle euclidienne ou ℓ_2, $\|x\|_2 = (\,\sum_i x_i^2\,)^{1/2}$ et la norme vectorielle ℓ_1 $\|x\|_1 = \sum_i |x_i|$ induisent les normes matricielles (1) et (3) respectivement.

A notre connaissance le problème de l'approximation

$$\min_{\Delta} \quad |||\Delta\,|||_1 \quad \text{sous} \quad \mathrm{rang}(A-\Delta) = r \tag{4}$$

associée à la norme (3) n'a pas été considéré à ce jour et nous en présentons quelques aspects.

Dans le paragraphe qui suit nous rappelons quelques résultats concernant la décomposition en valeurs et vecteurs singuliers (SVD), puis nous présentons les liens entre la norme de Frobenius, la SVD et certains problèmes d'approximation de matrices. Nous introduisons ensuite la nouvelle décomposition et un algorithme (coûteux) pour l'obtenir. Nous terminons par un exemple avant de conclure.

2 La décomposition en valeurs et vecteurs singuliers

La décomposition en valeurs et vecteurs singuliers d'une matrice réelle A de dimension (m,n) et de rang r est de la forme $A = U\Sigma V^T$ avec U et V des matrices unitaires

$$U^T A V = \begin{bmatrix} \bar{\Sigma} & 0 \\ 0 & 0 \end{bmatrix} \qquad \text{où} \quad \bar{\Sigma} = \mathrm{diag}(\sigma_1, \,\ldots\,, \sigma_r)$$
$$\text{avec} \quad \sigma_1 \geq \sigma_2 \,\ldots\, \geq \sigma_r > 0$$

On peut décomposer U en $[\;U_1\;U_2\;]$ avec U_1 ayant r colonnes et décomposer V de façon analogue pour avoir

$$A = U_1 \bar{\Sigma} V_1^T = \sum_1^r \sigma_k u_k v_k^T \tag{5}$$

qui est parfois appelé la factorisation en valeurs et vecteurs singuliers Les colonnes de U_1 forment une base orthogonale de $\Re(A)$, l'image de A et les colonnes de V_1 une base orthogonale de $\Re(A^T)$. Les colonnes de U_2 et V_2 sont alors des bases arbitraires des espaces complémentaires.

La factorisation (5) permet d'écrire simplement la pseudo-inverse A^+ de A: $A^+ = V_1\bar{\Sigma}^{-1}U_1^T$, une inverse généralisée spécifique de A. On rappelle qu'une inverse généralisée de A est une matrice A^- vérifiant $AA^-A = A$.

3 Norme de Frobenius et SVD

Nous commençons par le problème suivant. Soit A une matrice réelle de dimension (m,n) et de rang r, on cherche une perturbation Δ de norme minimale qui fait chuter le rang de A de un:

$$\min_{\Delta} \quad |||\Delta\,|||_2 \quad \text{s.t.} \quad \text{rang}(A-\Delta) = \text{rang}(A) - 1 \qquad (6)$$

Nous allons supposer que Δ est de rang un. Ceci est vrai pour la norme de Frobenius mais ne l'est pas nécessairement pour d'autres normes.

Bien que la solution soit bien connue et découle facilement de résultats plus généraux, nous allons l'établir car la démonstration utilisée est similaire à celle que nous allons utiliser plus loin pour trouver la solution dans le cas de la norme 'ℓ_1' (3). Nous allons utiliser le lemme suivant et son corollaire.

Lemme 1: Pour une matrice A donnée, toute matrice Δ de rang un qui est telle que rang(A-Δ) = rang(A) -1, est de la forme

$$\Delta = \tfrac{Axy^TA}{y^TAx} \qquad \text{avec} \qquad y^TAx \neq 0 \qquad \Box$$

Corollaire 1: La matrice Δ de rang un peut aussi s'écrire

$$\Delta = \tfrac{ab^T}{b^TA^-a} \qquad \text{avec} \qquad b^TA^-a \neq 0 \qquad .$$

$a \in \Re(A)$, $b \in \Re(A^T)$ et A^- une inverse généralisée de A.$\Box$

Comme on suppose que l'optimum de (6) est atteint en une matrice de rang un, en utilisant le corollaire et en prenant $A^- = A^+$, (6) se réécrit

$$\min_{a,b} \; |||\frac{ab^T}{b^TA^+a}\,|||_2 \;\; \text{où } b^TA^+a \neq 0,\; a \in \Re(A),\; b \in \Re(A^T)$$

comme $|||ab^T\,|||_2 = \|a\|_2\;\|b\|_2$, son optimum se déduit de

$$\max_{a,b} |b^TA^+a| \text{ avec } \|a\|_2 = \|b\|_2 = 1,\; a \in \Re(A),\; b \in \Re(A^T)$$

Nous montrons maintenant que l'optimum de ce dernier problème est atteint en $a = u_r$ and $b = v_r$ ce qui conduit à $\Delta = \sigma_r u_r v_r^T$:

$$\begin{array}{rcl} |b^TA^+a| & = & |b^TV_1\bar{\Sigma}^{-1}U_1^Ta| \\ & = & |y^T\bar{\Sigma}^{-1}z| \quad \text{avec} \quad y = V_1^Tb, \quad z = U_1^Ta \\ & = & |\sigma_1^{-1}y_1z_1 + \;..\; + \sigma_r^{-1}y_rz_r| \\ & \leq & \sigma_r^{-1}(|y_1z_1| + + \;..\; + |y_rz_r|) \\ & \leq & \sigma_r^{-1}\|y\|_2\;\|z\|_2 \quad \text{(inégalité de Schwarz)} \\ & = & \sigma_r^{-1} \end{array}$$

le maximum est donc inférieur à σ_r^{-1} mais comme par ailleurs $|v_r^TA^+u_r| = \sigma_r^{-1}$ cette valeur est atteinte pour $a = u_r$ et $b = v_r$.

Nous n'avons fait que vérifier que Δ avait la forme attendue, mais en utilisant le lagrangien il est facile d'établir ce résultat et en tout cas de montrer que l'optimum s'obtient à partir de la SVD de A. Grâce au lemme 1 le problème se réécrit

$$\max_{x,y} \; |y^TAx| \quad \text{with} \quad \|Ax\|_2 \leq 1, \quad \|A^Ty\|_2 \leq 1$$

où nous avons remplacé les contraintes égalités par des inégalités car cela ne modifie pas le problème. Le lagrangien s'écrit :

$$\ell(x,y) = -y^TAx + \tfrac{\lambda}{2}(\|Ax\|_2 - 1) + \tfrac{\mu}{2}(\|A^tY\|_2 - 1)$$

et ses points stationnaires, qui vérifient les conditions nécessaires du premier ordre, doivent satisfaire:

$$\begin{array}{l} -A^Ty + \lambda A^TAx = 0, \quad \lambda \geq 0, \quad \|Ax\|_2 = 1 \\ -Ax + \mu AA^Ty = 0, \quad \mu \geq 0, \quad \|A^Ty\|_2 = 1 \end{array}$$

On observe qu'en un point stationnaire $\lambda = \mu = |y^TAx|$ et en posant $u = Ax$, $v = A^Ty$ et $\sigma = 1/\lambda = 1/\mu$ ces conditions deviennent

$$\begin{array}{lll} A^Tu = \sigma v, & \|u\|_2 = 1, & \sigma > 0 \\ Av = \sigma u, & \|v\|_2 = 1 & \end{array} \qquad (7)$$

On reconnait le couple d'équations que doivent satisfaire les valeurs et vecteurs singuliers $\{\ \sigma, u, v\ \}$ aussi connues sous le nom d'équations de Schmidt. Comme $\sigma = 1/|y^TAx|$ doit être minimisé on retient le triplet $\{\ \sigma_r, u_r, v_r\ \}$ et il est facile de vérifier que les conditions suffisantes du second ordre sont satisfaites par u_r et v_r si la valeur singulière minimale σ_r est simple.

Nous avons donc montré que l'optimum de

$$\min_{\text{rang}(\Delta)=1} \; |||\Delta\,|||_2 \quad \text{sous} \quad \text{rang}(A-\Delta) = \text{rang}(A) - 1$$

est atteint en $\Delta = \sigma_r u_r v_r^T$ et a pour valeur σ_r la valeur singulière minimale de A. Si on remplace maintenant A par la matrice $A - \sigma_r u_r v_r^T$ et que l'on se repose le même problème de minimisation, on trouve pour solution $\Delta = \sigma_{r-1}u_{r-1}v_{r-1}^T$. La décomposition en valeurs et vecteurs singuliers peut ainsi être obtenue de proche en proche en résolvant une suite de problèmes d'optimisation.

C'est ce que nous nous proposons de faire en utilisant la norme 'ℓ_1' (3) à la place de la norme de Frobenius (1).

4 La décomposition 'ℓ_1'

4.1 La brique de base

Soit une matrice A de dimension (m,n) et de rang r, on cherche la perturbation Δ de norme 'ℓ_1' minimale qui fait chuter son rang de un. On cherche la solution de

$$\min_{\Delta} \quad |||\Delta\,|||_1 \quad \text{sous} \quad \text{rang}(A-\Delta) = \text{rang}(A) - 1 \qquad (8)$$

On suppose que l'optimum est atteint en une matrice Δ de rang un. Nous n'avons pas réussi à établir ce résultat sauf si A est de rang plein. En utilisant le corollaire 1 et le fait que $|||ab^T\,|||_1 = \|a\|_1\;\|b\|_1$, le problème (8) devient

$$\max_{a,b} \; |b^TA^-a|, \;\; \|a\|_1 \leq 1, \; \|b\|_1 \leq 1, \; a \in \Re(A), \; b \in \Re(A^T)$$

où, sans changer l'optimum on a remplacé les contraintes égalités par des inégalités. On réécrit maintenant les conditions $a \in \Re(A)$ et $b \in \Re(A^T)$ en utilisant les matrices U_2 et V_2 introduites plus haut (5), leurs colonnes sont des bases orthogonales arbitraires des espaces complémentaires de $\Re(A)$ et $\Re(A^T)$ respectivement. $a \in \Re(A)$ devient $U_2^T a = 0$ et $b \in \Re(A^T)$ devient $V_2^T a = 0$ et le problème (8) est alors équivalent à

$$\max_{a,b} \; |b^T A^- a| \;,\; \|a\|_1 \leq 1,\; \|b\|_1 \leq 1,\; U_2^T a = 0,\; V_2^T b = 0 \tag{9}$$

dont le lagrangien est
$$\ell(x, y, \lambda_1, \lambda_2, \Lambda_1, \Lambda_2) = -|b^T A^- a| + \lambda_1(\|a\|_1 - 1) + \lambda_2(\|b\|_1 - 1) + \Lambda_1^T U_2^T a + \Lambda_2^T V_2^T b$$
Pour écrire les conditions à satisfaire par ses points stationnaires on introduit la sous-différentielle, une extension de la notion de gradient en des points où il n'est pas défini

$$\begin{aligned} \partial \; \|z\|_1 \; &= \{v | v^T z = \|z\|_1,\; \|v\|_\infty \leq 1\} \\ &= \{v | v_i = \pm 1 \;\text{ si }\; z_i \neq 0 \;\text{ and }\; |v_i| \leq 1 \;\text{ sinon}\} \end{aligned}$$

Avec ces notations les points stationnaires satisfont:
$$\begin{aligned} &-\text{sign}(b^T A^- a) A^{-T} b + \lambda_1 \partial \; \|a\|_1 + U_2 \Lambda_1 = 0, \\ &-\text{sign}(b^T A^- a) A^- a + \lambda_2 \partial \; \|b\|_1 + V_2 \Lambda_2 = 0, \\ &\|a\|_1 = 1,\; \|b\|_1 = 1, \lambda_1 \geq 0,\; \lambda_2 \geq 0,\; U_2^T a = 0,\; V_2^T b = 0 \end{aligned}$$

Comme $a^T \partial(a) = \|a\|_1$ et $\text{sign}(b^T A^- a)(b^T A^- a) = |b^T A^- a|$, en pré-multipliant la première relation par a^T, et la seconde par b^T on a $\lambda_1 = \lambda_2 = |b^T A^- a|$. Les deux premières relations deviennent alors
$$\begin{aligned} &-A^{-T} b + \lambda_1 \partial \; \|a\|_1 + U_2 \Lambda_1 = 0,\; \lambda_1 > 0 \\ &-A^- a + \lambda_1 \partial \; \|b\|_1 + V_2 \Lambda_2 = 0, \end{aligned}$$
en posant $\sigma = 1/\lambda_1 = 1/\lambda_2$ et en observant que $AA^- a = a$ puisque $a \in \Re(A)$ et $AV_2 = 0$ et les propriétés analogues pour b et V_2, on obtient les conditions suivantes en pré-multipliant ces deux relations par A^T et A respectivement

$$\begin{aligned} A^T \partial \|a\|_1 &= \sigma b, \quad \|a\|_1 = 1, \quad \sigma \geq 0 \\ A \;\; \partial \; \|b\|_1 &= \sigma a, \quad \|b\|_1 = 1 \end{aligned} \tag{10}$$

Le couple a, b vérifiant ces conditions et associé au σ le plus petit est génériquement unique et la matrice de rang un $\sigma a b^T$ est alors une candidate à l'optimum de (8). Nous n'avons pas trouver les conditions suffisantes que ce point doit satisfaire.

Ces relations (10) sont le pendant des équations de Schmidt (7) qui caractérisent les triplets singuliers de la matrice A. Alors qu'il est facile de combiner les relations (7) pour faire apparaître les équations caractérisant les valeurs et vecteurs propres des matrices AA^T and $A^T A$, rien de tel ne semble faisable ici et nous avons seulement été capable de résoudre (10) à l'aide d'un algorithme de recherche exhaustive. Mais avant d'en arriver à un tel algorithme il nous faut établir le résultat suivant:

Lemme 2: La matrice optimal de rang un $\Delta = \frac{ab^T}{b^T A^- a} = \sigma a b^T$ déduite de la solution du problème (9) est le produit d'un vecteur colonne par un vecteur ligne ayant chacun au moins $r - 1$ composantes nulles. □

Démonstration: Nous montrons que l'optimum (a_o, b_o) de (9) est tel que a_o a au plus m-$(r-1)$ composantes non nulles et b_o au plus n-$(r-1)$ composantes non nulles. Pour b_o fixé, le vecteur a est solution de

$$\max b_o^T A^- a \text{ sous } \|a\|_1 = 1,\; U_2^T a = 0$$
dans lequel on reconnaît un programme linéaire en a pour lequel l'optimum estgénériquement atteint en une solution de base admissible. Dans notre cas cela signifie que l'optimum a au plus $m - r + 1$ composantes non nulles puisque l'on dénombre $m - r + 1$ contraintes linéaires égalités. Un raisonnement analogue permet d'établir le résultat annoncé pour b. □

4.2 L'algorithme de recherche exhaustive

Nous proposons un algorithme qui donne toutes les solutions (σ, a, b) qui vérifient (10):

$$\begin{aligned} A^T \partial \|a\|_1 &= \sigma b, \qquad \|a\|_1 = 1 \qquad \sigma \geq 0 \\ A \partial \; \|b\|_1 &= \sigma a, \qquad \|b\|_1 = 1, \end{aligned}$$

pour A une matrice (m,n) de rang r. A (σ, a, b) on associe la matrice de rang un $\Delta = \sigma a b^T$.

On considère toutes les matrices carrées d'ordre $(r-1)$ que l'on peut extraire de A. En échangeant les lignes d'une part et les colonnes d'autre part, on amène chacune de ces sous-matrices dans le coin supérieure gauche et on la note $A_{1,1}$. On résout alors (10) pour chacune de ces permutations. Comme le lemme précédent nous a appris que a et b ont au moins r-1 composantes nulles, on suppose que ce sont celles de a_1 et b_1, les sous-vecteurs associés à $A_{1,1}$. Les composantes restantes dans a_2 et b_2 sont alors génériquement non nulles. On décompose de la même façon $\partial\|a\|_1$ en deux sous-vecteurs ∂a_1 et $\text{sign}(a_2)$ et de même pour $\partial\|b\|_1$. Comme $a_1 = 0$, les composantes de $\partial\|a\|_1$ sont dans l'intervalle $(-1, 1)$. Comme les composantes de a_2 sont supposées différentes de zéro, les composantes correspondantes dans $\partial\|a\|_1$ sont égales à ± 1 d'où la notation $\text{sign}(a_2)$ pour désigner la fonction signe usuelle;

Pour chaque permutation, on décompose donc la matrice A en 4 blocs $A_{i,j}$ et la seconde équation de (10) $A\partial\|b\|_1 = \sigma a$ se transforme en deux sous-relations qui s'écrivent, avec les notations que nous venons d'introduire

$$\begin{aligned} A_{1,1} \partial b_1 + A_{1,2} \;\text{sign}(b_2) \;\; &= \;\; \sigma a_1 = 0 \\ A_{2,1} \partial b_1 + A_{2,2} \;\text{sign}(b_2) \;\; &= \;\; \sigma a_2 \end{aligned}$$

Si $A_{1,1}$ n'est pas inversible, on passe à la permutation suivante, sinon la première de ces deux relations permet de calculer ∂b_1 pour chacune des $2^{(n-r)}$ valeurs possibles de $\text{sign}(b_2) \in [\pm 1 \quad \pm 1 \;\; .. \pm 1 \;]^T$. Pour chaque $\text{sign}(b_2)$ on vérifie si $\|\partial b_1\|_\infty \leq 1$ et si c'est la cas on a obtenu un couple cohérent ∂b_1, $\text{sign}(b_2)$. Pour chacun de ces couples on utilise la seconde relation pour en déduire $\sigma > 0$ et a_2 avec $\|a_2\|_1 = 1$. On passe alors à la première équation de (10) $A^T \partial\|a\|_1 = \sigma b$ qui donnent également deux relations supplémentaires

$$\begin{aligned} \partial a_1^T A_{1,1} + \;\text{sign}(a_2)^T A_{2,1} \;\; &= \;\; \sigma b_1^T = 0 \\ \partial a_1^T A_{1,2} \; + \;\text{sign}(a_2)^T A_{2,2} \;\; &= \;\; \sigma b_2^T \end{aligned}$$

Comme $\text{sign}(a_2)$ est maintenant connu, on vérifie si $\|\partial a_1\|_\infty \leq 1$ dans la première relation. Si tel est le cas, on déduit $\sigma' > 0$, b_2 de la seconde. Il reste alors à voir si $\sigma' = \sigma$ et si les signes des composantes de b_2 sont les mêmes que ceux de $\text{sign}(b_2) \in [\pm 1 \quad \pm 1 \quad .. \pm 1\ ]^T$ retenu plus haut dans la première relation issue de la seconde équation de (10). Si toutes ces conditions sont satisfaites, on a trouvé un point stationnaire du Lagrangien.

Il y a en général plusieurs points stationnaires mais leur nombre n'est ni fixe ni prévisible contrairement à ce qui est le cas pour (7). On ne retient que la solution associée au σ le plus petit. Elle semble génériquement unique.

On peut développer un autre algorithme de recherche exhaustive qui donne, bien entendu, le même résultat. Dans le développement précédent, on cherche pour chaque permutation, la perturbation Δ dont seul le bloc $\Delta_{2,2}$ est non nul. On recherche donc la sous-matrice $\Delta_{2,2}$ qui fait chuter le rang de A de un quand elle est soustraite de $A_{2,2}$.

Si $A_{1,1}$ est inversible une telle matrice Δ existe toujours. La sous-matrice $\Delta_{2,2}$ associée vaut $A_{2,2} - A_{2,1}A_{1,1}^{-1}A_{1,2}$ et peut être vue comme le complément de Schur de $A_{1,1}$ dans A. On montre facilement qu e $\Delta_{2,2}$ est de rang un qui peut donc se mettre sous la forme $\sigma a_2 b_2^T$ avec $\|a_2\|_1 = 1$ et $\|b_2\|_1 = 1$. Mais rien ne garantit que le triplet σ, a_2, b_2 vérifie les conditions imposées par (10). Les rares cas où ces conditions sont satisfaites, mènent aux mêmes matrices que l'algorithme précédent. Cela semble notamment toujours être le cas pour le $\Delta_{2,2}$ de norme 'l_1' minimale

5 Un exemple

Pour illustrer la décomposition proposée nous l'appliquons à une matrice carrée d'ordre 4 inversible. On peut remarquer que le lemme 2 appliqué plus généralement à une matrice inversible d'ordre n nous dit que la matrice $\Delta = \sigma ab^T$ de rang un et de norme 'l_1' minimale qui fait chuter son rang de un est une matrice ayant juste une composante non nulle. La solution à ce problème spécifique s'obtient facilement directement. Après avoir soustrait cette matrice de A, la nouvelle matrice est d'ordre n et de rang $n-1$ et le lemme 2 nous apprend que la matrice $\Delta = \sigma ab^T$ de norme 'l_1' minimale qui fait passer son rang à $n-2$ est le produit de deux vecteurs σa et b ayant au plus deux composantes non nulles chacun. En continuant ainsi on en déduit que la k-ième matrice dans la décomposition est le produit de deux vecteurs ayant chacun au plus k composantes non nulles.

Pour la matrice de Vandermonde d'ordre 4

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ .2 & .4 & .6 & .8 \\ .04 & .16 & .36 & .64 \\ .008 & .064 & .216 & .512 \end{bmatrix}$$

la décomposition que l'on obtient et de la forme $A = GDH^T$ où dans les colonnes de G de norme ℓ_1 unité se trouvent les vecteurs a_i, dans les colonnes de H de norme ℓ_1 unité se trouvent les vecteurs b_i et les σ_i sont sur la diagonale de D, $A = \sum_i \ \sigma_i a_i b_i^T$

$$G = \begin{bmatrix} 0 & 0 & -0.5938 & -0.4615 \\ 0 & 0.5682 & 0 & -0.2592 \\ -1 & 0 & 0.1781 & -0.1661 \\ 0 & -0.4318 & 0.2280 & -0.1131 \end{bmatrix}$$
$$D = \begin{bmatrix} 0.0100 & 0 & 0 & 0 \\ 0 & 0.2247 & 0 & 0 \\ 0 & 0 & 2.0157 & 0 \\ 0 & 0 & 0 & 7.9121 \end{bmatrix}$$
$$H = \begin{bmatrix} 0 & -0.7000 & -0.4051 & -0.1411 \\ 1 & 0 & -0.2405 & -0.1950 \\ 0 & 0.3000 & 0 & -0.2739 \\ 0 & 0 & 0.3544 & -0.3900 \end{bmatrix}$$

6 Conclusions

Les résultats que nous avons proposés, peuvent aussi être utilisés pour résoudre ce que l'on pourrait appeler le problème des moindres déviations totales par analogie avec le problème des moindres carrés totaux. Si le problème des moindres déviations consistent à résoudre

$$\min_x \ \|Ax - b\|_1$$

avec A une matrice (m,n) de rang n et x et b des vecteurs de dimensions adéquates, qui s'écrit aussi $\min_{r,x} \ \|r\|_1$ sous $r = Ax - b$, celui des des moindres déviations totales correspond à résoudre

$$\min_{x,\ r,\ R} \ ||| \ [R\ r] \ |||_1 \quad \text{sous} \quad (A+R)x = b + r$$

et a déjà été considère dans [9].

Références

[1] R.A. Horn and C.R. Johnson *Matrix Analysis.* Cambridge University Press., 1985.

[2] R. Fletcher, *Practical Methods of Optimization.* John Wiley and sons., 1987.

[3] R.B. Bapat, *Linear Algebra and Linear Models.* Springer, Universitext, 1999, p. 114.

[4] G.H. Golub and C.F. Van Loan. *Matrix Computations.* John Hopkins University Press, 1983.

[5] P. Dewilde and E.F. Deprettere. "Singular Value Decomposition : an introduction" *SVD and Signal Processing* pp. 3-41, North-Holland, 1988.

[6] D. G. Luenberger. *Introduction to Linear and Nonlinear Programming.* Addison Wesley, 1973.

[7] G.W. Stewart and J.G. Sun. *Matrix Perturbation Theory.* Academic Press, 1990.

[8] D.P. Bertsekas. *Convex Analysis and Optimization.* Athena Scientific, 2003.

[9] M.R. Osborne and G.A. Watson. An analysis of the total approximation problem in separable norms and an algorithm for the total ℓ_1 problem. *SIAM J. Sci. Stat. Comput.*, vol. 6, 2, 410–424, april 1985.