
Table des Matières

I Propriétés statistiques du problème des moindres carrés	1
1 Introduction	1
2 Notions de Statistique	2
2.1 indépendance des variables aléatoires	2
2.2 Espérance d'une variable aléatoire	2
2.3 Propriétés de l'espérance mathématique	2
2.4 Variance et covariance d'une variable aléatoire	3
2.5 Propriété de la variance	3
2.6 Inégalité de Markov	4
3 Méthodes numériques	7
4 Bibliothèques numériques	10
4.1 Produit de Kronecker	12
5 Résolution des problèmes des moindres carrés	13
5.1 Sketching matrice	13
5.2 Approximation des moindres carrés	16
II Matrices aléatoires	19
1 Introduction	19
2 Préliminaires	19
3 Produit matriciel aléatoire	20
4 Matrices aléatoires	24
4.1 Projection Gaussienne	25

4.2	Transformé d'Hadamard aléatoire	25
4.3	Transformée de Fourier randomisée sous-échantillonnée	26
4.4	Leverage score échantillonné	26
4.5	Count-sketch	27
4.6	Matrices de code sous-échantillonnées	28
4.7	Comparaison entre les différentes matrices aléatoires	28
5	Approximation des moindres carrés	29
6	Conclusion	30
III	Conditionnement des problèmes des moindres carrés	31
1	Introduction	31
2	Formule explicite du conditionnement des moindres carrés	33
3	Estimation statistique du conditionnement	36
IV	Résolution des problèmes larges des moindres carrés	39
1	Introduction	39
2	Résolution des problèmes des moindres carrés	40
2.1	Description d'algorithme 1	44
2.2	Description d'algorithme 2	46
2.3	Description d'algorithme 3	47
3	Application numérique	48
Conclusion et perspectives		i
Annexe 1		iii

Liste des publications

1 - Publication Internationale

Fast randomized linear least squares.

Nadjet RABEHI et Mekki TERBECHE,

Journal of Statistics and Management Systems, Taylor and Francis,

ISSN 0972-0510 (Print), ISSN 2169-0014 (Online) Vol. 21 (2018), No. 1, pp. 107-123.

DOI : 10.1080/09720510.2017.1367481

2 - Article soumis pour publication internationale

Randomized method for Resolution of large Least Squares Problem.

N. Rabehi , L. BELAIB et M. TERBECHE.

Soumis le 27 Février 2018 au Asian-European Journal of Mathematics (AEJM).

3 - Publications nationales

Modification du noyau intégrale de Stokes

N. RABEHI, M. AARIZOU

Bulletin des Sciences Géographiques N 21, ISSN 1112- 3745. pp 2- 7 (juillet 2008).

Choix d'un Modèle Géospatial Global pour la Détermination Précise du Géoïde en Algérie.

N. RABEHI, M. KRERI, M. TOUABET et M. TERBECHE

Bulletin des Sciences Géographiques N 27, ISSN 1112- 3745, pp 33 - 39 (2012).

Etude qualitative de la précision des Modèles Numériques de Terrain globaux : Cas de l'Algérie.

M. HADDAD, N. RABEHI et H. TAIBI

Bulletin des Sciences Géographiques N 29, ISSN 1112- 3745, 1er semestre (2014)

Evaluation des Surfaces d'Anomalies Gravimétriques à l'Air Libre en Méditerranée Occidentale

M. Haddad, N. Rabehi, H. Taibi, H. Bousmaha et H. Hachemi

Revue Nature et Technologie. C- Sciences de l'Environnement N 14, pp 09 - (15Janvier 2016)

4 - Actes et Proceedings Edités à l'Etranger

Stokes' Kernel Modifications

N. RABEHI, A. BELHADJ et M. TERBECHE

TS04B - Heights, Geoid and Gravity, 6132, FIG Working 2012, Rome, Italy (6-10 May 2012).

Table des figures

IV.1	Comportement du conditionnement des composantes x_i de la solution du problème calculé avec $S^t HD$ comme matrice aléatoire.	50
IV.2	Variation des composantes de la solution des moindres carrés x_i dans différents cas	51
IV.3	Conditionnement des composantes x_i du problème des moindres carrés linéaires	52
IV.4	Variation du conditionnement des composantes x_i calculé à partir de la combinaison de count-sketch et la projection Gaussienne	53

Abréviations

- FJLT* : Fast subspace Johnson Lindenstrauss Transform
- JLT* : Johnson Lindenstrauss Transform
- LLS* : Moindre carrés linéaire
- SVD* : Décomposition en Valeurs singulières
- BLAS* : Basic Linear Algebra Subprograms
- LAPACK* : Linear Algebra PACKage
- ScalAPACK* : Scalable LAPACK
- Algol* : Langage de programmation
- FFT* : Fast Fourier Transform
- GOCE* : Gravity field and steady state Ocean circulation Explorer
- GPS* : Global Positioning System
- A^t : transposée de la matrice A
- A^\dagger : Pseudo inverse de la matrice A
- $\text{rang}(A)$: rang de la matrice A
- I_d : matrice identité

Résumé

Dans cette thèse, nous nous intéressons à la résolution des problèmes des moindres carrés de grande taille .

Afin d'accélérer la résolution de ce problème on est conduit à utiliser les matrices aléatoires, cette approche a reçue une grande intention dans les années récentes. Les matrices aléatoires se divise en deux parties : matrices échantillonnage et de projection, dans chaque partie on trouve des matrices aléatoires avec différentes estimations du temps de calcul. La solution obtenue par différentes méthodes aléatoires permet un gain de temps et de mémoire dans le calcul. Dans cette thèse, différentes méthodes aléatoires sont analysées et trois algorithmes sont proposés qui démontre la performance de cette théorie. Par la suite, nous optons pour le calcul de la solution en utilisant une combinaison entre l'échantillonnage moyen et la matrice aléatoire d'Hadamard tout en respectant la moyenne statistique. Une application est introduite pour le calcul des coefficients harmoniques du modèle géopotentiel .

Introduction générale

La méthode des moindres carrés est une technique fondamentale d'algèbre linéaire. Elle découle initialement de la nécessité d'adapter un modèle mathématique linéaire à des observations données [1]. Elle se formule par la minimisation de la somme des carrés des éléments du vecteur résiduel.

La résolution des grands systèmes linéaires denses apparaît dans de nombreuses applications scientifiques et techniques telles que l'exploration des données, les applications électromagnétiques, la géophysique et la géodésie. Il y a deux méthodes pour résoudre le problème des moindres carrés, la méthode directe, par exemple : les équations normales, la méthode de Householder et la décomposition en valeurs singulières (SVD). Des travaux approfondis ont été développés afin d'analyser ces méthodes dans [7, 8, 44, 14, 30]. Le principal inconvénient de la méthode directe est le temps requis en particulier lorsqu'on utilise des matrices denses et larges.

Pour accélérer la solution du problème, nous introduisons l'approximation aléatoire du problème des moindres carrés, ce qui conduit l'utilisation des matrices aléatoires. Par conséquent, le problème devient simple et il se résoud plus rapidement.

Sarlos[39] et Drineas et al. [9, 10] ont introduit les premiers algorithmes aléatoires, basés sur l'application de la transformée d'Hadamard aléatoire aux colonnes de la matrice d'entrée afin de créer un problème des moindres carrés de taille plus petite qui peut être résolu exactement et dont la solution se rapproche de celle du problème original avec une très forte probabilité. Cela a été suivi par le travail de Rokhlin et Tygert [38], qui utilisaient une transformée de Fourier aléatoire (SRFT) échantillonnée pour former un préconditionneur, et par la suite utilisaient un solveur itératif standard pour résoudre le problème.

On considère le problème des moindres carrés comme suit : $\min_{x \in \mathbb{R}^d} \|Ax - b\|_2$ où $A \in \mathbb{R}^{n \times d}$, $d \ll n$ et $b \in \mathbb{R}^n$.

Si A est de rang complet, x est simplifié par $x = (A^t A)^{-1} A^t b$, cette solution est calculée par les méthodes déterministes dans $O(nd^2)$ temps, mais en utilisant l'approche aléatoire on peut résoudre ce problème dans $O(nd \log(d) + \text{poly}(\frac{d}{\varepsilon}))$ temps, cette dernière donne la solution plus rapidement pour $d \ll n$ et ε pas suffisamment petit.

Les matrices aléatoires peuvent être utilisées immédiatement pour obtenir une résolution du problème de faible précision, ou elles peuvent être utilisées indirectement pour développer un préconditionneur en plus d'un algorithme numérique itératif afin d'acquérir une solution de haute précision [32, 10, 44]. Il est important d'évaluer la qualité numérique de la solution calculée. Dans ce contexte, la notion du conditionnement de la solution est recommandée. Le conditionnement de la solution permet de mesurer la sensibilité numérique de la solution aux perturbations des données et d'évaluer ainsi la qualité numérique de la solution calculée. De nombreux résultats sur le conditionnement de la solution des moindres carrés ont été publiés dans [2, 4, 3, 5]

Cette thèse traite le thème de la résolution des problèmes de moindres carrés linéaires. Elle comporte quatre chapitres.

Le premier chapitre est un rappel des notions de base que nous utilisons dans ce travail, des définitions qui sont utile par la suite ainsi que l'historique de la théorie des moindres carrés.

Dans le deuxième chapitre, on étudie les matrices aléatoires et les propriétés de chacune. Nous abordons dans ce chapitre les deux méthodes aléatoires : matrices d'échantillonnage et de projection tout en tenant compte du temps de calcul. À la fin de ce chapitre on va comparer les deux approches dans le but de trouver une solution rapide et stable au problème avec une probabilité élevée.

Dans le troisième chapitre, nous introduisons le calcul du conditionnement de la solution afin d'exprimer la qualité de la solution du problème des moindres carrés.

Dans le quatrième chapitre nous présentons notre approche en décrivant nos algorithmes qui permettent de calculer rapidement la solution. Le premier algorithme combine l'échantillonnage moyen et la matrice aléatoire d'Hadamard afin de déterminer la solution de haute précision, ce dernier tient compte de la moyenne statistique. Pour consolider nos résultats nous avons combiné la méthode de count-sketch avec la transformation d'Hadamard aléatoire (SRHT), tandis que le troisième algorithme utilise la projection Gaussienne en combinaison avec la méthode count-sketch. Les différents algorithmes proposés utilisent la décomposition QR. Une application numérique résolvant un problème des moindres carrés de géodésie

spatiale en plus d'une étude du conditionnement ont été réalisées dans le but d'une détermination précise des coefficients harmoniques géopotentiels du champ de pesanteur tout en intégrant des données issues du satellite GOCE afin d'illustrer la performance de nos algorithmes. Nous terminons cette thèse par une conclusion générale et des perspectives.

Chapitre I

Propriétés statistiques du problème des moindres carrés

1 Introduction

Le problème des moindres carrés est l'un des problèmes les plus importants dans l'algèbre linéaire, il est basé sur la nécessité d'adapter un modèle mathématique linéaire aux données d'observations, il s'exprime comme suit :

$$\min_x \|Ax - b\|, \quad A \in \mathbb{R}^{n \times d}, b \in \mathbb{R}^n. \quad (\text{I.1})$$

tel que : $r = b - Ax$ qui satisfait $A^t r = 0$

Gauss a donné la première définition du principe des moindres carrés en tant qu'une procédure statistique. Il a supposé que les erreurs n'étaient pas corrélées et normalement distribuées avec une moyenne zéro et de variance égale à un. Par la suite, Gauss a renforcé sa théorie du principe des moindres carrés dans deux mémoires 'Theoria Combinationis' [17, 18] et [15, 16], il a montré que l'optimalité de l'estimation des moindres carrés sans supposer une distribution particulière des erreurs aléatoires.

En 1996, Björck a traité les méthodes numériques de résolution des problèmes des moindres carrés dans [6], alors que Higham [19] a étudié la précision et la stabilité des résultats. On peut trouver les méthodes moderne de calcul avec des exemples d'applications pratiques

dans Hansen et al. [20] qui a fournit une version détaillée des problèmes des moindres carrés linéaires.

2 Notions de Statistique

2.1 indépendance des variables aléatoires

Définition 2.1 *On dit que deux variables aléatoires X et Y sont indépendantes si pour tout (x, y) de valeurs possible de (X, Y) ,*

$$P(X = x \text{ et } Y = y) = P(X = x) P(Y = y)$$

ce qui signifie que $X = x$ et $Y = y$ sont indépendantes.

2.2 Espérance d'une variable aléatoire

Définition 2.2 *Soit X une variable aléatoire, l'espérance $E[X]$ est définie par*

$$E[X] = \sum_{x \in X(\Omega)} x \cdot \Pr[X = x] \quad (\text{I.2})$$

telle que $X(\Omega)$ est l'image de la variable aléatoire X sur l'espace d'échantillon Ω . Rappelons que X est une fonction, $E[X]$ peut être exprimée sous forme d'une somme sur le domaine Ω . D'un autre coté, pour les espaces d'échantillons finis Ω nous obtenons :

$$E[X] = \sum_{\omega \in \Omega} X(\omega) \Pr[\omega]$$

2.3 Propriétés de l'espérance mathématique

Notons quelques propriétés de l'espérance mathématique
L'espérance mathématique est linéaire

$$\begin{aligned} E[X + Y] &= E[X] + E[Y] \\ E[\lambda X] &= \lambda E[X] \quad \text{pour tout } \lambda \in \mathbb{R} \end{aligned} \quad (\text{I.3})$$

Si X et Y sont indépendants alors : $E[XY] = E[X]E[Y]$

2.4 Variance et covariance d'une variable aléatoire

Définition 2.3 (*Variance d'une variable aléatoire*) Soit X une variable aléatoire, la variance $\text{var}[X]$ est définie comme

$$\text{var}[X] = E[(X - E[X])^2] \quad (\text{I.4})$$

La variance mesure la moyenne du carré de la différence $X - E[X]$.

2.5 Propriété de la variance

$$\text{var}[X] = E[X^2] - [E[X]]^2$$

En conséquence,

$$\text{var}[X] \leq E[X^2]$$

La variance de X est positive, elle est nulle si et seulement si X est constante. La variance n'est pas linéaire sauf si les variables aléatoires impliquées sont indépendantes. En effet, si les variables aléatoires X et Y sont indépendantes alors[44]

$$\begin{aligned} \text{var}[X + Y] &= \text{var}[X] + \text{var}[Y] \\ \text{var}[\lambda X] &= \lambda^2 \text{var}[X], \quad \text{pour tout } \lambda \in \mathbb{R} \end{aligned}$$

Définition 2.4 L'écart type noté σ est défini comme la racine carrée de la variance.

$$\sigma(X) = \sqrt{\text{var}(X)}$$

Remarque 2.1 - Une variable aléatoire d'espérance nulle est dite centrée.

- Une variable aléatoire de variance égale à 1 est dite réduite.

Définition 2.5 La matrice de variance-covariance de la solution estimée \hat{x} est donnée par la formule

$$\text{var}(\hat{x}) = \sigma^2 (A^t A)^{-1} \quad (\text{I.5})$$

En effet, $\text{var}(\hat{x}) = (A^t A)^{-1} A^t \text{var}(\hat{b}) A (A^t A)^{-1} = \sigma^2 (A^t A)^{-1}$

Définition 2.6 la variance est une forme quadratique sur l'espace vectoriel des variables aléatoires réelles. La covariance de deux variables aléatoires X et Y , notée $\text{Cov}(X, Y)$ est la

forme bilinéaire symétrique associée. Elles sont données par les formules :

$$\begin{aligned} \text{Cov}(X, Y) &= E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y] \\ \text{et on a} \\ \text{var}(X \pm Y) &= \text{var}(X) \pm 2\text{Cov}(X, Y) + \text{var}(Y) \end{aligned}$$

Définition 2.7 Si X et Y sont deux variables aléatoires de variances non nulles, on définit le coefficient de corrélation de X et Y comme étant

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma(X)\sigma(Y)} \quad (\text{I.6})$$

tel que $\sigma(X)$ et $\sigma(Y)$ sont les écarts types de X et Y respectivement.

2.6 Inégalité de Markov

Lemme 2.1 Soit X une variable aléatoire non négative, pour tout $\alpha > 0$,

$$\Pr[X \geq \alpha] \leq \frac{E[X]}{\alpha} \quad (\text{I.7})$$

C'est une inégalité très simple à appliquer et n'a besoin que d'une limite supérieure de l'espérance de X .

Une formulation équivalente à celle de (I.7), soit X une variable aléatoire non négative, pour tout $k > 1$,

$$\Pr[X \geq k.E[X]] \leq \frac{1}{k} \quad (\text{I.8})$$

ou équivalent,

$$\Pr[X \leq k.E[X]] \leq 1 - \frac{1}{k} \quad (\text{I.9})$$

La probabilité qu'une variable aléatoire dépasse k fois ses attentes est au plus $\frac{1}{k}$.

Preuve 2.1 Afin de prouver l'inégalité de Markov, nous allons montrer que :

$$\Pr[X \geq t] \leq \frac{E[X]}{t} \quad (\text{I.10})$$

Assumons que $k = \frac{t}{E[X]}$, pour tout $t > 0$. Afin de prouver l'inégalité ci-dessus, nous définissons la fonction suivante

$$f(X) = \begin{cases} 1 & \text{if } X \geq t \\ 0 & \text{autre} \end{cases}$$

avec une espérance $E[f(X)] = 1.Pr[X \geq t] + 0.Pr[X < t] = Pr[X \geq t]$

Revenons à la définition de la fonction, nous avons $f(X) \leq \frac{X}{t}$. Prenons l'espérance des deux cotés, nous obtiendrons :

$$\begin{aligned} E[f(X)] &\leq E\left[\frac{X}{t}\right] = \frac{E[X]}{t}, \\ Pr[X \geq t] &\leq \frac{E[X]}{t} \end{aligned} \quad (I.11)$$

Définition 2.8 (Modèle linéaire) Le modèle statistique linéaire suppose que le vecteur $b \in \mathbb{R}^n$ d'observation est relié à $x \in \mathbb{R}^d$ par la relation linéaire

$$Ax = b + \varepsilon$$

Où $A \in \mathbb{R}^{n \times d}$ est une matrice et ε est le vecteur d'erreur aléatoire.

Dans le modèle linéaire standard nous avons :

$$E[\varepsilon] = 0, \quad var(\varepsilon) = \sigma^2 I$$

Ceci veut dire que les variables aléatoires ε_i sont incorrélées et leurs moyennes égale à zéro.

Définition 2.9 (Modèle de Gauss-Markov)[6] Dans le modèle de Gauss-Markov on suppose la relation linéaire $Ax = z$ tel que $A \in \mathbb{R}^{n \times d}$, $x \in \mathbb{R}^d$ vecteur d'inconnus, et z une constante inconnue. Soit $b = z + \varepsilon \in \mathbb{R}^n$ vecteur d'observations, où ε est un vecteur d'erreur aléatoire tel que

$$E[\varepsilon] = 0, \quad var(\varepsilon) = \sigma^2 V$$

où, $V \in \mathbb{R}^{n \times n}$ une matrice symétrique, définie non négative et σ^2 une constante inconnue.

Dans le cas standard, les erreurs sont supposées non corrélées et avec la même variance i.e. $V = I_n$

Nous allons montrer quelques propriétés qui seront utiles par la suite.

Lemme 2.2 Soit $B \in \mathbb{R}^{r \times n}$ une matrice et y un vecteur aléatoire, tel que $E[y] = \mu$ et $\sigma^2 V$ sa matrice de variance. Alors l'espérance et la variance de By s'écrivent comme

$$E[By] = B\mu, \quad var(By) = \sigma^2 BVB^t \quad (I.12)$$

Dans le cas spéciale $V = I$, $B = c^t$ et c est un vecteur ligne, alors :

$$var(c^t y) = \mu^2 \|c\|_2^2$$

Preuve 2.2 la première propriété s'implique de la définition de l'espérance mathématique.
Montrons la deuxième égalité

$$\begin{aligned} \text{var}(By) &= E \left[B(y - \mu)(y - \mu)^t B^t \right] \\ &= B E \left[(y - \mu)(y - \mu)^t \right] B^t \\ &= \sigma^2 BVB^t \end{aligned}$$

Théorème 2.1 (Théorème de Gauss-Markov) Considérons le modèle linéaire de Gauss-Markov $Ax = z$ où $A \in \mathbb{R}^{n \times d}$ de rang d . Soit $b = z + e$ où e est un vecteur aléatoire de moyenne 0 et de variance $\sigma^2 I$.

Alors la meilleure estimation linéaire sans biais de x est le vecteur \hat{x} qui minimise la somme des carrés $\|Ax - b\|_2^2$. Ce vecteur est unique et égale à la solution des équations normales.

Plus généralement, $c^t x$ est le meilleur estimateur linéaire sans biais de toute fonction linéaire $\theta = c^t x$. La matrice de variance de \hat{x} est $\text{var}(\hat{x}) = \sigma^2 (A^t A)^{-1}$ alors que, l'estimateur sans biais de σ^2 est donné par

$$s^2 = \frac{\hat{r}^t \hat{r}}{m - n} \quad (\text{I.13})$$

avec $\hat{r} = b - A\hat{x}$

Preuve 2.3 Soit $\hat{\theta} = d^t b$ un estimateur sans biais de $\theta = c^t x$, alors

$$E[\hat{\theta}] = d^t E[b] = d^t Ax = c^t x$$

Nous avons $A^t d = c$, du lemme (2.2) il en résulte que $\text{var}(c) = \sigma^2 \|d\|_2^2$.

On doit minimiser $d^t d$ ainsi que $A^t d = c$.

Soit : $Q = d^t d - 2z^t (A^t d - c)$, où z est le vecteur multiplicateur de Lagrange.

Pour que Q soit minimum, il suffit que

$$\frac{\partial Q}{\partial d} = 2(d^t - z^t A^t) = 0 \Rightarrow d = Az$$

alors $A^t Az = A^t d = c$, donc les colonnes de A sont linéairement indépendantes.

Puisque $x \neq 0$ alors $Ax \neq 0$ et donc $x^t A^t Ax = \|Ax\|_2^2 > 0$.

Par conséquent, la matrice $A^t A$ est définie positive et non singulière, nous obtenons $z = (A^t A)^{-1} c$ ainsi le meilleure estimateur linéaire sans biais est :

$$d^t b = c^t (A^t A)^{-1} A^t b = c^t \hat{x}$$

Où \hat{x} est la solution des équations normales.

Il reste à montrer que le même résultat est obtenu si la somme des carrés $Q(x) = (b - Ax)^t (b - Ax)$ est minimisée. Dérivons par rapport à x

$$\frac{\partial Q}{\partial x} = -2A^t (b - Ax) = 0$$

Ce qui donne les équations normales. On peut montrer que c'est un minimum car

$$\|b - Ay\|_2^2 = \|b - Ax\|_2^2 + \|A(x - y)\|_2^2 \geq \|b - Ax\|_2^2$$

si x satisfait les équations normales.

3 Méthodes numériques

Les problèmes des moindres carrés peuvent être résolus par diverses méthodes, comme exemples : les équations normales, la triangulation de Householder et la décomposition en valeurs singulières (SVD), chaque méthodes ayant des avantages et des inconvénients.

Parmi les méthodes directes de résolution du problème des moindres carrés, on trouve la méthode des équations normales.

Définition 3.1 (*équation normale*) Soit $A \in \mathbb{R}^{n \times d}$ une matrice de rang d , la solution du problème des moindres carrés satisfait l'équation normale

$$A^t A x = A^t b \quad (\text{I.14})$$

La matrice $A^t A \in \mathbb{R}^{d \times d}$ est symétrique, elle est définie positive car le rang(A) = d . Alors la solution du problème des moindres carrés est unique, elle s'exprime comme

$$x = (A^t A)^{-1} A^t b \quad (\text{I.15})$$

Elle a l'avantage d'être facile à programmer que les autres méthodes directes. Dans la littérature, cette méthode devrait fournir une solution avec une meilleure précision que certaines méthodes itératives, bien que ce gain de précision pourrait être très inférieur au niveau de bruit d'observation. Un autre avantage de cette méthode est que cela nous permet d'obtenir la matrice de covariance $(A^t A)^{-1}$ qui est parfois nécessaire aux physiciens. Ceci explique pourquoi elle est souvent favorisée par les statisticiens. Cependant, la résolution des équations normales n'est pas toujours stable en présence d'erreur d'arrondi. En plus, elle nécessite

des capacités de calcul élevé pour l'assemblage des équations normales et aussi de grandes capacités de mémoire pour stocker le $d \times d$ matrice symétrique dans la mémoire de l'ordinateur.

Théorème 3.1 (*Décomposition QR*) *Soit $A \in \mathbb{R}^{n \times d}$ une matrice, $\text{rang}(A) = d$, A admet une décomposition QR tel que :*

$$A = Q \begin{pmatrix} R \\ 0 \end{pmatrix}, \quad Q = (Q_1 \quad Q_2) \quad (\text{I.16})$$

Alors la solution unique x au $\min_x \|Ax - b\|_2$ et son vecteur résidu correspondant au $r = b - Ax$ sont obtenus par

$$\begin{pmatrix} d_1 \\ d_2 \end{pmatrix} = Q^t b, \quad Rx = d_1, \quad r = Q \begin{pmatrix} 0 \\ d_2 \end{pmatrix} \quad (\text{I.17})$$

alors $\|r\|_2 = \|d_2\|_2$

tel que $Q \in \mathbb{R}^{n \times n}$ est une matrice orthogonale et $R \in \mathbb{R}^{n \times d}$ une matrice triangulaire supérieure tel que les éléments diagonaux sont non-négatifs. La matrice R est même appelé *R-facteur*. La factorisation QR est une méthode standard de résolution des problèmes de moindres carrés [30].

Preuve 3.1 *Puisque la matrice Q est orthogonale, nous avons*

$$\begin{aligned} \|Ax - b\|_2^2 &= \|Q^t(Ax - b)\|_2^2 \\ &= \left\| \begin{pmatrix} Rx \\ 0 \end{pmatrix} - \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} \right\|_2^2 \\ &= \|Rx - d_1\|_2^2 + \|d_2\|_2^2 \end{aligned}$$

Si $Rx = d_1$, la première partie de la formule est minimisée. En utilisant l'orthogonalité de Q nous avons

$$b = Qd = Q_1d_1 + Q_2d_2 = Ax + r$$

puisque $Q_1d_1 = Q_1Rx = Ax$, il suit que : $r = Q_2d_2$

Cependant, si A est de rang qui n'est pas complet, il s'avère que la factorisation QR a des problèmes de stabilité. Dans ce cas, on est conduit vers la méthode de décomposition en valeurs singulières (SVD) [6, 8].

Définition 3.2 (SVD) Soit la matrice $A \in \mathbb{C}^{n \times d}$ tel que $\text{rang}(A) = r$, la décomposition en valeurs singulières (SVD) de la matrice A est donnée par :

$$A = U\Sigma V^* \quad (\text{I.18})$$

Où, $U \in \mathbb{R}^{n \times n}$ et $V \in \mathbb{R}^{d \times d}$ des matrices orthogonales qui contiennent les vecteurs singuliers de la matrice A , V^* est le conjugué transposé de V . $\Sigma \in \mathbb{R}^{n \times d}$ une matrice diagonale tel que la diagonale est composée des valeurs singulières de la matrice A en ordre décroissant tel que leurs nombre est égale au rang de la matrice A .

Théorème 3.2 Considérons le problème général des moindres carrés

$A \in \mathbb{C}^{n \times d}$, $\text{rang}(A) = r \leq \min(n, d)$, La solution de ce problème est unique, elle s'écrit en terme de SVD sous la forme

$$x = V \begin{pmatrix} \Sigma_r^{-1} & 0 \\ 0 & 0 \end{pmatrix} U^H b \quad (\text{I.19})$$

Preuve 3.2 Soit

$$z = V^H x = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}$$

et

$$c = U^H b = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$$

Où $z_1, c_1 \in \mathbb{C}^r$ alors

$$\begin{aligned} \|b - Ax\|_2 &= \left\| U^H (b - AVV^H x) \right\|_2 \\ &= \left\| \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} - \begin{pmatrix} \Sigma_1^{-1} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \right\|_2 \\ &= \left\| \begin{pmatrix} c_1 - \Sigma_1^{-1} z_1 \\ c_2 \end{pmatrix} \right\|_2 \end{aligned}$$

Ainsi, la norme résiduelle est minimisée pour z_2 arbitraire et $z_1 = \Sigma_r^{-1} c_1$. Le choix de $z_2 = 0$ minimise $\|z\|_2$, et donc $\|x\|_2 = \|Vz\|_2$

L'approche SVD est à la fois précise et robuste dans le cas où le rang de la matrice est incomplet. L'approche QR est moins coûteuse que la SVD, mais elle est légèrement moins robuste dans le cas où le rang de la matrice est incomplet. L'étude des méthodes disponibles

est donnée dans [6, 25].

On note quelques inégalités impliquant les valeurs singulières :

Soient $A, B \in \mathbb{R}^{m \times n}$, pour tout $i = 1, \dots, \min\{m, n\}$,

$$|\sigma_i(A) - \sigma_i(B)| \leq \|A - B\|_2$$

Si $A \in \mathbb{R}^{p \times m}$ et $B \in \mathbb{R}^{m \times n}$ pour tout $i = 1, \dots, \min\{m, n\}$

$$\sigma_i(AB) \leq \sigma_i(A) \sigma_i(B)$$

Parmi les inconvénients des méthodes directes de résolution des problèmes des moindres carrés, il y a le temps requis pour la résolution du problème, en particulier lors de l'utilisation de grandes matrices. La randomisation est une technique qui permet d'obtenir des solutions beaucoup plus rapides que par les méthodes déterministes. L'approche de randomisation peut être appliquée à des matrices creuses.

Dans ce chapitre, nous donnons un aperçu sur les notions de base d'algèbre linéaire qui sont très utiles pour la suite de notre travail.

Remarque 3.1 (*SVD et les moindres carrés*) *Si la décomposition en valeurs singulières (SVD) $A = U \Sigma V^t \in \mathbb{R}^{n \times d}$ est connue, alors la norme minimum du théorème 3.2 est la solution des moindres carrés du $\|Ax - b\|_2$ est donnée par*

$$x = A^\dagger b = V \begin{pmatrix} \Sigma_r^{-1} & 0 \\ 0 & 0 \end{pmatrix} c, \quad c = U^t b \quad (\text{I.20})$$

Où le rang(A) = $r \leq d$ et $\Sigma_r = \text{diag}(\sigma_1, \dots, \sigma_r)$, V et $U^t b$ sont requis.

4 Bibliothèques numériques

L'une des premières collections de logiciels de haute qualité a été une série d'algorithmes écrit en Algol 60 paru en 1971. Il comprenait onze sous-programmes pour les systèmes linéaires, les moindres carrés linéaires et la programmation linéaire ainsi que dix-huit routines pour le calcul des valeurs propres algébrique.

En 1974, EISPACK est apparu comme une collection de sous-routines écrites en language Fortran IV contenant le calcul des valeurs propres et les vecteurs propres des matrices. EISPACK était basé principalement sur les procédures d'Algol.

LINPACK, réalisé en 1979, est une collection de sous-routines écrites en Fortran pour résoudre les systèmes d'équations linéaires ainsi que le problème des moindres carrés linéaire. L'une des caractéristiques les plus importantes de LINPACK est que les sous-routines effectuent autant de calculs que possible par des appels aux sous-programmes d'algèbre linéaire de base (BLAS).

BLAS, connu sous le nom de niveau 1 BLAS, a été introduit en 1979 par Lawson. BLAS est un ensemble de fonctions standardisées réalisant des opérations de base d'algèbre linéaire comme des additions de vecteurs, des produits scalaires ou des multiplications de matrices. Le niveau 2 BLAS sont des opérations impliquant une matrice et un ou plusieurs vecteurs, alors que le niveau 3 de BLAS, introduit en 1990, a été dérivé du niveau 2 BLAS en remplaçant les vecteurs par des matrices. Ils permettent d'atteindre des performances proches de l'optimum sur une grande variété d'architectures.

La routine de multiplication matrice-matrice (GEMM) est le noyau du niveau 3 de BLAS qui se rapproche le plus des performances de pointe.

La collection LAPACK de sous-programmes a été publiée en 1992 et est conçue pour remplacer et intégrer les algorithmes dans LINPACK et EISPACK. LAPACK est une bibliothèque écrite en fortran dédié à l'algèbre linéaire numérique, elle est développé par l'université du Tennessee.

Un certain nombre d'algorithmiques avancées qui ont été faites après LINPACK et EISPACK ont été écrits et incorporés. Les sous-programmes sont restructurés pour atteindre une efficacité beaucoup plus grande sur les ordinateurs haute performance modernes.

La bibliothèque LAPACK a énormément simplifié les calculs matriciels. Plusieurs formes spéciales de matrices sont supportées par LAPACK.

ScaLAPACK est une extension de la bibliothèque logicielle LAPACK conçu pour être efficace. Les sous programmes de ScaLAPACK ressemble beaucoup à ceux de LAPACK.

LAPACK95 est une interface Fortran 95 de la bibliothèque LAPACK écrite en Fortran 77. Il améliore l'interface utilisateur originale du paquet LAPACK, en profitant des simplifications considérables que Fortran-95 permet.

Définition 4.1 (*L'inégalité de Schwarz*)

Soit X et Y deux variables aléatoires réelles définies sur (Ω, P) .

alors $|\text{Cov}(X, Y)| \leq \sigma(X)\sigma(Y)$ si et seulement s'il existe $\lambda \in \mathbb{R}$ et $c \in \mathbb{R}$ tel que $Y = \lambda X + c$.

4.1 Produit de Kronecker

Définition 4.2 Soient $A \in \mathbb{R}^{n \times d}$ et $B \in \mathbb{R}^{p \times q}$, le produit $A \otimes B$ nomé produit de kronecker est définit par la matrice $np \times dq$ tel que

$$A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & \dots & a_{1d}B \\ a_{21}B & a_{22}B & \dots & a_{2d}B \\ \vdots & \vdots & \dots & \vdots \\ a_{n1}B & a_{n2}B & \dots & a_{nd}B \end{pmatrix}$$

La résolution des problèmes tout en prenant en considération le produit de Kronecker peuvent économiser le stockage mémoire ainsi que minimiser le nombre d'opérations.

On note quelques propriétés du produit de Kronecker

$$\begin{aligned} (A + B) \otimes C &= (A \otimes C) + (B \otimes C) \\ A \otimes (B + C) &= (A \otimes B) + (A \otimes C) \\ A \otimes (B \otimes C) &= (A \otimes B) \otimes C \\ (A \otimes B)^t &= A^t \otimes B^t \end{aligned}$$

Si A, B, C, D sont des matrices tel que les produits AC et BD sont définis alors

$$(A \otimes B)(C \otimes D) = AC \otimes BD$$

Si les matrices A et B sont carrés et non singulières alors

$$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$$

Soient A^\dagger et B^\dagger les pseudo-inverse des matrices A et B respectivement, nous avons

$$(A \otimes B)^\dagger = A^\dagger \otimes B^\dagger$$

On va introduire une fonction reliée au produit de Kronecker qui convertit une matrice à un

vecteur.

Définition 4.3 Pour toute matrice $A = (a_1, a_2, \dots, a_d) \in \mathbb{R}^{n \times d}$

Vec est un vecteur formé en empilant les colonnes de $A = (a_1, a_2, \dots, a_d) \in \mathbb{R}^{n \times d}$ en un vecteur.

$$vec(A) = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_d \end{pmatrix}$$

Remarque 4.1 Si $A \in \mathbb{R}^{n \times d}$, $B \in \mathbb{R}^{p \times q}$ et $X \in \mathbb{R}^{q \times d}$ alors

$$(A \otimes B) Vec(X) = Vec(BXA^t)$$

En effet,

$$(BXA^t)_k = BX(A^t)_k = B \sum_{i=1}^n a_{ki} X_i$$

5 Résolution des problèmes des moindres carrés

5.1 Sketching matrice

Définition 5.1 (43) Soit la matrice $A \in \mathbb{R}^{n \times d}$, $S \in \mathbb{R}^{d \times s}$ une sketching matrice. On dit que $C \in \mathbb{R}^{m \times s}$ est un sketch de A s'il vérifie : $C = AS$.

Le sketch d'une matrice est linéaire, la taille de C est petite par rapport à A , mais A préserve plusieurs propriétés importantes.

Considérons le problème des moindres carrés. Pour toute $n \times d$ matrice A et $n \times 1$ vecteur b , le système est sur-déterminé puisque $n \gg d$, on cherche la solution $x \in \mathbb{R}^d$ qui minimise $\|Ax - b\|_2$.

Définition 5.2 (Sous espace d'injection) Soit la matrice $A \in \mathbb{R}^{n \times d}$, $p \in [1, \infty]$ et $\varepsilon \in (0, 1)$, $\phi \in \mathbb{R}^{s \times n}$ est un sous espace d'injection de Ω_p , si $s = O(\text{poly}(d))$ indépendent de n , et

$$(1 - \varepsilon) \|y\|_p \leq \|\phi y\|_p \leq (1 + \varepsilon) \|y\|_p, \forall y \in \Omega_p \quad (\text{I.21})$$

Notons que le sous espace d'injection Ω dans \mathbb{R}^s avec $s > 0$ est une structure préservant $\phi : \Omega \rightarrow \mathbb{R}^s$, où la signification de la préservation de la structure varie en fonction de

l'application. Nous nous intéressons aux prolongement linéaires de l'espace vectoriel normé $\Omega_p = (\Omega, \|\cdot\|_p)$.

Basé sur les propriétés du sous espace d'injection, nous pouvons définir l_2 sous espace d'injection qui est utilisé dans l'algèbre linéaire aléatoire et a été fourni par Woodruff [42]. En utilisant l_2 sous espace d'injection, on peut travailler avec SA au lieu de $A \in \mathbb{R}^{n \times d}$, $s \ll n$, donc nous travaillons avec une matrice plus petite qui réduit le temps et la complexité du problème.

Soit la matrice $A \in \mathbb{R}^{n \times d}$, $d \ll n$. Toute sketching matrice S peut être utilisée pour résoudre le problème approximativement tant qu'il satisfait la propriété du sous-espace d'injection $C = S^t A$ est une sketch de la matrice A . La taille de C est plus petite que A .

Définition 5.3 $(1 \pm \varepsilon)$ l_2 sous espace d'injection des colonnes de la matrice A est la matrice S tel que :

$$\|SAx\|_2^2 = (1 \pm \varepsilon) \|Ax\|_2^2$$

Définition 5.4 Soit $S \in \mathbb{R}^{s \times n}$ une matrice qui satisfait les propriétés du sous espace d'injection, on considère le problème :

$$\tilde{x} = \min_x \|S^t Ax - S^t b\|_2 \quad (\text{I.22})$$

alors

$$\|S^t Ax - S^t b\|_2 \leq (1 + \varepsilon) \min_x \|Ax - b\|_2 \quad (\text{I.23})$$

L'objectif de cette théorie est de trouver la solution x qui minimise $\|Ax - b\|_2$, mais avec les méthodes aléatoires, nous pouvons montrer que :

Lemme 5.1 le vecteur \tilde{x} est la meilleure solution avec une forte probabilité tel que

$$\|Ax - b\|_2 \leq (1 + \varepsilon) \min_x \|Ax - b\|_2 \quad (\text{I.24})$$

Preuve 5.1 Effectivement, à partir de la définition du sous espace d'injection nous avons

$$\frac{1}{\eta} \|Ax - b\|_2 \leq \|S^t(Ax - b)\|_2 \leq \eta \|Ax - b\|_2, \quad \forall x.$$

Ainsi

$$\frac{1}{\eta} \|A\tilde{x} - b\|_2 \leq \|S^t(A\tilde{x} - b)\|_2 \quad \text{et} \quad \|S^t(Ax - b)\|_2 \leq \eta \|Ax^* - b\|_2;$$

Puisque \tilde{x} est optimal alors,

$$\|S^t A \tilde{x} - S^t b\|_2 \leq \|S^t(Ax - b)\|_2,$$

et

$$\begin{aligned} \frac{1}{\eta} \|A \tilde{x} - b\|_2 &\leq \|S^t A \tilde{x} - S^t b\|_2 \leq \|S^t A x^* - S^t b\|_2 \leq \eta \|A x^* - b\|_2 \\ &\Rightarrow \|A \tilde{x} - b\|_2 \leq \eta^2 \|A x^* - b\|_2. \end{aligned}$$

Cet approche s'effectue à travers deux étapes suivantes :

1. Construire une matrice S préservant le sous-espace d'injection
2. Résoudre le problème des moindres carrés de taille réduite

$$\tilde{x} = \min_x \|S^t A x - S^t b\|_2 \quad (\text{I.25})$$

qui peut être résolu dans $O(sd^2)$ temps afin d'obtenir une solution approximative du problème original.

Lemme 5.2 (30) $A \in \mathbb{R}^{n \times d}$ ($A = USV^t$), $n \gg d$ et $\text{rang}(A) = d$ alors :

$$\begin{aligned} \text{rang}(SA) &= \text{rang}(SU) = \text{rang}(A) = \text{rang}(U) = d \\ \|\Sigma_{SU} - \Sigma_{SU}^{-1}\| &\leq \varepsilon \\ (SA)^\dagger &= V \Sigma^{-1} (SU)^\dagger \\ \|(SU)^\dagger - (SU)^t\|_2 &\leq \|\Sigma_{SU} - \Sigma_{SU}^{-1}\|_2 \end{aligned}$$

Preuve 5.2 Nous commençons par montrer la première formule du lemme.

En effet, nous avons :

$$\begin{aligned} |1 - \sigma_i^2(SU)| &= |\sigma(U^t U) - \sigma(U^t S^t S U)| \\ &\leq \|U^t U - U^t S^t S U\|_2 \quad (\text{puisque } U^t U = I, \text{ et utilisons les propriétés de JL}) \\ &\leq \varepsilon \end{aligned}$$

Donc toutes les valeurs singulières de SU , sont proches de 1. D'où $\text{rang}(SA)$ et le $\text{rang}(SU)$

sont conservés tel que :

$$\begin{aligned}
 \left\| \Sigma_{SU} - \Sigma_{SU}^{-1} \right\| &= \max_{ij} \left| \sigma_i(SU) - \frac{1}{\sigma_i(SU)} \right| \\
 &\leq \max_i \left| \sigma_i(SU) - \frac{1}{\sigma_i(SU)} \right| \\
 &= \max_i \frac{|1 - \sigma_i(SU)^2|}{|\sigma_i(SU)|} \\
 &\leq \frac{\varepsilon}{\sqrt{1-\varepsilon}} \quad (\text{puisque } |1 - \sigma_i^2(SU)| \leq \varepsilon)
 \end{aligned}$$

Elle dépend de la valeur de ε , la borne est proche de ε .

Montrons la troisième équation,

$$\begin{aligned}
 (SA)^\dagger &= (SU \sum V^t)^\dagger \\
 &= V (SU \sum)^\dagger ; \quad (V \text{ unitaire}) \\
 &= V (U_{SU} \sum_{SU} V_{SU}^\dagger \sum)^\dagger \\
 &= V (\sum_{SU} V_{SU}^t \sum)^+ U_{SU}^t ; \quad (U_{SU}^t U_{SU} = I) \\
 &= V \sum^{-1} V_{SU} \sum_{SU}^{-1} U_{SU}^t ; \quad (\text{Puisque } \sum, \sum_{SU}, V_{SU} \text{ de rang complet}) \\
 &= V \sum^{-1} (SU)^\dagger ; \quad (V_{SU} \sum_{SU}^{-1} U_{SU}^t = (SU)^\dagger)
 \end{aligned}$$

Montrons la quatrième inégalité

$$\begin{aligned}
 \left\| (SU)^+ - (SU)^t \right\|_2 &= \left\| (U_{SU} \sum_{SU} V_{SU}^t)^+ - (U_{SU} \sum_{SU} V_{SU}^t)^t \right\|_2 \\
 &= \left\| V_{SU} (\sum_{SU}^{-1} - \sum_{SU}) U_{SU}^t \right\|_2 \\
 &\leq \left\| \sum_{SU}^{-1} - \sum_{SU} \right\|_2
 \end{aligned}$$

5.2 Approximation des moindres carrés

La théorie d'utilisation des matrices aléatoires est divisée en deux parties : matrices d'échantillonnage et de projection aléatoire.

Soit X une matrice sketch, par exemple une matrice d'échantillonnage ou une matrice de projection aléatoire. On va résoudre le problème

$$x_{opt} = \min_x \| X(Ax - b) \|_2 \quad (\text{I.26})$$

En dérivant le carré de la fonction par rapport à x et en mettant la dérivée à zéro, on obtient l'équation suivante :

$$A^t X^t X A x - A^t X^t X b = 0.$$

Alors

$$x_{opt} = (A^t X^t X A)^{-1} A^t X^t X b. \quad (I.27)$$

Étant donné que AX est de rang complet, par conséquent,

$$Ax_{opt} = A(A^t X^t X A)^{-1} A^t X^t X b$$

Il existe essentiellement deux mesures de qualité pour la solution issue des méthodes aléatoires :

- $\tilde{x}_{opt} \approx x_{opt}$
- $\| A\tilde{x}_{opt} - b \| \approx \| Ax_{opt} - b \|$

Le lemme suivant fournit les conditions pour que \tilde{x}_{opt} fonctionne bien par rapport à x_{opt} .

Lemme 5.3 *Considérons le problème $Z = \min_x \| Ax - b \|_2$ avec $x_{opt} = A^+b$, soit $A = U_A \Sigma_A V_A^T$ et $b^\perp = (I - U_A U_A^T)b$. Pour tout $\varepsilon \geq 0$, si les deux conditions suivantes sont satisfaites*

$$\begin{aligned} \text{ConditionI} &: \sigma_{\min}^2(XU_A) \geq \frac{1}{\sqrt{2}} \\ \text{ConditionII} &: \| U_A^t X^t X b^\perp \|_2^2 \leq (\frac{\varepsilon}{2})Z^2 \end{aligned} \quad (I.28)$$

alors la solution au problème $\tilde{x}_{opt} = (XA)^\dagger Xb$, satisfait

$$\| A\tilde{x}_{opt} - b \| \leq (1 + \varepsilon)Z$$

Preuve 5.3 A partir de l'inégalité triangulaire

$$\begin{aligned} \| A\tilde{x}_{opt} - b \|_2 &= \| A\tilde{x}_{opt} - b + Ax_{opt} - Ax_{opt} \|_2 \\ &\leq \| Ax_{opt} - b \|_2 + \| A\tilde{x}_{opt} - Ax_{opt} \|_2 \end{aligned}$$

Du moment que : $\| Ax_{opt} - b \|_2 = Z$, il suffit de montrer que le deuxième terme est borné par εZ .

Nous avons :

$$Ax_{opt} = U_A U_A^t b \quad \text{et} \quad Ax_{opt} = U_A (U_A^t X^t X U_A)^{-1} U_A^t X^t X b$$

Notons que XU_A est de rang complet.

soit $U_A y_{opt} = A(\tilde{x}_{opt} - x_{opt})$, nous avons :

$$\| U_A y_{opt} \|_2^2 = y_{opt}^t U_A^t U_A y_{opt} = \| y_{opt} \|_2^2$$

En outre, notons que

$$U_A^t X^t X U_A y_{opt} = U_A^t X^t X (A \tilde{x}_{opt} - A x_{opt}) = -U_A^t X^t X b^\perp$$

alors

$$\| U_A^t X^t X b^\perp \|_2 \geq \| y_{opt} \|_2 \geq \frac{1}{2} \| A(\tilde{x}_{opt} - x_{opt}) \|_2 \quad (\text{I.29})$$

par conséquent

$$\| A \tilde{x}_{opt} - A x_{opt} \|_2 \leq 2 \left(\frac{\varepsilon}{2} \right) Z = \varepsilon Z$$

Définition 5.5 Johnson- Lindenstrauss Lemma [25] Soit $\varepsilon \in (0, 1)$, un ensemble de points χ de n points dans \mathbb{R}^d , il existe une fonction linéaire $\phi : \mathbb{R}^n \rightarrow \mathbb{R}^s$ avec $s = c \log \frac{n}{\varepsilon^2}$, où $c > 0$ une constante globale, tel que

$$(1 - \varepsilon) \| x - y \|_2^2 \leq \| \phi(x) - \phi(y) \|_2^2 \leq (1 + \varepsilon) \| x - y \|_2^2, \forall x, y \in \chi \quad (\text{I.30})$$

On dit que ϕ est Johnson- Lindenstrauss si elle satisfait la condition précédente avec une constante probabilité.

La matrice SA est connue par le sketch matrice et S le sketching matrice de A qui doit être linéaire. Cette théorie nous donne la solution plus rapidement que les méthodes directes. Le problème qui se pose est que si A est dense, le produit matriciel va être lent. Cependant Sarlos [38] a démontré qu'on peut prendre S de la famille des matrices aléatoires. Ceci est structuré dans la transformation de Johnson- Lindenstrauss rapide.

Chapitre II

Matrices aléatoires

1 Introduction

Les méthodes classiques ou nommée directes comme exemples, la décomposition de Cholesky, décomposition en QR et la décomposition en valeurs singulières calculent la solution du problème des moindres carrés dans $O(mn^2)$ temps. Les méthodes aléatoires résolvent ce problème en utilisant des matrices aléatoires. Cette théorie est dévisée en deux parties : matrices d'échantillonnage aléatoire, elle consiste en un petit nombre de lignes soigneusement échantillonnées de A (et les éléments correspondants de b). Tandis que l'utilisation des matrices de projection aléatoire consiste en un petit nombre de combinaison linéaire des lignes de A et des éléments de b .

2 Préliminaires

Les matrices aléatoires sont un outil utile et important en algèbre linéaire. Soit la matrice $A \in \mathbb{R}^{n \times d}$, $d \ll n$, tout sketching matrice S satisfait les propriétés du sous espace d'injection [42, 43] avec C le sketch matrice de A , $C = S^t A \in \mathbb{R}^{s \times d}$ alors la taille de C est petite que A . Nous pouvons trouver dans [43] des détails de la méthode de sketching.

La transformation de Johnson Lindenstrauss (JLT) [25] joue un rôle dans le développement du sous espace d'injection basé sur les matrices aléatoires, elle est définie comme suit : (II.1).

Définition 2.1 (25) (Transformée de Johnson Lindenstrauss (JLT)) Soit donnée $\varepsilon > 0$ et n points $\{x_i\}_{i=1}^n$, $x_i \in \mathbb{R}^d$, ε -JLT noté $\pi \in \mathbb{R}^{r \times d}$ est une projection des points dans \mathbb{R}^r tel que

$$(1 - \varepsilon) \|x_i\|_2^2 \leq \|\pi(x_i)\|_2^2 \leq (1 + \varepsilon) \|x_i\|_2^2 \quad (\text{II.1})$$

Afin de donner un aperçu des matrices de projection aléatoire, nous nous intéressons au calcul rapide de (JLT).

Définition 2.2 (25) (Fast subspace Johnson Lindenstrauss Transform (FJLT)) Soit $\varepsilon > 0$, $U \in \mathbb{R}^{n \times d}$ matrice orthogonale de d vecteurs dans \mathbb{R}^n . FJLT projectent les vecteurs de $\mathbb{R}^n \rightarrow \mathbb{R}^r$ tel que l'orthogonalité de U est préservée, et elle est rapide.

Alors $\pi \in \mathbb{R}^{r \times n}$ est ε -FJLT si :

- $\|I_d - U^t \pi^t \pi U\|_2 \leq \varepsilon$
- $\forall x \in \mathbb{R}^{n \times d}$, on peut calculer πx dans $O(nd\log(r))$ temps.

L'objectif de cette théorie est de trouver une solution x qui minimise $\|Ax - b\|_2$. Cependant, nous pouvons trouver un vecteur de minimisation \tilde{x} en haute probabilité avec la théorie de randomisation.

$$\|S^t A \tilde{x} - S^t b\|_2 \leq (1 + \varepsilon) \min_x \|Ax - b\|_2 \quad (\text{II.2})$$

Généralement, il existe deux types de matrices aléatoires : matrices de projection et les matrices d'échantillonnage, nous les présentons ci dessous :

La projection aléatoire est une méthode utilisée dans l'approximation des moindres carrés. Le principe de la projection aléatoire est qu'on projette des points de données d'un espace de haute dimension \mathbb{R}^n sur un sous-espace de dimension inférieure \mathbb{R}^d ($d \ll n$) en utilisant une matrice aléatoire S .

Les méthodes de projection aléatoire sont efficaces sur le plan des calculs et suffisamment précises en pratique pour réduire la dimension des ensembles de données de grande dimension. L'application de la projection aléatoire est une tâche courante dans de nombreuses applications d'exploration des données. D'autre part, les méthodes d'échantillonnage des colonnes ont besoin d'accéder aux données pour l'approximation.

3 Produit matriciel aléatoire

Le produit matriciel aléatoire est un problème d'algèbre linéaire numérique. L'approche est basée sur l'échantillonnage afin d'approximer le produit de deux matrices $A \in \mathbb{R}^{m \times n}$

et $B \in \mathbb{R}^{n \times p}$. Cet algorithme de multiplication matricielle aléatoire est utilisé de manière essentielle dans l'analyse de l'approximation des moindres carrés, ainsi que les algorithmes d'approximation de rang inférieurs.

Le produit matriciel AB s'écrit sous forme d'une somme comme :

$$AB = \sum_{t=1}^n A_{*t} B_{t*} \quad (\text{II.3})$$

Nous exprimons un algorithme aléatoire simple pour approximer le produit AB .

Considérons le produit matriciel défini par (II.3), soit la probabilité $\{p_k\}_{k=1}^n$, et un nombre c de colonnes et lignes à choisir, alors on peut estimer le produit AB par [9] :

$$\sum_{t=1}^c \frac{1}{cp_{i_t}} A_{*i_t} B_{*i_t} \quad (\text{II.4})$$

Autrement dit, l'estimateur (II.4) peut être considéré comme le produit des deux matrices C et R tel que

$$CR = \sum_{t=1}^c C_{*t} R_{t*} = \sum_{t=1}^c \left(\sqrt{\frac{1}{cp_{i_t}}} A_{*i_t} \right) \left(\sqrt{\frac{1}{cp_{i_t}}} B_{*i_t} \right) = \frac{1}{c} \sum_{t=1}^c \frac{1}{p_{i_t}} A_{*i_t} B_{*i_t} \quad (\text{II.5})$$

Par conséquent, l'approximation d'un produit matriciel utilise l'échantillonnage et la mise à l'échelle dans l'algorithme de multiplication. Définissons $S \in \mathbb{R}^{n \times c}$ une matrice rééchelonné d'échantillonnage, elle s'écrit

$$S_{i_t} = \frac{1}{\sqrt{cp_{i_t}}}$$

Alors on peut écrire $C = AS$ et $R = S^t B$ tel que $CR = ASS^t B \approx AB$

Lemme 3.1 *Soient C et R deux matrices définies précédemment alors*

$$\begin{aligned} E[(CR)_{ij}] &= (AB)_{ij} \\ \text{Var}[(CR)_{ij}] &\leq \frac{1}{c} \sum_{k=1}^n \frac{A_{ik}^2 B_{kj}^2}{p_k} \end{aligned} \quad (\text{II.6})$$

Preuve 3.1 *Soient i, j fixé. Pour tout $t = 1, \dots, c$, on définit*

$$X_t = \left(\frac{A_{*i_t} B_{i_t*}}{cp_{i_t}} \right)_{ij} = \frac{A_{iit} B_{i_tj}}{cp_{i_t}}, \quad \forall t$$

$$E[X_t] = \sum_{k=1}^n p_k \frac{A_{ik}B_{kj}}{cp_k} = \frac{1}{c} \sum_{k=1}^n A_{ik}B_{kj} = \frac{1}{c} (AB)_{ij}$$

Puisque $(CR)_{ij} = \sum_{t=1}^c X_t$ alors

$$E[(CR)_{ij}] = E\left[\sum_{t=1}^c X_t\right] = \sum_{t=1}^c E[X_t] = (AB)_{ij}$$

et

$$Var[(CR)_{ij}] = Var\left[\sum_{t=1}^c X_t\right] = \sum_{t=1}^c Var[X_t]$$

Utilisons

$$Var[(CR)_{ij}] \leq E[X_t^2] = \sum_{k=1}^n \frac{A_{ik}^2 B_{kj}^2}{c^2 p_k}$$

on obtient

$$Var[(CR)_{ij}] = \sum_{t=1}^c Var[X_t] \leq c \sum_{k=1}^n \frac{A_{ik}^2 B_{kj}^2}{c^2 p_k} = \frac{1}{c} \sum_{k=1}^n \frac{A_{ik}^2 B_{kj}^2}{p_k}$$

Dans cette partie nous nous intéressons aux bornes supérieure de la matrice d'erreur $\|AB - CR\|_F^2$ où CR est une approximation du produit matriciel AB défini précédemment

Théorème 3.1 (30) Soient deux matrices C et R telles que CR est une approximation de AB alors

$$E[\|AB - CR\|_F^2] \leq \sum_{k=1}^n \frac{\|A_{*k}\|_2^2 \|B_{k*}\|_2^2}{cp_k} \quad (\text{II.7})$$

En outre, si

$$p_k = \frac{\|A_{*k}\|_2 \|B_{k*}\|_2}{\sum_{k'=1}^n \|A_{*k'}\|_2 \|B_{k'*}\|_2} \quad (\text{II.8})$$

pour tout $k = 1, \dots, n$, alors

$$E[\|AB - CR\|_F^2] \leq \frac{1}{c} \left(\sum_{k=1}^n \|A_{*k}\|_2 \|B_{k*}\|_2 \right)^2 \quad (\text{II.9})$$

Le choix de $\{p_k\}_{k=1}^n$ minimise $E[\|AB - CR\|_F^2]$

Preuve 3.2 Rappelons que le produit CR est une estimation de AB , $E[(AB - CR)_{ij}] = 0$ alors

$$E[\|AB - CR\|_F^2] = \sum_{i=1}^m \sum_{j=1}^p E[(AB - CR)_{ij}^2] = \sum_{i=1}^m \sum_{j=1}^p Var[(CR)_{ij}^2]$$

utilisons le lemme 3.1 , on obtient

$$\begin{aligned} E \left[\|AB - CR\|_F^2 \right] &\leq \frac{1}{c} \sum_{k=1}^n \sum_{j=1}^p \frac{1}{p_k} \left(\sum_i A_{ik}^2 \right) \left(\sum_j B_{kj}^2 \right) \\ &= \frac{1}{c} \sum_{k=1}^n \frac{1}{p_k} \|A_{*k}\|_2^2 \|B_{k*}\|_2^2 \end{aligned}$$

Remplaçant p_k par sa valeur donné dans l'équation (II.8) alors,

$$E \left[\|AB - CR\|_F^2 \right] \leq \frac{1}{c} \left(\sum_{k=1}^n \|A_{*k}\|_2 \|B_{k*}\|_2 \right)^2$$

Maintenant on veut trouver $\{p_k\}_{k=1}^n$ qui minimise $E \left[\|AB - CR\|_F^2 \right]$. Définissons la fonction tel que $\sum_{k=1}^n P_k = 1$ alors

$$f(p_1, \dots, p_n) = \sum_{k=1}^n \frac{1}{p_k} \|A_{*k}\|_2^2 \|B_{k*}\|_2$$

On introduit le multiplicateur de Lagrange λ et on définit la fonction suivante :

$$g(p_1, \dots, p_n) = f(p_1, \dots, p_n) + \lambda \left(\sum_{k=1}^n P_k - 1 \right)$$

Dérivons la fonction g telle que

$$0 = \frac{\partial g}{\partial p_k} = \frac{-1}{p_k^2} \|A_{*k}\|_2^2 \|B_{k*}\|_2 + \lambda$$

Ainsi

$$p_k = \frac{\|A_{*k}\|_2 \|B_{k*}\|_2}{\sqrt{\lambda}} \quad (\text{II.10})$$

Par la suite, on va résoudre $\sum_{k=1}^n p_k = 1$ pour avoir la valeur de $\sqrt{\lambda}$ et on la remplace dans (II.10) on obtient

$$p_k = \frac{\|A_{*k}\|_2 \|B_{k*}\|_2}{\sum_{k'=1}^n \|A_{*k'}\|_2 \|B_{k'*}\|_2}$$

Théorème 3.2 Soit $\{p_k\}_{k=1}^n$ une probabilité presque optimale pour tout $\beta \in (0, 1]$,

$$p_k \geq \frac{\beta \|A^{(k)}\|_2 \|B_{(k)}\|_2}{\sum_{k'} \|A^{(k')}\|_2 \|B_{(k')}\|_2}$$

alors

$$E \left[\|AB - CR\|_F^2 \right] \leq \frac{1}{\beta c} \|A\|_F^2 \|B\|_F^2$$

Preuve 3.3 Nous avons

$$\begin{aligned} E \left[\|AB - CR\|_F^2 \right] &= \sum_{i=1}^m \sum_{j=1}^p E \left[(AB - CR)_{ij}^2 \right] \\ &= \sum_i \sum_j \text{Var} \left[(CR)_{ij} \right] \\ &= \frac{1}{c} \sum_{k=1}^n \frac{1}{p_k} \sum_{i=1}^m \sum_{k=1}^n A_{ik}^2 \sum_{k=1}^n B_{kj}^2 - \frac{1}{c} \sum_{i,j} (AB)_{ij}^2 \\ &= \frac{1}{c} \sum_{k=1}^n \frac{1}{p_k} \|A^{(k)}\|^2 \|B_{(k)}\|^2 - \frac{1}{c} \|AB\|_F^2 \\ &\leq \frac{1}{\beta c} \sum_{k=1}^n \frac{1}{p_k} \|A^{(k)}\| \|B_{(k)}\| \end{aligned}$$

En utilisant l'inégalité de Cauchy Schwartz on aura

$$E \left[\|AB - CR\|_F^2 \right] \leq \frac{1}{\beta c} \|A\|_F^2 \|B\|_F^2$$

4 Matrices aléatoires

Dans cette partie, on se concentrera sur les matrices d'échantillonnage aléatoires et les matrices de projection aléatoire afin de résoudre le problème linéaire d'approximation des moindres carrés.

Étant donnée une $n \times d$ matrice A , avec $n \gg d$, et un vecteur b , le problème d'approximation des moindres carrés se base sur la recherche du vecteur

$$x_{opt} = \min_x \|Ax - b\|_2$$

Les méthodes classiques, y compris la décomposition de Cholesky, les factorisation QR , et la décomposition en valeurs singulières, calcule la solution en temps $O(nd^2)$.

Les Méthodes aléatoires résolvent ce problème en multipliant la matrice A par une matrice aléatoire. Les matrices d'échantillonnage consistent à travailler avec un petit nombre échantillonées et remis à l'échelle de la matrice A (ainsi que les éléments de b), tandis que pour les méthodes de projection aléatoire utilise un nombre de combinaisons linéaires des lignes de A et des éléments de b . Si on résout le sous système induit, alors les approximations des

erreurs relatives très fines à la solution du problème original sont obtenues.

En plus, on peut utiliser les matrices aléatoires pour calculer un préconditionneur du problème original afin d'obtenir une très haute précision des approximations. Ces implémentations numériques s'exécutent en $O(nd^2)$ temps, elles sont plus rapides que les méthodes numériques déterministes existantes.

On peut représenter l'échantillonnage aléatoire avec une matrice d'échantillonnage aléatoire S , si l'échantillonnage aléatoire est implémenté en choisissant des colonnes c , alors la matrice S de $n \times c$ a des entrées $S_{ij} = \frac{1}{\sqrt{cp_i}}$

Alternativement, étant donné une matrice $A \in \mathbb{R}^{n \times d}$, on peut s'intéresser à une projection aléatoire en multipliant A par une matrice de projection aléatoire $n \times l$, ainsi sélectionnant l combinaisons linéaires des colonnes de A . Il y a plusieurs façons de construire une telle matrice. Johnson et Lindenstrauss [25] considèrent une projection orthogonale sur un espace de dimension aléatoire, où $l = O(\log m)$

4.1 Projection Gaussienne

Soit $S \in \mathbb{R}^{n \times s}$ une matrice de projection Gaussienne aléatoire, elle s'exprime comme

$$S = \frac{1}{\sqrt{s}} G \quad (\text{II.11})$$

où G est échantillonné à partir de $N(0, 1)$. La projection Gaussienne est connue comme étant la transformée de Johnson Lindenstrauss.

Elle est simple à implémenter parmi les inconvénients de cette méthode est la complexité pour effectuer la multiplication matricielle d'autant plus si la matrice A est creuse.

4.2 Transformé d'Hadamard aléatoire

La matrice Randomisé d'Hadamard (HD) [11] a été introduite comme une version rapide du lemme de Johnson Lindenstrauss, elle est définie par :

$$S = HD \quad (\text{II.12})$$

ou $H \in \mathbb{R}^{n \times n}$ est la matrice de Walsh Hadamard telles que : $H = \frac{1}{\sqrt{n}} H_n$, et H_n , H_2 sont définies respectivement par $H_n = \begin{bmatrix} H_{\frac{n}{2}} & H_{\frac{n}{2}} \\ H_{\frac{n}{2}} & -H_{\frac{n}{2}} \end{bmatrix}$, $H_2 = \begin{bmatrix} +1 & +1 \\ +1 & -1 \end{bmatrix}$, $D \in \mathbb{R}^{n \times n}$ est une matrice

diagonale dont les entrées sont fixées à $+1$ avec une probabilité $1/2$ et à -1 avec probabilité $1/2$.

A noté que la transformée d’Hadamard aléatoire est une transformation orthogonale, alors $HDD^tH^t = H^tD^tDH = I_n$, appliquant HD prend $O(nd\log(d))$ temps.

Rappelons que la transformée d’Hadamard aléatoire sous-échantillonnée (SRHT) est une matrice de la forme $S = \frac{1}{\sqrt{s}}RHD$, où $R \in \mathbb{R}^{s \times n}$ restreint un vecteur de dimension n à s coordonnées. L’application de SRHT prend $O(nd\log(s))$ complexité en temps, ce qui rend le SRHT une version rapide de JLT (FJLT). SRHT est plus efficace que la projection Gaussienne

4.3 Transformée de Fourier randomisée sous-échantillonnée

La transformée de Fourier aléatoires sous-échantillonnée (SRFT)[21,37] est définie par :

$$\Omega = \sqrt{\frac{n}{l}} DFS \quad (\text{II.13})$$

où D est une $n \times n$ matrice diagonale dont les entrées sont des variables aléatoires indépendantes et uniformément réparties sur le cercle d’unité complexe, F est la transformée de Fourier discrète unitaire $n \times n$, dont les entrées prennent les valeurs

$$F_{pq} = \frac{1}{\sqrt{n}} e^{-2\pi i(p-1)(q-1)/n}$$

pour $p, q = 1, 2, \dots, n$, et S est une $n \times l$ matrice qui échantillonne les coordonnées l à partir de n uniformément au hasard.

Rokhlin et al. [37] indiquent que l’algorithme utilisé avec la transformée de Fourier randomisée sous-échantillonnée (SRFT) fonctionne mieux que toute autre transformations randomisées.

4.4 Leverage score échantillonné

Leverage Score échantillonné de la matrice A est la norme au carré de la matrice contenant ses vecteurs singuliers. Il sélectionne chaque colonne de A avec une probabilité proportionnelle à sa moyenne[38].

Soit $A \in \mathbb{R}^{n \times s}$ telle qu’elle est démontré dans [39], pour toute constante $\varepsilon > 0$, et avec une probabilité d’au moins $\frac{2}{3}$, les vecteurs u'_i sont définis comme u_i pour tout $i \in [n]$: $u'_i = (1 \pm \varepsilon) u_i$. Le temps d’application de cette méthode est $O(nnz(A) \log n + r^3 \log^2 r + r^2 \log n)$,

où r est le rang de A . Le calcul du Leverage Score échantillonné est aussi cher que le calcul de la SVD. Dans [43] nous pouvons trouver une étude de cette technique et une expérience numérique.

Cette technique a été utilisé historiquement pour la détection des valeurs aberrantes dans les diagnostiques des régressions statistique, ils ont également été cruciaux dans le récent développement d'algorithmes matriciels randomisés.

Définition 4.1 *La cohérence γ des lignes de la matrice A est donnée par*

$$\gamma = \max_{i \in \{1, \dots, m\}} l_i$$

i.e., c'est le plus grand leverage score statistique de A .

4.5 Count-sketch

Les propriétés théoriques du count-sketch sont établies par Woodruff [42], qui l'a appliqué pour accélérer le calcul matriciel. Il existe différentes manières d'implémenter count-sketch : map-reduce fashion et streaming fashion. Il peut être écrit comme le produit de deux matrices

$$F = SD \in \mathbb{R}^{s \times n} \quad (\text{II.14})$$

Où $S \in \mathbb{R}^{s \times n}$ à chaque colonne choisie indépendamment et uniformément de \mathbb{R}^s et $D \in \mathbb{R}^{n \times n}$ est une matrice diagonale avec des entrées diagonales choisies indépendamment et uniformément par ± 1 .

La méthode count-sketch est très efficace, en particulier lorsque A est creuse et comparée à la projection gaussienne, elle nécessite un plus grand s pour atteindre la même précision. Le calcul du count-sketch d'une matrice est $O(nnz(A) + \text{poly}(\frac{d}{\varepsilon}))$ temps et de $O(\text{poly}(\frac{d}{\varepsilon}))$ de mémoire.

Remarque 4.1 *Si S est une projection Gaussienne, SRHT, count-sketch, où leverage score échantillonné matrice, et $s = \text{poly}(\frac{d}{\varepsilon})$ pour tout paramètre d'erreur $\varepsilon \in]0, 1]$, alors*

$$\| A\tilde{x} - b \|_2^2 \leq (1 + \varepsilon)^2 \min_x \| Ax - b \|_2^2 \quad (\text{II.15})$$

est garanti [39]

4.6 Matrices de code sous-échantillonnées

La matrice de code sous-échantillonnées est définie par :

$$\Omega = \sqrt{\frac{2^r}{l}} D S \phi \quad (\text{II.16})$$

ou D est une matrice aléatoire diagonale $n \times n$, ces entrées sont indépendantes de signes aléatoires, i.e. variable aléatoire uniformément répartie sur $\{\pm 1\}$, S est l'échantillonneur vers le bas uniformément aléatoire, un $n \times 2^r$ matrice telle que les n lignes sont sélectionnées au hasard à partir d'un $2^r \times 2^r$ matrice d'identité. ϕ est le $2^r \times l$ code matrice, généré en utilisant $[l, r]$ schéma de codage linéaire avec la cartographie BPSK et mis à l'échelle par $2^{-r/2}$ de sorte que toutes les colonnes ont une norme unitaire. Il est basé sur des systèmes de codage de correction d'erreur populaires dans la communication.

Remarque 4.2 *Concernons la définition de la propriété du sous-espace d'injection, il est facile de montrer que \tilde{x} est une solution de faible précision. Cependant, on peut obtenir une solution de haute précision par une approche basée sur un préconditionneur aléatoire.*

4.7 Comparaison entre les différentes matrices aléatoires

Lorsque nous utilisons SRFT comme matrice randomisée pour calculer la solution approximative du problème LLS, nous avons le choix d'utiliser une FFT sous-échantillonnée. L'utilisation de la transformée de Fourier discrète F implique l'application d'une matrice complexe. Elle implique des comptages d'opérations et des pénalités de mémoire sur les phases lorsqu'il est appliqué sur une matrice réelle. Donc quand nous calculons $Y = A\Omega$, nous utilisons $O(nd\log(l))$ via un FFT sous-échantillonné.

La transformée randomisée SRHT peut être effectuée en $O(nd\log(s))$ temps, ce qui rend le SRHT plus efficace que la projection Gaussienne. De plus, les matrices d'Hadamard n'existent que pour les tailles qui sont en facteur de 2. Pour SRHT, on sait que les matrices d'Hadamard sont carrées ($2^r \times 2^r$ pour tout entier r). Donc, nous avons besoin de sous-échantillonner les colonnes en utilisant une matrice d'échantillonnage aléatoire pour obtenir la matrice d'Hadamard sous-échantillonnée qui est de taille $2^r \times d\log(d)$.

Nous remarquons que la conception d'une matrice de code sous-échantillonnée est similaire à celle des matrices SRFT et SRHT. La matrice de code décrite par Ubaru et al. [40,41] présente plusieurs avantages :

- Contrairement à SRHT, où une matrice d’Hadamard est appliquée à la matrice d’entrée et que les colonnes sont ensuite sous-échantillonées, nous n’avons pas besoin de sous-échantillonnage pour les matrices de code. Nous pouvons choisir un code de longueur désirée (nombre de colonnes) et les appliquer à la matrice.
- Plus important encore, nous pouvons montrer théoriquement que pour certaines matrices de code, nous avons seulement besoin de colonnes $O(d)$.
- La matrice de code présente plusieurs avantages dans le calcul parallèles et distribués.
- Les matrices de code avec certaines propriétés satisfont la transformé de Johnson Lindenstrauss (JLT) et les propriétés du sous-espace d’injection.
- Cependant, pour les matrices de code, on peut directement former une matrice $2^r \times l$ en utilisant un schéma de codage de longueur l (l est proche du rang d). Donc, nous n’avons pas besoin de sous-échantillonner.

En pratique, selon le travail d’Ubaru et al.[40], la plupart de ces matrices aléatoires (SRHT, SRFT, matrices de codes) ont des performances similaires lorsqu’elles sont appliquées à des matrices relativement denses.

5 Approximation des moindres carrés

Considérons le problème d’approximation des moindres carrés sur-déterminé, où la matrice A de $n \times d$ et $n \gg d$, il n’y a en général, aucun vecteur x tel que $Ax = b$, et il est courant de quantifier le meilleur en cherchant un vecteur x_{opt} telle que la norme euclidienne de l’erreur résiduelle soit faible, ie. pour résoudre le problème d’approximation des moindres carrés

$$x_{opt} = \min_x \| ZAx - Zb \|_2$$

Z doit satisfaire les deux conditions suivantes :

$$\begin{aligned} \sigma_{min}^2(ZU_A) &\geq \frac{1}{\sqrt{2}} \\ \| U_A^t Z^t Z b^\perp \|_2^2 &\leq \frac{\varepsilon}{2} \| Ax_{opt} - b \|_2^2 \end{aligned}$$

Notons que : $\| b^\perp \|_2 = \| Ax_{opt} - b \|_2$, pour certains $\varepsilon \in (0, 1)$, le vecteur solution \tilde{x}_{opt} au problème d’approximation des moindres carrés satisfait l’erreur relative suivante :

$$\begin{aligned} \| b - A\tilde{x}_{opt} \|_2 &\leq (1 + \varepsilon) \| b - Ax_{opt} \|_2 \\ \| x_{opt} - \tilde{x}_{opt} \|_2 &\leq \sqrt{\varepsilon} (k(A) \sqrt{\gamma^{-2} - 1}) \| x_{opt} \|_2 \end{aligned} \tag{II.17}$$

Où $k(A)$ est le conditionnement de A et

$$\gamma = \|UU^t b\|_2 / \|b\|_2$$

est un paramètre définissant la quantité de la masse de b dans l'espace des colonnes de A .

6 Conclusion

Les dernières années ont vu le développement d'algorithmes aléatoires en haute performance en applications informatiques. Cet intérêt accru est motivé par le fait que des algorithmes sont capables de surpasser les méthodes déterministes tout en fournissant des résultats exactes. En plus d'être plus facile à analyser.

Le principal avantage de ces algorithmes est qu'ils peuvent conduire à une solution beaucoup plus rapide en effectuant un plus petit nombre d'opérations ponctuelles. En conséquence, ils permettent potentiellement aux scientifiques du domaine d'adresser à des simulations plus importantes (qui contribuent également à obtenir des résultats plus précis). Autres applications de ces techniques statistiques pour accélérer les calculs d'algèbre linéaire peuvent être trouvés par exemple pour résoudre des systèmes linéaires utilisant des méthodes de Monte Carlo [11], ou pour des conditions de calcul d'estimation.

Chapitre III

Conditionnement des problèmes des moindres carrés

1 Introduction

Les approches standards de résolution d'un système linéaire sur-déterminé $Ax = b$ trouvent des corrections minimales au vecteur b où à la matrice A de sorte que le système corrigé soit cohérent, comme la méthode des moindres carrés. Le conditionnement permet de mesurer la sensibilité numérique de la solution du problème aux perturbations de ses données car il est utile d'évaluer la qualité numérique de la solution calculée. Le conditionnement est initialement défini dans la littérature comme le facteur d'amplification maximum entre une petite perturbation dans les données et le changement résultant dans la solution [16]. Il est utile d'évaluer la qualité numérique d'une solution calculée. Dans ce chapitre, nous nous intéressons à la formule du conditionnement en exploitant les propriétés du produit de Kronecker des matrices. Par la suite nous donnerons son estimation statistique afin de comparer les deux solutions.

Le problème des moindres carrés linéaire correspond à

$$\min_{x \in \mathbb{R}^d} \|Ax - b\|_2 \quad (\text{III.1})$$

avec $A \in \mathbb{R}^{n \times d}, d \ll n$ et $b \in \mathbb{R}^n$. Assumons que $\text{rang}(A) = d$, le problème consiste à

trouver une solution $x \in \mathbb{R}^d$, tel que

$$x = (A^t A)^{-1} A^t b = A^\dagger b \quad (\text{III.2})$$

où A^\dagger dénote la matrice pseudo-inverse de A définie par $A^\dagger = (A^t A)^{-1} A^t$.

Dans ce chapitre, nous voulons définir le conditionnement de la solution déduit des moindres carrés aéatoires, donc nous allons travailler avec $S^t A$ et $S^t b$.

On considère :

$$\begin{aligned} g : \mathbb{R}^{s \times d} \times \mathbb{R}^s &\longrightarrow \mathbb{R}^d \\ (D, B) &\longmapsto g(D, B) = L^t x = L^t (D^t D)^{-1} D^t B \end{aligned} \quad (\text{III.3})$$

Où $D = S^t A \in \mathbb{R}^{s \times d}$ et $B = S^t b \in \mathbb{R}^s$, où $S \in \mathbb{R}^{n \times s}$ une sketching matrice. Soit x la solution du problème, elle peut être exprimée comme fonction $g(y)$ des données y . Si g est différentiable, le conditionnement de g à y est défini comme [24]

$$k(y) = \max_{z \neq 0} \frac{\| g'(y) \cdot z \|}{\| z \|} \quad (\text{III.4})$$

Nous considérons la norme du produit définie sur $\mathbb{R}^{s \times d} \times \mathbb{R}^s$ par :

$$\| (D, B) \|_F = \sqrt{\| D \|_F^2 + \| B \|_2^2} \quad (\text{III.5})$$

La norme de Frobenius qui est définie dans l'espace des données est appropriée pour estimer l'erreur directe obtenue lorsque le problème des moindres carrés est résolu via l'équation normale. Tandis que la norme euclidienne est définie dans l'espace de la solution \mathbb{R}^d .

Dans ce chapitre, nous considérons le problème du calcul de la quantité $L^t x$, où x est la solution du problème linéaire des moindres carrés.

Puisque g est une fonction Fréchet différentiable alors D et B , $(D^t D)^{-1}$ sont aussi Fréchet différentiable au voisinage de (D, B) donc g est Fréchet dérivable au voisinage de (D, B) .

Selon [24] le conditionnement de g au (D, B) est donné par :

$$k(L, D, B) = \max_{(\Delta D, \Delta B) \neq 0} \frac{\| g'(D, B)(\Delta D, \Delta B) \|_2}{\| (\Delta D, \Delta B) \|_F} \quad (\text{III.6})$$

$k(L, D, B)$ s'appelle le conditionnement absolu de $L^t x$.

2 Formule explicite du conditionnement des moindres carrés

Proposition 2.1 (36) *Le conditionnement de la solution des moindres carrés est donné par :*

$$k(L, D, B) = \left\| \left[-x^t \otimes L^t D^\dagger + \left(r^t \otimes \left(L^t (D^t D)^{-1} \right) \right), L^t D^\dagger \right] \right\|_2 \quad (\text{III.7})$$

$$k(L, D, B) = \|C\|_2^{1/2} \quad (\text{III.8})$$

où C est une matrice symétrique définie par :

$$k(L, D, B) = \sqrt{(1 + \|x\|_2^2) \|L^t (D^t D)^{-1} D^t\|_2^2 + \|r\|_2^2 \|L^t (D^t D)^{-1}\|_2^2}$$

Preuve 2.1 *Considérons la fonction (III.3), soit $D = S^t A$ une matrice de rang complet, g' une fonction Fréchet dérivable. On applique la fonction linéaire $g'(D, B)$ aux variables $(\Delta D, \Delta B)$, tel que :*

$$\begin{aligned} g' : \mathbb{R}^{s \times d} \times \mathbb{R}^s &\longrightarrow \mathbb{R}^d \\ (\Delta D, \Delta B) &\longmapsto L^t (D^t D)^{-1} D^t (\Delta B - \Delta D x) + L^t (D^t D)^{-1} \Delta D^t r \end{aligned} \quad (\text{III.9})$$

La fonction g' peut être exprimée comme

$$g'(D, B)(\Delta D, \Delta B) = L^t D^\dagger (\Delta B - \Delta D x) + L^t (D^t D)^{-1} \Delta D^t r$$

$g'(D, B)(\Delta D, \Delta B)$ est l'image de $(\Delta D, \Delta B)$ par la fonction linéaire $g'(D, B)$, où D^\dagger dénote la matrice pseudo-inverse de D définie par, $D^\dagger = (D^t D)^{-1} D^t$ et r le vecteur du résidu $r = B - D x$.

On introduit l'opérateur de vectorisation des matrices vec . Soit M une matrice telle que

$$\text{vec}(g'(D, B)(\Delta D, \Delta B)) = M \begin{pmatrix} \text{vec}(\Delta D) \\ \text{vec}(\Delta B) \end{pmatrix} \quad (\text{III.10})$$

$$\begin{aligned} g'(D, B)(\Delta D, \Delta B) &= \text{vec}(g'(D, B)(\Delta D, \Delta B)) \\ &= \text{vec}\left(L^t (D^t D)^{-1} \Delta D^t r + L^t (D^t D)^{-1} D^t (\Delta B - \Delta D x)\right) \\ &= \left(-x^t \otimes L^t D^\dagger\right) \text{vec}(\Delta D) + \left(r^t \otimes \left(L^t (D^t D)^{-1}\right)\right) \text{vec}(\Delta D^t) + L^t D^\dagger \Delta B \end{aligned}$$

et $\text{vec}(\Delta B) = \Delta B$.

En utilisant (III.10) nous obtenons :

$$g'(D, B)(\Delta D, \Delta B) = \left[-x^t \otimes L^t D^\dagger + \left(r^t \otimes \left(L^t (D^t D)^{-1} \right) P \right), L^t D^\dagger \right] \begin{bmatrix} \text{vec}(\Delta D) \\ \Delta B \end{bmatrix}$$

Où

$$M = \left[-x^t \otimes L^t D^\dagger + \left(r^t \otimes \left(L^t (D^t D)^{-1} \right) P \right), L^t D^\dagger \right] \in \mathbb{R}^{d \times (nm+m)} \quad (\text{III.11})$$

Nous avons

$$\|(\Delta D, \Delta B)\|_F = \left\| \begin{bmatrix} \text{vec}(\Delta D) \\ \Delta B \end{bmatrix} \right\|_2$$

En utilisant l'équation (III.6), nous aurons :

$$k(L, D, B) = \|M\|_2.$$

Par conséquent, le conditionnement s'écrit via le produit de Kronecker comme (III.7).

Montrons la deuxième formule (III.8) de la proposition (2.1). Considérons l'équation (III.11).

Nous avons :

$$\|M\|_2 = \|M^t\|_2 = \|MM^t\|_2^{1/2}$$

et

$$M^t = \begin{bmatrix} -x \otimes H^t + P^t \left(r \otimes \left((DD^t)^{-1} \right)^t L \right) \\ H^t \end{bmatrix}$$

avec

$$H = L^t D^\dagger$$

Utilisons les propriétés du produit de Kronecker,

$$P(x \otimes D^t) = D^t \otimes x$$

$D^t D$ est une matrice symétrique donc $((D^t D)^{-1})^t = (D^t D)^{-1}$ par suite

$$\begin{aligned} MM^t &= x^t x \otimes HH^t + \left(r^t \otimes \left(L^t (D^t D)^{-1} P \right) \right) \left(P^t \left(r \otimes (D^t D)^{-t} \right) \right) L \\ &\quad + (x^t \otimes H) \left(P^t \left(r \otimes (D^t D)^{-t} L \right) \right) + \left(r^t \otimes \left(L^t (D^t D)^{-1} P \right) \right) (-x \otimes H^t) + HH^t \\ MM^t &= (1 + x^t x) HH^t + \|r\|_2^2 L^t (D^t D)^{-2} L - 2r^t H^t \otimes L^t (D^t D)^{-1} x \end{aligned}$$

Puisque $y^t \otimes z = zy^t$ alors

$$MM^t = (1 + x^t x) HH^t + \|r\|_2^2 L^t (D^t D)^{-2} L - 2L^t (D^t D)^{-1} x r^t H^t,$$

remplaçant $H = L^t D^\dagger = L^t (D^t D)^{-1} D^t$, on obtient :

$$\begin{aligned} MM^t &= (1 + x^t x) L^t (D^t D)^{-1} D^t D (D^t D)^{-1} L + \|r\|_2^2 L^t (D^t D)^{-2} L - \\ &\quad 2L^t (D^t D)^{-1} x r^t D (D^t D)^{-1} L \\ &= L^t (D^t D)^{-1} \left[(1 + x^t x) D^t D + \|r\|_2^2 I_n - 2x r^t D \right] (D^t D)^{-1} L. \end{aligned}$$

Nous avons

$$x r^t D = x(B - Dx)^t D = x(B^t - x^t D^t) D = xB^t D - x x^t D^t D$$

$$\begin{aligned} D^t B &= D^t D x \\ B^t D &= (D^t B)^t = x^t D^t D \\ x r^t D &= 0. \end{aligned}$$

Par conséquent :

$$\begin{aligned} MM^t &= L^t (D^t D)^{-1} \left[(1 + x^t x) D^t D + \|r\|_2^2 I_n \right] (D^t D)^{-1} L \\ &= L^t (D^t D)^{-1} \left[(1 + \|x\|_2^2) D^t D + \|r\|_2^2 I_n \right] (D^t D)^{-1} L \end{aligned}$$

Notons que $MM^t = C$ alors

$$k(L, D, B) = \|C\|_2^{1/2}$$

La formule explicite du conditionnement des moindres carrés $g(D, B)$ est définie par

$$k(L, D, B) = \sqrt{(1 + \|x\|_2^2) \|L^t (D^t D)^{-1} D^t\|_2^2 + \|r\|_2^2 \|L^t (D^t D)^{-1}\|_2^2} \quad (\text{III.12})$$

où $r = B - Dx$.

Dans la plupart des cas, L est défini comme étant la matrice identité I_d , cela correspond au conditionnement de la solution des moindres carrés, elle est exprimée par

$$k(I_d, D, B) = \sqrt{(1 + \|x\|_2^2) \|(D^t D)^{-1} D^t\|_2^2 + \|r\|_2^2 \|(D^t D)^{-1}\|_2^2} \quad (\text{III.13})$$

L peut prendre la forme d'un vecteur canonique $\{e_i^t\}_{i=1}^d$ de \mathbb{R}^d si nous sommes intéressés par

le conditionnement des composantes de x

$$k_i = \sqrt{(1 + \|x\|_2^2) \|e_i^t (D^t D)^{-1} D^t\|_2^2 + \|r\|_2^2 \|e_i^t (D^t D)^{-1}\|_2^2} \quad (\text{III.14})$$

Si nous résolvons le problème des moindres carrés par la méthode de factorisation QR , et en substituant dans l'équation (III.12) les équations suivantes :

$$\begin{aligned} \|L^t D^\dagger\|_2 &= \|R_D^{-t} L\|_2 \\ \|L^t (D^t D)^{-1}\|_2 &= \|R_D^{-1} (R_D^{-t} L)\|_2 \end{aligned}$$

considérant que $L = I_d$ tel que I_d matrice d'identité, nous obtenons la formule du conditionnement de la solution du problème randomisée

$$k_{LS} = \sqrt{\|R_D^{-1}\|_2 (\|R_D^{-1}\|_2^2 \|r\|_2^2 + \|x\|_2^2 + 1)}. \quad (\text{III.15})$$

3 Estimation statistique du conditionnement

L'estimation statistique du conditionnement est une méthode statistique appliquée aux matrices spécifiquement au problème des moindres carrés [27, 28].

On suppose que $g : \mathbb{R}^p \rightarrow \mathbb{R}$ est une fonction différentiable, calculons le gradient de g en fonction de y . Par la suite on utilise le développement de Taylor

$$g(y + \delta z) = g(y) + \delta (\nabla g(y))^t + O(\delta^2),$$

tel que $\delta \in \mathbb{R}$ est petit et $z \in \mathbb{R}^p$. Si z est sélectionné uniformément et aléatoire dans la sphère unité S_{p-1} [38]. L'espérance de la valeur estimée $\nu = \frac{|(\nabla g(y))^t z|}{\omega_p}$ satisfait

$$E(\nu) = \|\nabla g(y)\|_2$$

Où ω_p est le facteur de Wallis défini par

$$\left\{ \begin{array}{ll} \omega_1 = 1 & \\ \omega_2 = \frac{2}{\pi} & \\ \omega_p = \frac{1.3.5...(p-2)}{2.4.6...(p-1)} & \text{si } p \text{ impaire} > 2 \\ \omega_p = \frac{2}{\pi} \frac{1.3.5...(p-2)}{2.4.6...(q-1)} & \text{si } p \text{ paire} > 2 \end{array} \right. \quad (\text{III.16})$$

Le facteur de Wallis peut être approximée avec précision (démonstration dans annexe 1) par

$$\omega_p \approx \sqrt{\frac{2}{\pi \left(p - \frac{1}{2}\right)}}$$

On définit

$$\left| \frac{g(y + \delta z) - g(y)}{\delta \omega_p} \right|$$

comme premier ordre du conditionnement, qui peut être estimée $\|\nabla g(y)\|_2$ avec une grande probabilité pour la fonction g . Par conséquent, on définit le conditionnement estimé par :

$$\xi(q) = \frac{\omega_p}{\omega_q} \sqrt{\left|(\nabla g(y))^t z_1\right|^2 + \left|(\nabla g(y))^t z_2\right|^2 + \dots + \left|(\nabla g(y))^t z_q\right|^2}$$

D'un autre coté, nous pouvons estimer le conditionnement de la solution des problèmes des moindres carrés en utilisant la théorie des petits échantillons [37] qui fournit des estimations statistiques du conditionnement. Basé sur cette théorie Baboulin et al.[5] ont estimé le conditionnement statistique de la solution des moindres carrés.

On considère la solution x exprimer sous forme de fonction $x = g(y)$. en utilisant le théorème de Taylor. L'erreur δx de la solution $x(A, b)$ s'exprime comme suit :

$$\Delta x = x'(A, b) \cdot (\Delta A, \Delta b) + O\left(\|(\Delta A, \Delta b)\|_F^2\right)$$

telle que $x'(A, b) \cdot (\Delta A, \Delta b)$ est l'image de $(\Delta A, \Delta b)$ par la fonction linéaire $x'(A, b)$. Le conditionnement de la solution x correspond à la norme de $x'(A, b)$. Selon [27], $\|\Delta x\|_2$ est estimé par

$$\xi(q) = \frac{\omega_q}{\omega_n} \sqrt{|z_1^t \Delta x|^2 + |z_2^t \Delta x|^2 + \dots + |z_q^t \Delta x|^2}$$

Où z_1, \dots, z_q sont des vecteurs orthogonaux aléatoires sélectionnés uniformément de façon aléatoire à partir de la sphère unitaire de n dimensions.

Kenny et Laub [28] ont étudié l'estimation statistique du conditionnement appliquée au problème des moindres carrés, ils exploitent le fait que la fonction est Fréchet dérivable ainsi que la factorisation des donnés. Selon [27, 28] l'estimation statistique du \bar{k}_{LS} du problème des moindres carrés est exprimé comme

$$\bar{k}_{LS} = \frac{\omega_q}{\omega_n} \left(\sum_{i=1}^q k_i^2 \right)^{\frac{1}{2}} \quad (\text{III.17})$$

Où, q est considéré comme le nombre d'échantillons aléatoires et k_i le conditionnement de $z_i^t x$, on peut le calculer en utilisant l'expression comme

$$k_i = \sqrt{\|R^{-1}R^{-t}z_i\|_2^2 \|r\|_2^2 + \|R^{-t}z_i\|_2^2 (\|x\|_2^2 + 1)}$$

Chapitre IV

Résolution des problèmes larges des moindres carrés

1 Introduction

Dans ce chapitre, nous allons travailler sur trois méthodes aléatoires. Trois algorithmes différents afin de trouver la solution la plus approximée au problème des moindres carrés dont les matrices incluses sont de grandes tailles. Comme première partie du travail nous devons effectuer un échantillonnage aléatoire moyen d'une projection afin de respecter la moyenne statistique et par la suite, accélérer le calcul de la solution des moindres carrés.

Afin de résoudre le problème des moindres carrés avec précision, nous proposons une approche basée sur un préconditionneur aléatoire, puis nous itérons avec une méthode d'itération. Pour trouver la matrice de préconditionnement, nous calculons la factorisation QR de A , malheureusement cette étape est coûteuse. Dans ce cas, Woodruff [42] a proposé l'utilisation des méthodes aléatoires pour trouver R approximativement.

Soit S une matrice aléatoire, nous formons $L = S^t A = QR$ telle que QR est la décomposition de L . Ensuite, nous choisissons R^{-1} comme matrice de préconditionnement de A [30]. Par la suite, nous appelons une méthode d'itération avec R^{-1} comme préconditionneur. Finalement, nous déterminons \tilde{x} tel que $R\tilde{x} = y$.

Le choix approprié du préconditionneur est pris en compte car le conditionnement de la matrice AR^{-1} est petit. Par la suite, l'utilisation d'une méthode itérative tel que le gradient conjugué nécessite quelques itérations pour obtenir y .

La résolution du problème s'effectue en deux étapes :

- Construire une matrice S préservant le sous-espace d'injection.
- Résoudre le problème des moindres carrés de taille réduite,

$$\min_x \| S^t A \tilde{x} - S^t b \|_2 \quad (\text{IV.1})$$

où la solution $\tilde{x} = (S^t A)^\dagger S^t b$.

Le problème peut être résolu en $O(sd^2)$ temps. Ainsi, obtenir une solution approximative au problème original.

2 Résolution des problèmes des moindres carrés

Dans ce chapitre, nous intéressons aux propriétés de la matrice aléatoire $S^t H D$ pour obtenir rapidement la meilleure solution approximée du problème original. Nous commençons par le lemme 2.1.

Lemme 2.1 (30) *Soit $U \in \mathbb{R}^{n \times d}$ une matrice orthogonale, et soit HD la matrice d'Hadamard, alors pour une probabilité d'au moins de $19/20$, nous avons*

$$\|(H D U)_{(i)}\|_2 \leq \sqrt{\frac{2d \log(40nd)}{n}}$$

Preuve 2.1 Pour $i \in \{1, \dots, n\}$ et $j \in \{1, \dots, d\}$, $(H D U)_{ij}$ s'écrit comme $\sum_{k \in [n]} \alpha_k U_{kj}$ où $\alpha_k = \pm 1$ est une variable indépendante aléatoire.

Prenons $s = \sqrt{\frac{2 \log(40nd)}{n}}$ nous avons :

$$P \left[|(H D U)_{ij}| > s \right] \leq \frac{1}{20nd}.$$

Puisque les colonnes de U sont orthonormales alors

$$\| U \|_2 = \sqrt{\sum_{k \in [n]} |U_{kj}|^2} = 1$$

Pour tout i, j et avec une probabilité au moins de $19/20$ nous avons :

$$|(H D U)_{ij}| \leq s, \quad \forall i \in \{1, \dots, n\}, j \in \{1, \dots, d\}$$

Ceci implique que

$$\forall i \in \{1, \dots, n\}, \|(H D U)_{(i)}\| \leq s \sqrt{d}.$$

Théorème 2.1 Considérons le problème d'approximation des moindres carrés sur-déterminés, $A \in \mathbb{R}^{n \times d}$: $A = U \Sigma V^t$ est le SVD de A , soit $b \in \mathbb{R}^n$ et $b^\perp = U_A U_A^t b$ est une partie de b , U_A contenant d vecteurs de A et $\varepsilon \in (0, 1)$. Soit S une matrice d'échantillonnage rééchelonnées uniformément, HD est la matrice aléatoire d'Hadamard.

Alors, la matrice $S^t HD$ satisfait les deux conditions suivants :

$$\sigma_{\min}^2(S^t H D U_A) \geq \frac{1}{\sqrt{2}} \quad (\text{IV.2})$$

$$\|U_A^t (S^t HD)^t S^t H D b^\perp\|_2^2 \leq \frac{\varepsilon \|b^\perp\|_2^2}{2} \quad (\text{IV.3})$$

Preuve 2.2 D'après le lemme 2.1 prouvé précédemment, pour toute matrice orthogonale U et $n \times n$ matrice d'Hadamard HD , et avec une probabilité d'au moins 0.95, nous avons

$$\|(H D U)_{(i)}\|_2^2 \leq \frac{2 \log(40nd)}{n}, \forall i \in \{1, \dots, n\} \quad (\text{IV.4})$$

Pour tout $i \in \{1, \dots, n\}$, nous avons

$$\begin{aligned} |1 - \sigma_i^2(S^t H D U_A)| &= |\sigma_i(U_A^t D H^t H D U_A) - \sigma_i(U_A^t D H^t S S^t H D U_A)| \\ &\leq \|U_A^t D H^t H D U_A - U_A^t D H^t S S^t H D U_A\|_2 \end{aligned}$$

Considérons la matrice $(H D U_A)^t$, du moment que HD et U_A sont des matrices orthogonales. Par suite, $\|H D U_A\|_2 = 1$ et $\|H D U_A\|_F = \sqrt{d}$.

Soit $\beta = \frac{1}{2 \log(40nd)}$, alors

$$\frac{1}{n} \geq \beta \frac{\|(H D U_A)_{(i)}\|_2^2}{\|H D U_A\|_F^2}, \forall i \in [n] \quad (\text{IV.5})$$

Selon le théorème lié à la norme spectrale et en supposant que $\varepsilon = 1 - \frac{1}{\sqrt{2}}$ et $\delta = \frac{1}{10}$ donc avec une probabilité $\geq 1 - \delta \simeq 0.9$, nous avons :

$$\|U_A^t D H^t H D U_A - U_A^t D H^t S S^t H D U_A\|_2 \leq 1 - \frac{1}{\sqrt{2}} \quad (\text{IV.6})$$

donc

$$\sigma_{\min}^2(S^t H D U_A) \geq \frac{1}{\sqrt{2}}. \quad (\text{IV.7})$$

Démontrons la deuxième inégalité (IV.3) du théorème 2.1

Rappelons que $b^\perp = U_A^\perp U_A^{\perp t} b$ et $\| U_A^t D H^t H D b^\perp \|_2^2 = 0$, nous avons :

$$\| (S^t H D U_A)^t S^t H D b^\perp \|_2^2 = \| U_A^t D H^t S S^t H D b^\perp - U_A^t D H^t H D b^\perp \|_2^2 \quad (\text{IV.8})$$

En prenant en considération l'inégalité (IV.4) et $\| H D U_A \|_F = \sqrt{d}$, nous avons

$$\begin{aligned} E[\| (S^t H D U_A)^t S^t H D b^\perp \|_2^2] &\leq \frac{1}{\beta r} \| H D U_A \|_F^2 \| H D b^\perp \|_2^2 \\ &\leq \frac{d \| b^\perp \|_2^2}{\beta r} \end{aligned} \quad (\text{IV.9})$$

En appliquant l'inégalité de Markov alors avec une probabilité ≥ 0.9 on obtient :

$$\| (S^t H D U_A)^t S^t H D b^\perp \|_2^2 \leq \frac{10d \| b^\perp \|_2^2}{\beta r} \quad (\text{IV.10})$$

Si on prend $r \geq 20\beta^{-1}\frac{d}{\varepsilon}$, nous aurons

$$\| (S^t H D U_A)^t S^t H D b^\perp \|_2^2 \leq \frac{\| b^\perp \|_2^2 \varepsilon}{2} \quad (\text{IV.11})$$

Par conséquent, la solution \tilde{x} du problème en prenant $S^t H D$ comme matrice aléatoire satisfait une borne d'erreur relative à la forme décrite dans le théorème 2.2.

Théorème 2.2 [30] Considérons le problème d'approximation des moindres carrés, U_A contient les d vecteurs singuliers de A et $\varepsilon \in (0, 1)$. Nous supposons que $S^t H D$ satisfait aux conditions du théorème 2.1. Par la suite, avec une probabilité d'au moins 0.8, la solution $\tilde{x} \in \mathbb{R}^d$ du problème des moindres carrés satisfait :

$$\| A \tilde{x}_{opt} - b \|_2 \leq (1 + \varepsilon) \min_{x \in \mathbb{R}^d} \| A x - b \|_2 \quad (\text{IV.12})$$

Si nous supposons que $\| U_A U_A^t b \|_2 \geq \gamma \| b \|_2$, pour $\gamma \in (0, 1]$ alors

$$\| x_{opt} - \tilde{x}_{opt} \|_2 \leq \sqrt{\varepsilon} (k(A) \sqrt{\gamma^{-2} - 1}) \| x_{opt} \|_2 \quad (\text{IV.13})$$

Où $k(A)$ est le conditionnement de la matrice A .

Preuve 2.3 soit $z_{opt} \in \mathbb{R}^d$, $U_A z_{opt} = A(\tilde{x}_{opt} - x_{opt})$, $b^\perp = b - Ax_{opt}$

Toutefois,

$$\min_{x \in \mathbb{R}^d} \| S^t H D A x - S^t H D b \|_2^2 =$$

$$\begin{aligned} &= \min_{x \in \mathbb{R}^d} \| S^t H D (A x_{opt} + y) - S^t H D (A x_{opt} + b^\perp) \|_2^2 \\ &= \min_{y \in \mathbb{R}^d} \| S^t H D (A y) - S^t H D b^\perp \|_2^2 \\ &= \min_{z \in \mathbb{R}^d} \| S^t H D (U_A z) - S^t H D b^\perp \|_2^2 \end{aligned}$$

$$Or, \| S^t H D (U_A z_{opt}) - S^t H D b^\perp \|_2^2 =$$

$$\begin{aligned} &= \| S^t H D A (\tilde{x}_{opt} - x_{opt}) - S^t H D b^\perp \|_2^2 \\ &= \| S^t H D A \tilde{x}_{opt} - S^t H D A x_{opt} - S^t H D b^\perp \|_2^2 \\ &= \| S^t H D A \tilde{x}_{opt} - S^t H D b \|_2^2 \\ &= \min_{x \in \mathbb{R}^d} \| S^t H D A x - S^t H D b \|_2^2 \\ &= \min_{z \in \mathbb{R}^d} \| S^t H D U_A z - S^t H D b^\perp \|_2^2 \end{aligned}$$

alors,

$$\min_{z \in \mathbb{R}^d} \| S^t H D (U_A z) - S^t H D b^\perp \|_2^2 = \| S^t H D (U_A z_{opt}) - S^t H D b^\perp \|_2^2. \quad (\text{IV.14})$$

Ainsi, par l'équation normale de (IV.14), nous obtenons :

$$(S^t H D U_A)^t (S^t H D U_A) z_{opt} = (S^t H D U_A)^t (S^t H D b^\perp). \quad (\text{IV.15})$$

Prenons la norme des deux côtés de (IV.15), alors :

$$\frac{1}{2} \| z_{opt} \|_2^2 \leq \| (S^t H D U_A)^t (S^t H D U_A) z_{opt} \|_2^2.$$

Selon les équations (IV.15) and (IV.3) nous obtenons

$$\| z_{opt} \|_2^2 \leq \varepsilon Z^2. \quad (\text{IV.16})$$

Nous avons :

$$\begin{aligned} \| A \tilde{x}_{opt} - b \|_2^2 &= \| A \tilde{x}_{opt} - A x_{opt} + A x_{opt} - b \|_2^2 \\ &= \| A x_{opt} - b \|_2^2 + \| A x_{opt} - A \tilde{x}_{opt} \|_2^2 \\ &= Z^2 + \| -U_A z_{opt} \|_2^2 \end{aligned} \quad (\text{IV.17})$$

où $Z = \min_{x \in \mathbb{R}^d} \|Ax - b\|$

Alors :

$$\begin{aligned}
 \|A\tilde{x}_{opt} - b\|_2^2 &\leq Z^2 + \varepsilon Z^2 \\
 \|A\tilde{x}_{opt} - b\|_2^2 &\leq (1 + \varepsilon)Z^2 \\
 \|A\tilde{x}_{opt} - b\|_2 &\leq \sqrt{(1 + \varepsilon)}Z \\
 \|A\tilde{x}_{opt} - b\|_2 &\leq (1 + \varepsilon)Z \\
 \|A\tilde{x}_{opt} - b\|_2 &\leq (1 + \varepsilon)\min_{x \in \mathbb{R}^d} \|Ax - b\|
 \end{aligned} \tag{IV.18}$$

Prouvons la deuxième inégalité (IV.13) du théorème 2.2

Par hypothèse $\|U_A U_A^t b\|_2 \geq \gamma \|b\|_2$, pour $\gamma \in (0, 1]$, donc

$$\|b\|_2 \leq \gamma^{-1} \|U_A U_A^t b\|_2, \tag{IV.19}$$

Puisque $U_A U_A^t b = Ax_{opt}$ alors

$$\begin{aligned}
 Z^2 &= \|b\|_2^2 - \|U_A U_A^t b\|_2^2 \\
 Z^2 &\leq (\gamma^{-2} - 1) \|U_A U_A^t b\|_2^2
 \end{aligned}$$

$$\|U_A U_A^t b\|_2 = \|Ax_{opt}\|_2 \leq \|A\|_2 \|x_{opt}\|_2 = \sigma_{\max}(A) \|x_{opt}\|_2$$

Rappelons que :

$$\begin{aligned}
 A(x_{opt} - \tilde{x}_{opt}) &= U_A z_{opt} \\
 \text{alors } \|A(x_{opt} - \tilde{x}_{opt})\|_2^2 &= \|U_A z_{opt}\|_2^2
 \end{aligned}$$

Selon l'équation (IV.16) nous obtenons : $\|x_{opt} - \tilde{x}_{opt}\|_2^2 \leq \frac{\varepsilon}{\sigma_{\min}^2(A)} Z^2$,

donc

$$\begin{aligned}
 \|x_{opt} - \tilde{x}_{opt}\|_2 &\leq \frac{1}{\sigma_{\min}(A)} \sqrt{\varepsilon} \sqrt{(\gamma^{-2} - 1)} \|U_A U_A^t b\|_2 \\
 \|x_{opt} - \tilde{x}_{opt}\|_2 &\leq \frac{\sigma_{\max}(A)}{\sigma_{\min}(A)} \sqrt{\varepsilon} \sqrt{\gamma^{-2} - 1} \|x_{opt}\|_2 \\
 \|x_{opt} - \tilde{x}_{opt}\|_2 &\leq k(A) \sqrt{\varepsilon} \sqrt{\gamma^{-2} - 1} \|x_{opt}\|_2
 \end{aligned}$$

2.1 Description d'algorithme 1

Considérons le problème des moindres carrés [36]

$$Z = \min_{x \in \mathbb{R}^d} \|Ax - b\|_2, \tag{IV.20}$$

où $A \in \mathbb{R}^{n \times d}$, $d \ll n$, $b \in \mathbb{R}^n$. Nous supposons que n est une puissance de 2 et que $\text{rang}(A) = d$. L'équation (IV.20) peut être résolu plus efficacement par les méthodes aléatoires.

Pour notre algorithme (FAST SHD)[36], on considère une matrice S d'échantillonnage uniforme $n \times r$ et de rééchelonnement. Elle échantillonne uniformément les r contraintes. Elle est définie par une matrice diagonale qui a une valeur non nulle par ligne qui est égale à $\frac{1}{\sqrt{rp_i}}$ et zéro autrement. Puisque nous échantillonnons uniformément, nous substituons la probabilité par $p_i = \frac{1}{n}$. Par la suite, nous effectuons $L = S^t H D A$ avec $H D$ est matrice d'Hadamard aléatoire.

Dans cette partie, nous intéressons à la solution de haute précision du problème d'approximation des moindres carrés. Ensuite, nous calculons la factorisation QR de $L = S^t H D A = Q_L R_L$, où R_L est une matrice triangulaire supérieure, afin d'obtenir la matrice de préconditionnement R^{-1} . Cette matrice a été prouvé par woodruff [42] qu'elle représente la meilleure matrice du préconditionnement car $k(R^{-1})$ est petit.

Par conséquent, le problème initial de minimiser $\|Ax - b\|_2$ avec n contraintes revient à minimiser $\|S^t H D A R^{-1} y - S^t H D b\|_2$ définie sur r contraintes ($r < n$). Nous allons résoudre ce dernier problème par la méthode du gradient conjugué avec solution initiale $y^{(0)} = (S^t H D A R_L^{-1})^\dagger (S^t H D b)$. Ensuite, on déduit que

$$\tilde{x} = R_L^{-1} y$$

Entrée : $A \in \mathbb{R}^{n \times d}$, A de rang complet et $b \in \mathbb{R}^n$, $\varepsilon \in (0, 1)$

Sortie : \tilde{x} , solution approximative du problème

1. Soit $r = O(d(\log d)(\log n) + \frac{d(\log n)}{\varepsilon})$.
2. Soit S une $n \times r$ matrice d'échantillonage uniforme rééchelonnée.
3. Effectuer le Sketching $L = S^t H D A \in \mathbb{R}^{r \times d}$.
4. Calculer la factorisation QR de L : $[Q_L, R_L] = QR(L)$.
5. Calculer la matrice de préconditionnement R_L^{-1} .
6. Calculer la solution initiale $y^{(0)} = (S^t H D A R_L^{-1})^\dagger S^t H D b = Q_L^t (S^t H D b)$.
7. Utiliser une méthode itérative pour résoudre le problème.
8. Return $\tilde{x} = R_L^{-1} y^{(t)}$

ALG 1 – Fast SHD

2.2 Description d'algorithme 2

Cet algorithme [36] consiste à combiner deux méthodes aléatoires count-sketch et SRHT. Considérons le problème des moindres carrés

$$\min_{x \in \mathbb{R}^d} \|Ax - b\|_2, \quad (\text{IV.21})$$

où $A \in \mathbb{R}^{n \times d}$, $d \ll n$, $b \in \mathbb{R}^n$, $\text{rang}(A) = d$.

Puisque $d \ll n$ ce problème peut être résolu plus efficacement avec les méthodes aléatoires. Soit $S \in \mathbb{R}^{n \times s_2}$ tel que $D = S^t A$. Nous calculons la solution approximative :

$$\tilde{x} = \min_{x \in \mathbb{R}^d} \|S^t Ax - S^t b\|_2 \quad (\text{IV.22})$$

avec

$$S = S_{CS} S_{SRHT}$$

où, $S_{CS} \in \mathbb{R}^{n \times s_1}$ matrice aléatoire de count-sketch et $S_{SRHT} \in \mathbb{R}^{s_1 \times s_2}$, telles que $s_1 = O(s^2 \log^6 \frac{s}{\varepsilon} + \frac{s}{\varepsilon})$ et $s_2 = O(\frac{s}{\varepsilon} \log \frac{s}{\varepsilon})$.

s_1, s_2 assure l'erreur relative $(1 + \varepsilon)$.

Pour la valeur de s nous pouvons choisir un facteur déchantillonnage $\gamma > 1$, tel que $s = \gamma d$.

Par conséquent :

$$D = S_{SRHT}^t S_{CS}^t A = S^t A \in \mathbb{R}^{s_2 \times d} \text{ and } B = S_{SRHT}^t S_{CS}^t b = S^t b \in \mathbb{R}^{s_2} \quad (\text{IV.23})$$

Calculons la factorisation QR de D , telle que $D = Q_D R_D$, R_D est approximativement trouvée dans $O(nnz(A) + \text{poly}(\frac{d}{\varepsilon}))$ temps. Alors, pour simplifier la résolution du problème nous allons trouver la solution du problème

$$\min_{x \in \mathbb{R}^d} \|Dx - B\|_2 = R_D^\dagger Q_D^t B \quad (\text{IV.24})$$

où R_D^\dagger et Q_D^t sont respectivement la pseudo-inverse de R et la transposée de Q_D calculée à partir de la factorisation QR de $D = S^t A$.

Entrée : $A \in \mathbb{R}^{n \times d}$, $\text{rang}(A) = d$ et $b \in \mathbb{R}^n$.

Sortie : \tilde{x} , la solution du problème

1. Soit $s = \gamma d$, $\gamma > 1$
2. Prenons les paramètres $s_1 = s^2 \log^6 \frac{s}{\varepsilon} + \frac{s}{\varepsilon}$, et $s_2 = \frac{s}{\varepsilon} \log \frac{s}{\varepsilon}$;
3. Etablir $S_{CS} \in \mathbb{R}^{n \times s_1}$ count sketch matrice et $S_{SRHT} \in \mathbb{R}^{s_1 \times s_2}$
4. Effectuer le sketching : $D = S_{SRHT}^t S_{CS}^t A \in \mathbb{R}^{s_2 \times d}$ et $B = S_{SRHT}^t S_{CS}^t b \in \mathbb{R}^{s_2}$;
5. Calculer la factorisation QR de D : $[Q_D, R_D] = QR(D)$;
6. Calculer $\tilde{x} = R_D^\dagger Q_D^t B$.

ALG 2 – Fast SRHT CS

La combinaison de SRHT avec count-sketch peut être calculer aussi efficacement que le count-sketch uniquement. Théoriquement, la matrice S produite par la combinaison de count-sketch et SRHT vérifie la propriété du sous-espace d'injection avec une probabilité de 0.99.

2.3 Description d'algorithme 3

Considérons le problème des moindres carrés

$$\min_{x \in \mathbb{R}^d} \|Ax - b\|_2, \quad (\text{IV.25})$$

où, $A \in \mathbb{R}^{n \times d}$, $d \ll n$, $b \in \mathbb{R}^n$ et $\text{rang}(A) = d$.

Nous allons faire la même procédure utilisée au paravant avec la combinaison de count-sketch et SRHT. Dans ce cas, nous allons combiner la méthode de count-sketch et la projection Gaussienne [36] afin de minimiser

$$\tilde{x} = \min_{x \in \mathbb{R}^d} \|S^t Ax - S^t b\|_2, \quad (\text{IV.26})$$

telles que $S = S_{CS} S_{GP} \in \mathbb{R}^{n \times s_2}$ une sketching matrice avec $D = S^t A$.

$S_{CS} \in \mathbb{R}^{n \times s_1}$ matrice de count sketch et $S_{GP} \in \mathbb{R}^{s_1 \times s_2}$ matrice de Projection Gaussienne.
Prenons les paramètres s_1, s_2 comme suit :

$$s_1 = \frac{s}{\varepsilon} \log \frac{s}{\varepsilon} \quad \text{et} \quad s_2 = \frac{s}{\varepsilon} \log \frac{s}{\varepsilon}$$

Pour la valeur de s , nous pouvons choisir le facteur d'échantillonnage $\gamma > 1$, tel que $s = \gamma d$
En conséquence

$$D = S_{GP}^t S_{CS}^t A \in \mathbb{R}^{s_2 \times d} \quad \text{et} \quad B = S_{GP}^t S_{CS}^t b \in \mathbb{R}^{s_2} \quad (\text{IV.27})$$

En effet, calculons

$$\min_{x \in \mathbb{R}^d} \|Dx - B\|_2 = R_D^\dagger Q_D^t B, \quad (\text{IV.28})$$

où, R_D^\dagger et Q_D^t sont respectivement le pseudo-inverse de R et la transposée de Q_D calculée par la factorisation QR de $D = SA$.

Entrée : $A \in \mathbb{R}^{n \times d}$, $\text{rang}(A) = d$ et $b \in \mathbb{R}^n$.

Sortie : \tilde{x} , solution du problème

1. Soit $s = \gamma d$, $\gamma > 1$
2. Prenons : $s_1 = \frac{s}{\varepsilon} \log \frac{s}{\varepsilon}$, et $s_2 = \frac{s}{\varepsilon} \log \frac{s}{\varepsilon}$
3. Etablir $S_{CS} \in \mathbb{R}^{n \times s_1}$ matrice de count sketch et $S_{GP} \in \mathbb{R}^{s_1 \times s_2}$ matrice
4. Effectuer le Sketching : $D = S_{GP}^t S_{CS}^t A \in \mathbb{R}^{s_2 \times d}$ et $B = S_{GP}^t S_{CS}^t b \in \mathbb{R}^{s_2}$;
5. Etablir la décomposition QR de D : $[Q_D, R_D] = QR(D)$
6. Calculer $\tilde{x} = R_D^\dagger Q_D^t B$.

ALG 3 – Fast GP CS

3 Application numérique

Dans ce chapitre, nous décrivons notre approche de calcul des coefficients géopotentiels harmoniques afin de calculer le potentiel gravitationnel qui nécessite la résolution d'un large problème linéaire des moindres carrés. La théorie de randomisation nous aide à résoudre ce problème en minimisant le temps et la capacité de mémoire.

L'objectif de cette application est de calculer plus précisément que possible des coefficients harmoniques normalisés \bar{C}_{lm} et \bar{S}_{lm} pour une détermination du potentiel gravitationnel V tel que :

$$V(r, \theta, \lambda) = \frac{GM}{R} \sum_{i=0}^n \left(\frac{R}{r}\right)^{l+1} \sum_{m=0}^l \bar{P}_{lm}(\cos\theta) [\bar{C}_{lm} \cos m\lambda + \bar{S}_{lm} \sin m\lambda], \quad (\text{IV.29})$$

où r, θ, λ définissent respectivement le rayon géocentrique, la latitude et la longitude, GM est le produit de la constante gravitationnelle et la masse de la Terre, R est le rayon de la Terre, \bar{P}_{lm} est la fonction de Legendre normalisée de degré l et ordre m .

Pour cette étude, nous considérons 30 jours d'observations simulées avec un pas de $10''$ de la mission satellitaire correspondant au satellite GOCE (Gravity field and steady state Ocean circulation Explorer), Les mesures sont simulées à partir de la position du satellite, sa vitesse et son accélération obtenues par les stations sol (Laser, Doppler) ou autres satellites (GPS). Le nombre total d'observations est de $m = 777603$ et le nombre de coefficients harmoniques sphériques n' se calcule en fonction du degré du modèle géopotentiel l , on le détermine par la formule

$$n' = l^2 + 2l - 2$$

Dans cet application on va prendre $l = 20$ par la suite, nous avons 438 coefficients harmoniques à déterminer.

Dans la première étape, nous calculons les paramètres de rotation simulées du système céleste vers le système terrestre. Dans cette étape nous utilisons la subroutine TERCEL du package NOVAK. Puis, nous calculons la matrice de configuration A basé sur la théorie d'Austin et Reubelt [1]. Ensuite nous résolvons le problème des moindres carrés en prenant en considération les algorithmes 2 et 3 proposés précédemment.

Selon l'algorithme 1, nous formons la matrice randomisée L avec un échantillonnage uniforme et on calcule la projection aléatoire d'Hadamard. Ensuite, nous calculons QR de L . On note que la matrice de préconditionnement est l'inverse de R . Par la suite nous calculons la solution initiale du problème des moindres carrés par méthode itératif. La méthode d'itération utilisée ici est le gradient conjugué, ensuite nous calculons la solution approximative du problème original \tilde{x} qui est le produit du préconditionnement avec la solution trouvée précédemment.

Le conditionnement de la solution des moindres carrés mesure l'effet sur la solution des petits changements dans les données. Dans notre application, nous étudions le conditionnement des composantes de la solution LLS aléatoire en prenant en considération $S^t HD$ comme matrice

aléatoire, il est défini par :

$$k_i = \sqrt{(1 + \|x\|_2^2) \|e_i^t (V^t V)^{-1} V^t\|_2^2 + \|r\|_2^2 \|e_i^t (V^t V^{-1})\|_2^2} \quad (\text{IV.30})$$

ou $S^t HD = V$,

Ci-dessous une figure qui représente la variation du conditionnement des composantes x_i de la solution du problème des moindres carrés où la matrice aléatoire est $S^t HD$.

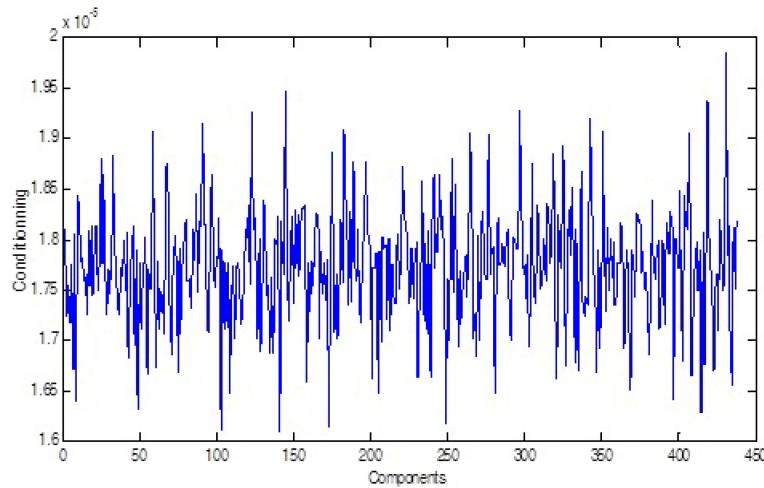


Figure IV.1 – Comportement du conditionnement des composantes x_i de la solution du problème calculé avec $S^t HD$ comme matrice aléatoire.

Nous remarquons que le conditionnement des composantes de la solution du problème aléatoire linéaire des moindres carrés est dans le voisinage de 10^{-5} .

Donc, nous donnons ici quelques avantages pour combiner la projection Gaussienne avec count-sketch [36] : la taille de la combinaison des deux matrices aléatoires est aussi petit que la projection Gaussienne, la complexité temporelle est beaucoup plus basse que la projection Gaussienne quand $d \ll n$, et l'implémentation est plus facile que SRHT avec count-sketch

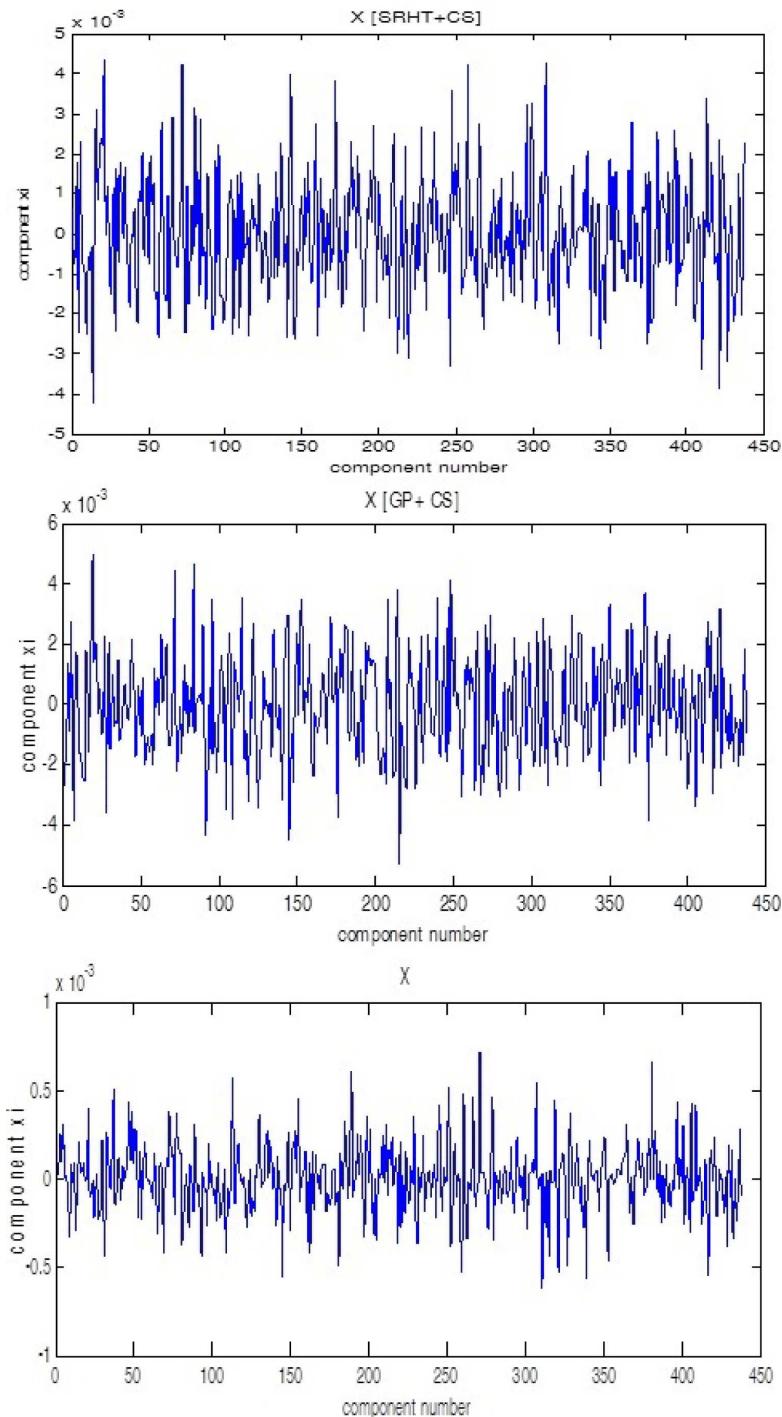


Figure IV.2 – Variation des composantes de la solution des moindres carrés x_i dans différents cas

Dans la figure (IV.2) nous présentons la variation des composantes de la solution des

moindres carrés linéaires x . La première figure indique x_i avec la combinaison de count-sketch et SRHT, la seconde montre celle de count-sketch avec la projection Gaussienne et la troisième figure montre cette variation avec la méthode des équations normales sans utiliser la randomisation.

Par la suite, un autre programme Matlab est écrit pour le calcul du conditionnement absolu des différentes composantes x_i des moindres carrés pour le cas où nous n'utilisons pas la randomisation, ensuite nous incluons la combinaison de la projection Gaussienne avec count-sketch

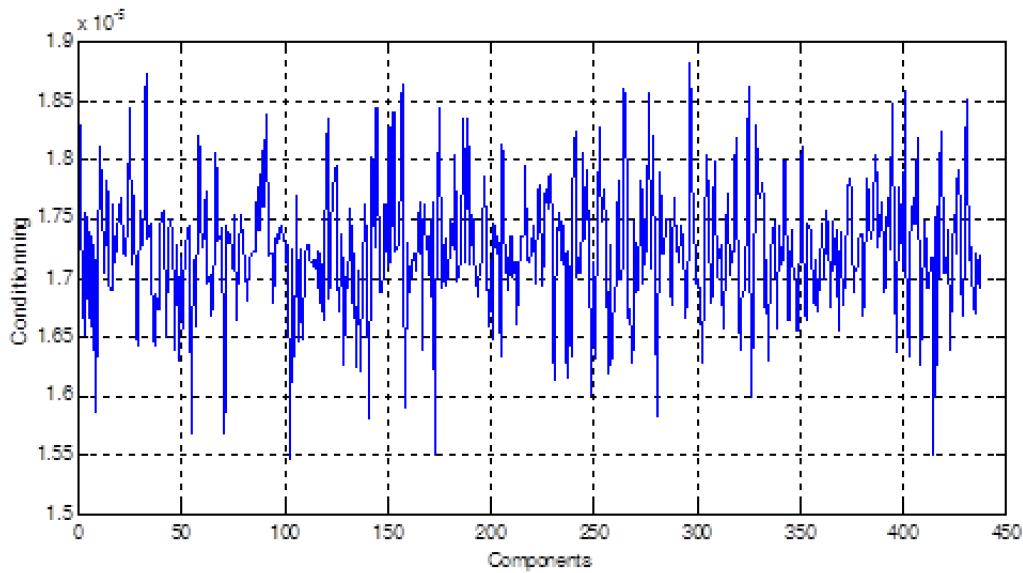


Figure IV.3 – Conditionnement des composantes x_i du problème des moindres carrés linéaires

Figure (IV.3) correspond au conditionnement des coefficients x_i , dans le cas où la randomisation n'est pas utilisée.

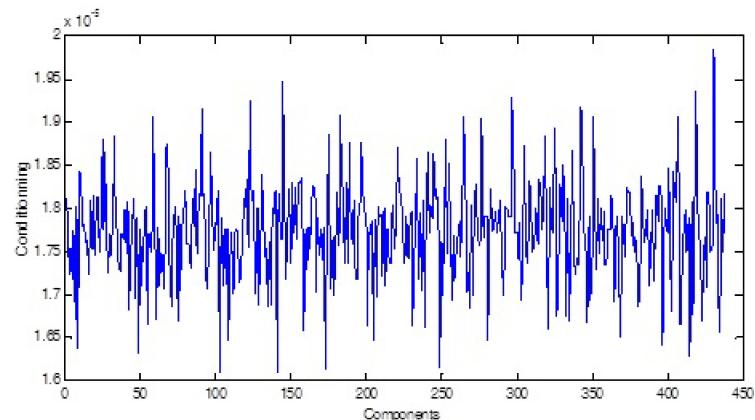


Figure IV.4 – Variation du conditionnement des composantes x_i calculé à partir de la combinaison de count-sketch et la projection Gaussienne

Figure (IV.4) définit le conditionnement des composantes x_i calculé par la combinaison du count-sketch et la projection Gaussienne : nous remarquons que le conditionnement varie de l'ordre de 10^{-5} . Notons que le conditionnement absolu dans ce cas égal à $2,79 \times 10^{-5}$.

Conclusion et perspectives

Les moindres carrés linéaires est un problème de calcul très importants, qui est originale de la nécessité d'ajuster un modèle mathématique linéaire à des observations données. Ils sont appliqués dans plusieurs domaines notamment en ingénierie tels que les statistiques, la géodésie, la photogrammétrie, le traitement du signal et le contrôle.

Nous avons montré dans cette thèse comment les calculs d'algèbre linéaire peuvent être accélérés en utilisant des techniques statistiques dans le cas d'un système linéaire. Nous avons étudié des transformations aléatoires de A qui nous permettent d'éviter de pivoter et par la suite réduire le volume des matrices. Ces transformations s'effectuent à un coût calculatoire très abordable tout en donnant une précision satisfaisante lorsqu'on les compare avec le pivotage.

L'utilisation des méthodes aléatoires pour l'analyse des données volumineuse sont très utilisées à cause de la rapidité du calcul en plus de sa simplicité à mettre en œuvre que les méthodes traditionnelles.

Le but de ce travail est de démontrer que les méthodes aléatoires représentent un outil puissant pour la résolution des problèmes des moindres carrés larges. Le choix entre les différentes matrices aléatoires afin d'obtenir la solution du problème des moindres carrés dépend de la recherche d'une solution avec une probabilité élevée. Dans cette thèse, nous avons analysé les différents types de matrices aléatoires. A noter que nous pouvons utiliser les matrices aléatoires sur des matrices relativement denses ou creuses.

Dans cette thèse, parmi les trois algorithmes définis dans le chapitre 4, nous déduisons que la combinaison utilisé dans le premier algorithme est la meilleure puisqu'elle tient compte d'une méthode aléatoire et qu'elle s'intéresse à l'aspect statistique. Dans cet algorithme nous

échantillonnons uniformément et rééchelonnons la projection d’Hadamard de la matrice A , dont le but est d’accélérer le calcul de la solution avec précision. Nous remarquons que l’algorithme 1 fonctionne mieux que les autres modèles.

Le conditionnement des moindres carrés est considéré comme une étape fondamental pour l’analyse de la sensibilité des solutions. Il a été appliqué à de nombreux problèmes d’algèbre linéaire tels que les systèmes linéaires, les moindres carrés linéaires ... Il mesure l’effet sur la solution des petits changements des données. Cette quantité dépend des normes choisies. Dans ce travail, l’estimation statistique du conditionnement a été obtenue via la dérivée de Fréchet et avec la méthode appelée des petits échantillons.

Dans les perspectives, nous s’intéresserons à l’application des méthodes aléatoires sur les systèmes linéaires dans le cas où le rang de la matrice est incomplet ainsi qu’aux problèmes des moindres carrés non linéaires. Par la suite appliquer cette théorie dans le domaine de détermination des orbites des satellites.

Annexe 1

Cet annexe montre la formule de Wallis défini dans le chapitre 3 pour les grandes valeurs de n .

Cependant, le facteur de Wallis ω_n est défini par l'équation (III.15) pour les petites valeurs de n . Alors que pour les grandes valeurs de n on ne peut pas utiliser cette formule donc on va l'approximer.

nous avons à partir des équations (III.15)

$$\omega_n \omega_{n+1} = \frac{2}{\pi n}$$

ω_n décroissante alors $\omega_n > \omega_{n+1}$

Nous avons

$$\frac{2}{\pi n} = \omega_n \omega_{n+1} < \omega_n^2$$

On va faire la racine carré des deux coté, on obtient

$$\sqrt{\frac{2}{\pi n}} < \omega_n$$

ce qui veut dire

$$\frac{2}{\pi(n-1)} = \omega_{n-1} \omega_n > \omega_n^2$$

Ces deux formules implique la propriété suivante

$$\sqrt{\frac{2}{\pi n}} < \omega_n < \sqrt{\frac{2}{\pi(n-1)}}$$

Par la suite, on peut approximer la valeur du facteur de Wallis par

$$\omega_n \approx \sqrt{\frac{2}{\pi(n - \frac{1}{2})}}$$

Bibliographie

- [1] G. Austin, T. Reubelt, *Raumliche Schwerefeldanalyse aus semi- kontinuierlichen Ephemeriden niedrigfliegender GPS- vermessener Satelliten vom Typ CHAMP, GRACE und GOCE*, Diplomarbeit from university of Stuttgart, Geodatisches Institut, (2000).
- [2] M. Arioli, M. Baboulin, S. Gratton, *Partial condition numbers for linear least squares problems*, Technical report RAL-TR-2004-029, (2004).
- [3] M. Baboulin, S. Gratton, R. Lacroix, A. Laub, *Efficient computation of condition estimates for linear least squares problems*, Research Report in INRIA N 8065, (2012).
- [4] M. Baboulin, J. Dongarra, S. Gratton, J. Langou, *Computing the conditioning of the components of a linear least squares solution*, Numer. Linear Algebra Application **16**, pp 517-533, (2009).
- [5] M. Baboulin, S. Gratton, R. Lacroix, A. Laub, *Statistical estimates for the conditioning of linear least squares problems*. Proceedings of the 10th International Conference on Parallel Processing and Applied Mathematics, PPAM (2013).
- [6] Å. Björck, R. Lacroix, A. Laub, *Numerical Methods for Least Squares Problems*, SIAM, Philadelphia, (1996).
- [7] Å. Björck, *Numerics of Gram-Schmidt orthogonalization*, Linear Algebra Appl, pp297-316 (1994).
- [8] Å. Björck, *Numerical Methods in Matrix Computations*, Springer International Publishing **59**, (2015).

- [9] P. Drineas, M. W. Mahoney, S. Muthukrishnan, T. Sarlos, *Faster Least Squares Approximation*, Technical Report. Numerische Mathematik, **117** pp 219-249, (2011).
- [10] P. Drineas, M W. Mahoney, S. Muthukrishnan, T. Sarlos , *Faster Least Squares Approximation*, Cornell University Library, (2010).
- [11] P. Drineas, R.Kannan , M.W. Mahoney, *Fast Monte Carlo Algorithms for matrices I : Approximating Matrix Multiplication*, SIAM Journal on Computing **346**, pp 132-157, (2006),
- [12] A.J. Geurts, *A contibution to the theory of condition*, Numerische Mathematik **39** , pp 85-96,(1982).
- [13] S. Gratton, *On the condition number of linear least squares problems in a weighted frobenius norm*,BIT **36** pp 523-530, (1996).
- [14] A. Graham, *Kronecker products and matrix calculus with application*, Wiley, New York,(1981).
- [15] C.F. Gauss, *Theory of the motion of the Heavenly bodies moving about the sun in Conic sections* republished in 1963 by Dover, Mineola, NY (1809). C. H.Davis translation.
- [16] C.F. Gauss, *the theory of the combination of Observations least subject to errors*. part 1, Part 2. SIAM, Philadelphia, PA, G. W.Stewart, translation 1995 (1821).
- [17] C.F. Gauss, *Theoria combinationis observationum erroribus minimis obnoxiae, pars prior*. In : werke, IV(ed.) . Königlichen Gesellschaft der Wissenschaften zu Göttingen, pp.1-26, (1880).
- [18] C.F. Gauss, *Theoria combinationis observationum erroribus minimis obnoxiae, pars posterior*. In : werke, IV(ed.) . Königlichen Gesellschaft der Wissenschaften zu Göttingen, pp.27-53, (1880).
- [19] N.J. Higham, *Accuracy and stability of Numerical algorithms*. 2nd edn.SIAM, Philadelphia (2002).
- [20] P.C. Hansen, V. Pereyra, G. Scherer, *Least Squares data fitting with Applications*.The Johns Hopkins University Press, Baltimore(2013).

- [21] N.Halko, P. G. Martinsson, J. A. Tropp , *Finding Structure with Randomness :Probabilistic Algorithms for Constructing Approximate matrix Decompositions*, Siam review, **53** No.2 pp 217-288, (2011).
- [22] J. T. Holodnak, I. Ipsen *Randomized approximation of the Gram matrix : Exact computation and probabilistic bounds*, SIAM J. Matrix Anal. Appl. **36**, N 1, pp 110-137, (2015).
- [23] W. Heiskanen, H. Moritz, *Physical Geodesy*,W.H. Freeman and Company San Francisco, (1976).
- [24] D. J. Higham, N. J. Higham , *Matlab Guide*, second edition, SIAM, (2000).
- [25] W. B. Johnson and J. Lindenstrauss , *Extensions of Lipschitz mappings into a Hilbert space*, Contemporary Mathematics, **26 :1** , pp 189-206, (1984).
- [26] L. Kenneth, Clarkson, P. D. Woodruff, *Low rank approximation and regression in input sparsity time*,STOC'13 proceedings of the forty fifth annual ACM symposium pp 81-90, (2013).
- [27] C.S. Kenny, A.J Laub, *Small sample statistical condition estimates for general matrix*,SIAM J Sci.Comput, **15(1)**, pp 36-61, (1994).
- [28] C.S. Kenny, A.J Laub and M. S. Reese, *Statistical condition estimtion for linear least squares*,SIAM J. Matrix Analysis and applications, **19(4)** , pp 906-923, (1998).
- [29] G. Kaplan, J. Bartlett, A. Monet, J.Bangert, W. Puatua *User's guide to NOVAS Version F3.1*, Naval Observatory Vector Astrometry Software, Fortran Edition, U.S. Naval observatory, (2011).
- [30] M. W. Mahoney , *Randomized algorithms for matrices and data*, Foundations and Trends in Machine in Machine Learning series. **3(2)**, pp123-224, (2011).
- [31] X. Meng, *Randomized algorithms for large scale strongly over determined linear regression problems*,,a dissertation submitted to the institute for computational and mathematical engineering and the committee on graduate studies of Stanford University, (2014).
- [32] R. Motwani, P. Raghavan, *Randomized algorithms*, Cambridge University Press, New York, (1995).

-
- [33] X. Meng, *Randomized algorithms for large scale strongly over determined linear regression problems*,a dissertation submitted to Stanford University (2014).
 - [34] L. Kenneth, Clarkson, P. D. Woodruff, *LOW rank approximation and regression in input sparsity time*, (2013).
 - [35] G. Kaplan, J. Bartlett, A. Monet, J. Bangert, W. Puatua , *User's Guide to NOVAS Version F3.1 Naval Observatory Vector Astrometry Software*,Fortran Edition, U.S. Naval Observatory, (2011).
 - [36] N. Rabehi et M. Terbeche, *Fast randomized linear least squares*. Journal of Statistics and Management Systems,Taylor and Francis, **21**, No. 1, pp. 107-123, (2018)
 - [37] V. Rokhlin and M. Tygert, *A fast randomized algorithm for overdetermined linear least-squares regression*, Technical Report Proc Natl Acad Sci USA **105(36) : 13212-7** ,(2008).
 - [38] T. Sarlos, *Improved approximation algorithms for large matrices via random projections*, Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science, FOCS06, IEEE Computer Society,Washington, DC, USA, pp143-152, (2006).
 - [39] W. Shusen , *A Practical Guide to Randomized Matrix Computations with MATLAB Implementations*, CORR,abs, 1505.07570 (2015).
 - [40] S. Ubaru, A. Mazumdar and Y. Saad, *Low Rank Approximation using Error Correcting Coding Matrices*, Proceedings of the 32nd International Conference on Machine Learning, JMLR : W&CP, **37** (2015).
 - [41] S. Ubaru, *Randomized techniques for matrix decomposition and estimating the approximate rank of a matrix*,A thesis submitted to the university of Minnesota, (2014).
 - [42] D. P. Woodruff, *Sketching as a Tool for Numerical Linear Algebra*,Foundations and Trends in Theoretical Computer Science. NOW Publishers, Boston, 2014.
 - [43] J.Yang , X. Meng, and M.W. Mahoney, *Implementing randomized matrix algorithms in parallel and distributed environments*, Available at arXiv :1502.03032 (2015).

Journal of Statistics and Management Systems

Fast randomized linear least squares

Nadjet Rabehi & Mekki Terbeche

To cite this article: Nadjet Rabehi & Mekki Terbeche (2018) Fast randomized linear least squares, *Journal of Statistics and Management Systems*, 21:1, 107-123

To link to this article: <https://doi.org/10.1080/09720510.2017.1367481>



Published online: 13 Feb 2018.



Submit your article to this journal



View related articles



View Crossmark data

Fast randomized linear least squares

Nadjet Rabehi *

*Department of Mathematics
University of Oran 1
Ahmed Ben Bella
Oran
Algeria*

and

*Centre de Développement des Satellites
PB 4065, Ibn Rochd USTO Bir el Djir
Oran 31130
Algeria*

Mekki Terbeche [†]

*Department of Mathematics
University of Oran 1
Ahmed Ben Bella
Oran
Algeria*

Abstract

Least squares approximation is a technique for finding an approximate solution to a system of linear equations that has no exact solution. In this paper, we are concerned with fast randomized approximation of Linear Least Squares (LLS) problems. Two algorithms are presented and compared: the first employing the combination of count-sketch with Subsampled Randomized Hadamard transform and in the second we combine count-sketch with a Gaussian projection. Both algorithms make use of QR factorization. The condition number of randomized LLS is computed. Finally, an application in space geodesy is presented for computing the geopotential harmonic coefficients with the aim of computing the gravitational potential.

Subject Classification: 93E24

^{*}E-mail: rabezin@gmail.com

[†]E-mail: mekki.terbeche@gmail.com

Keywords: Dense linear least squares, Randomized linear algebra, Randomized Johnson–Lindenstrauss lemma, Randomized algorithm, Condition number, Geopotential harmonic coefficient

1. Introduction

The resolution of large dense linear systems appears in many engineering and scientific applications of computational sciences such as in data mining, electromagnetic applications, geophysics and geodesy.

Least-squares is a fundamental technique in linear algebra which is often used in such applications. In the last, they used the least-squares techniques for the analysis of survey data and in astronomical calculations where the problem is formulated as the minimization of the sum of squares of the elements of the residual vector.[21, 20]

Least squares problems can be solved by various methods including the normal equations, Householder triangularization and Singular Value Decomposition (SVD), each method having particular advantages in certain situations.

A classical way of solving a LLS problem is the method of normal equations: it has the advantage of being easy to code and is faster than other direct methods. This explains why it is often favoured by statisticians. However, solving the normal equations is not always stable in the presence of rounding errors, and thus for many years, numerical analysts have instead recommended QR factorization as the standard method for LLS problems [17]. However, if A is close to rank deficient, it turns out that QR factorization has less than ideal stability properties, and there are good reasons to turn to the SVD method [19]. The SVD approach is both accurate and robust to rank-deficiency. The QR approach is less expensive than SVD, but it is slightly less robust at determining the rank. A comprehensive survey of the methods available for solving LLS problems is given in [6, 17]. Among the disadvantages of the least squares method is the time required for the resolution of the problem especially when using large matrices. Randomized, approximations of LLS problems, involving the use of a sketching matrix, permit much more rapid solutions to be obtained than via deterministic or conventional methods. Randomization is especially useful for generating simpler and faster algorithms. The sketching methods can be applied to sparse matrices as well, and in this case, iterative methods are often preferable. We note here that we are interested only to direct methods. The most straightforward use of these methods is to construct a subspace preserving the embedding matrix and then to

solve the resulting reduced-size problem exactly, thereby obtaining an approximate solution to the original problem.

Consider the LLS problem $\min_{x \in \mathbb{R}^d} \|Ax - b\|_2$ where $A \in \mathbb{R}^{n \times d}$, $d \ll n$ and $b \in \mathbb{R}^n$. If A has full column rank, x simplifies to $x = (A^t A)^{-1} A^t b$; this minimizer can be computed deterministically in $O(nd^2)$ time, but with randomization and approximation, this problem can be solved in $O\left(nd \log(d) + \text{poly}\left(\frac{d}{\varepsilon}\right)\right)$ time, which is much faster for $d \ll n$ and ε not too small. In addition to compute LLS solutions efficiently, an important issue is to assess the numerical quality of the computed solution. In this context, the notion of the condition number is recommended. The condition number offers a theoretical framework that is used to measure the numerical sensitivity of a problem's solution to perturbations in its data and thereby to assess the numerical quality of a computed solution. Many results on normwise LLS condition number have been published in [2, 3, 6, 8, 17].

This paper is organised as follows: In section 2, a brief review is presented of properties of the sketching matrix, subspace embedding and different methods of randomization of least squares problem, all of which play an essential role throughout our study. In section 3, two algorithms are presented which enable the rapid computation of a LLS solution. The first algorithm combines the count-sketch method with the Subsampled Randomized Hadamard Transform (SRHT) whereas the second algorithm makes use of Gaussian projection in combination with the count-sketch method; both algorithms make use of QR factorization. In section 4, a review is given of the condition number of randomized LLS with the objective of efficiently computing LLS solution. In section 5, a numerical experiment in the area of space geodesy is presented the two algorithms are applied to the computation of the harmonic geopotential coefficients for the determination of the gravitational potential V . The paper is concluded with some remarks and perspectives in section 6.

2. Preliminaries

First we must go through some definitions.

Definition 1: (Subspace embedding) Given a matrix $b \in \mathbb{R}^{n \times d}$, $p \in [1, \infty]$ and $\varepsilon \in (0, 1)$, $\phi \in \mathbb{R}^{s \times n}$ is a subspace embedding of Ω_p , if $s = O(\text{poly}(d))$ independent of n , and

$$(1 - \varepsilon) \|y\|_p \leq \|\phi y\|_p \leq (1 + \varepsilon) \|y\|_p, \forall y \in \Omega_p \quad (1)$$

Note that a subspace embedding of Ω into \mathbb{R}^s with $s > 0$ is a structure preserving mapping $\phi: \Omega \rightarrow \mathbb{R}^s$, where the meaning of structure preserving varies depending on the application. We are interested to linear embeddings of the normed vector space $\Omega_p = (\Omega, \|\cdot\|_p)$, the subspace Ω paired with the l_p norm $\|\cdot\|_p$.

Based on properties of the subspace embedding, we can define the l_2 regression via subspace embedding which is used in randomized linear algebra and has been provided by Woodruff [14].

Let matrix $A \in \mathbb{R}^{n \times d}$, $d \ll n$. Any sketching matrix S can be used to solve a LLS problem approximately as long as it satisfies the subspace embedding property and $C = SA \in \mathbb{R}^{n \times s}$ be a sketch of A . The size of C is much smaller than A .

Definition 2: Let $S \in \mathbb{R}^{s \times n}$ be any subspace embedding matrix and $\tilde{x} = \operatorname{argmin}_x \|SAx - Sb\|_2$

$$\text{then } \|SAx - Sb\|_2 \leq (1 + \varepsilon) \operatorname{argmin}_x \|Ax - b\|_2 \quad (2)$$

The objective of this theory is to find a solver x that minimizes $\|Ax - b\|_2$, but with randomization we can find a minimizer vector x' with high probability so that

$$\|Ax' - b\|_2 \leq (1 + \varepsilon) \operatorname{min}_x \|Ax - b\|_2 \quad (3)$$

Using an l_2 subspace embedding, one can work with $SA \in \mathbb{R}^{s \times d}$ instead of $A \in \mathbb{R}^{n \times d}$, $s \ll n$, so we are working with a smaller matrix that reduces the time, space complexity of many algorithms.

In some more detail, this approach performs the following two steps:

1. Construct a subspace preserving embedding matrix S with distortion $1 \pm \frac{\varepsilon}{4}$,
2. Solve the reduced sized LLS problem

$$\tilde{x} = \min_x \|SAx - Sb\|_2 \quad (4)$$

which can be solved in $O(sd^2)$ time. Thereby obtaining an approximate solution to the original problem.

Sketching is a useful and important tool in randomized matrix algorithms. It can be used in two ways projection or sampled. We can find in [18] a review and history of randomized method.

Random Projection is one category of methods used for least-squares approximation. A random projection algorithm projects data points from a high-dimensional space \mathbb{R}^n onto a lower-dimensional subspace \mathbb{R}^d ($d \ll n$)

using a random matrix S . Random projection methods are computationally efficient and sufficiently accurate in practice for dimensionality reduction of high-dimensional data-sets. Moreover, since the complexity of many geometric algorithms depends significantly on the dimension, applying random projection as pre-processing is a common task in many data-mining applications. As opposed to column-sampling methods, which need to access data for approximating. The key idea of random mapping comes from the Johnson-Lindenstrauss Lemma [7].

Remark 1: Johnson- Lindenstrauss Lemma [16] Given $\varepsilon \in (0, 1)$, a point set χ of N points in \mathbb{R}^d , there is a linear map $\phi: \mathbb{R}^m \rightarrow \mathbb{R}^s$ with $s = C \log \frac{N}{\varepsilon^2}$, where $C > 0$ is a global constant, such that

$$(1 - \varepsilon) \|x - y\|^2 \leq \|\phi(x) - \phi(y)\|^2 \leq (1 + \varepsilon) \|x - y\|^2, \forall x, y \in \chi \quad (5)$$

We say a mapping has Johnson- Lindenstrauss property if it satisfies the above condition with a constant probability.

The Johnson-Lindenstrauss transform (JLT) is defined as: let $\varepsilon > 0$, $\delta < 1$ and an matrix S formed by $S = \frac{1}{\sqrt{s}} G \in \mathbb{R}^{s \times n}$, such that $G_{ij} \sim N(0, 1)$ entries are independent normal random variables. If $s = \Omega\left(\varepsilon^{-2} \log d \log\left(\frac{1}{\delta}\right)\right)$ then S is a $JLT(\varepsilon, \delta, d)$. This projection is easy to implement and makes a very high quality sketch. It also well known as the transform Gaussian random projection [11]. For practical applications the $N(0, 1)$ entries can be replaced by random ± 1 variables.

Ailon and Chazelle [15] construct the so-called fast Johnson-Lindenstrauss transform (FJLT). Which is a product of three matrices $S = PHD$, where $P \in \mathbb{R}^{s \times m}$ is a sparse JL transform with approximately $O(s \log^2 N)$ non zeros; $H \in \mathbb{R}^{m \times m}$ is a normalized Walsh Hadamard matrix and $D \in \mathbb{R}^{m \times m}$ is a diagonal matrix with its diagonals drawn independently from $\{-1, \pm 1\}$ with probability $\frac{1}{2}$. This projection reduces time from $O(sm)$ to $O(m \log m)$.

A Subsampled Randomized Hadamard Transform (SRHT) is an $s \times n$ matrix of the form $S = \sqrt{\frac{n}{s}} RHD$, where $D \in \mathbb{R}^{n \times n}$ is a diagonal matrix whose entries are independent random signs, $H \in \mathbb{R}^{n \times n}$ is a Walsh Hadamard matrix. We set $H = \frac{1}{\sqrt{n}} H_n$, where H_n is defined by $H_n = \begin{bmatrix} H_{\frac{n}{2}} & H_{\frac{n}{2}} \\ H_{\frac{n}{2}} & -H_{\frac{n}{2}} \end{bmatrix}$, and

$H_2 = \begin{bmatrix} +1 & +1 \\ +1 & -1 \end{bmatrix}$, $R \in \mathbb{R}^{s \times n}$ restricts an n -dimensional vector to s coordinates.

Note that applying SRHT takes $O(nd \log s)$ time complexity, which makes SRHT more efficient than Gaussian projection. Noted that applying the randomized Hadamard transform to the matrix A and vector b only takes $O(nd \ln d)$ time.

The Leverage Score Sampling of a matrix A is the squared norms of the matrix containing its right singular vectors. It is to select each column of A with probability proportional to its leverage scores [11]. For $A \in \mathbb{R}^{n \times s}$ it was shown in [9] that for any constant $\varepsilon > 0$, and with probability at least $\frac{2}{3}$, outputs vectors u'_i to the leverage score u_i for all $i \in [n]$: $u'_i = (1 \pm \varepsilon)u_i$. The running time is $O(nnz(A)\log n + r^3 \log^2 r + r^2 \log n)$, where r is the rank of A .

Computing the leverage scores is as expensive as computing SVD, so it is not a practical way to sketch the matrix A itself. In [15] we can find a review of this technique and a numerical experiment.

The theoretical properties of count-sketch are established by Clarkson and Woodruff [9], who applied it to speed-up matrix computation. There are different ways to implementing count-sketch: map-reduce fashion and streaming fashion. The count-sketch can be written as the product of two matrices $F = SD \in \mathbb{R}^{s \times n}$, where $S \in \mathbb{R}^{s \times n}$ has each column chosen independently and uniformly from the s standard basis vectors of \mathbb{R}^s and $D \in \mathbb{R}^{n \times n}$ is a diagonal matrix with diagonal entries chosen independently and uniformly from ± 1 .

The count-sketch method is very efficient; especially when A is sparse and compared with Gaussian projection it requires larger s to attain the same accuracy. The total time cost of computing the count-sketch of a matrix is $O(nnz(A) + \text{poly}\left(\frac{d}{\varepsilon}\right))$ and the memory cost is $O\left(\text{poly}\left(\frac{d}{\varepsilon}\right)\right)$, which are both lower than the cost of exact LLS when $d \ll n$, nnz is the number of non-zero entries of A .

If S is a Gaussian projection, SRHT, count-sketch, or leverage score sampling matrix, and $s = \text{poly}\left(\frac{d}{\varepsilon}\right)$ for any error parameter $\varepsilon \in]0, 1]$, then

$$\|A\tilde{x} - b\|_2^2 \leq (1 + \varepsilon)^2 \min_x \|Ax - b\|_2^2 \quad (6)$$

is guaranteed [11]

By the subspace embedding property, it can be easily shown that \tilde{x} is a low-precision solution. We can find in [15] a table that defines a summary of data aware l_2 embeddings. Recall also that, one can obtain a high-precision solution by invoking an iterative algorithm. The principle of a methodology is to construct a subspace-preserving embedding matrix

and use that to construct a preconditioner for the original l_2 regression problem, thereby obtaining an approximate solution to the original problem.

Alternatively, for the LS problem, instead of solving the original problem $\min_x \|Ax - b\|_2$, we compute a randomized preconditioner and then to apply one of these iterative solvers on the preconditioned system.

3. Algorithm Description

One simple improvement is to combine firstly the count-sketch with SRHT and after with Gaussian Projection.

3.1 Description of Algorithm 1

Consider the LLS problem

$$\min_{x \in \mathbb{R}^d} \|Ax - b\|_2 \quad (7)$$

Where $A \in \mathbb{R}^{n \times d}$, $d \ll n$, $b \in \mathbb{R}^n$. Let $r = \text{rank}(A) = d$, the problem has full column rank. Since $d \ll n$ this problem can be solved more efficiently by sketching.

We draw a sketching matrix $S \in \mathbb{R}^{n \times s_2}$ such $D = S^t A$ and we compute the approximate solution to

$$\tilde{x} = \min_{x \in \mathbb{R}^d} \|S^t Ax - S^t b\|_2 \quad (8)$$

In this part we combine firstly SRHT with count-sketch and in second part the Gaussian projection with count-sketch.

Woodruff and al.[9] showed that using the combination SRHT and count-sketch S with $s_1 = O\left(s^2 \log^6 \frac{s}{\varepsilon} + \frac{s}{\varepsilon}\right)$ and $s_2 = O\left(\frac{s}{\varepsilon} \log \frac{s}{\varepsilon}\right)$ ensures $(1 + \varepsilon)$ relative error. For the value of s we can choose an oversampling factor $\gamma > 1$, Set $s = \gamma d$. [15]

So we draw a $n \times s_1$ count-sketch matrix S_{CS} and an $s_1 \times s_2$ matrix S_{SRHT} . Set the parameters as: $s_1 = s^2 \log^6 \frac{s}{\varepsilon} + \frac{s}{\varepsilon}$ and $s_2 = \frac{s}{\varepsilon} \log \frac{s}{\varepsilon}$; such that $S = S_{CS} S_{SRHT}$.

Accordingly

$$D = S_{SRHT}^t S_{CS}^t A \in \mathbb{R}^{s_2 \times d} \quad \text{and} \quad B = S_{SRHT}^t S_{CS}^t b \in \mathbb{R}^{s_2} \quad (9)$$

We calculate the QR decomposition of D , such that $D = Q_D R_D$, R_D is approximately find in $O\left(nnz(A) + \text{poly}\left(\frac{d}{\varepsilon}\right)\right)$ time. So to compute the $\min_{x \in \mathbb{R}^d} \|Ax - b\|_2$ we simplify the calculation to

$$\min_{x \in \mathbb{R}^d} \|Dx - b\|_2 = R_D^\dagger Q_D^t B \quad (10)$$

Where R_D^\dagger, Q_D^t are respectively the pseudo inverse of R and the transpose of Q_D computed by the QR decomposition of $D = SA$.

Algorithm 1 Fast SRHT CS

Input: $A \in \mathbb{R}^{n \times d}$, A is full column rank and $b \in \mathbb{R}^n$.

Output: \tilde{x} , the solution of problem

1. Set $s = \gamma d$, $\gamma > 1$
2. Set the parameters as: $s_1 = s^2 \log^6 \frac{s}{\varepsilon} + \frac{s}{\varepsilon}$, and $s_2 = \frac{s}{\varepsilon} \log \frac{s}{\varepsilon}$;
3. Draw a $n \times s_1$ count sketch matrix S_{CS} and an $s_1 \times s_2$ matrix S_{SRHT} ;
4. Perform Sketching: $D = S_{SRHT}^t S_{CS}^t A \in \mathbb{R}^{s_2 \times d}$ and $B = S_{SRHT}^t S_{CS}^t b \in \mathbb{R}^{s_2}$;
5. QR Decomposition: $[Q_D, R_D] = qr(D)$;
6. $\tilde{x} = R_D^\dagger Q_D^t B$

The combination of SRHT with count-sketch can be computed as efficiently as the count-sketch and the size is near the same to the size of SRHT. Theoretically, the matrix S produced by the combination of SRHT and count-sketch satisfy the subspace embedding property with probability 0,99.

3.2 Description of Algorithm 2

Consider the LLS problem

$$\min_{x \in \mathbb{R}^d} \|Ax - b\|_2 \quad (11)$$

Where $A \in \mathbb{R}^{n \times d}$, $d \ll n$, $b \in \mathbb{R}^n$. Let $r = \text{rank}(A) = d$, the problem has full column rank. Since $d \ll n$ this problem can be solved more efficiently by sketching.

We draw a sketching matrix $S \in \mathbb{R}^{n \times s_2}$ such $D = S^t A$ and we compute the approximate solution to

$$\tilde{x} = \min_{x \in \mathbb{R}^d} \|S^t Ax - S^t b\|_2 \quad (12)$$

The same procedure used before with the combination of SRHT and count-sketch. In this case we will combine Gaussian Projection and count-sketch.

So we draw a $n \times s_1$ count sketch matrix S_{CS} and an $s_1 \times s_2$ Gaussian Projection matrix S_{GP} . Set the parameters as: $s_1 = \frac{s}{\varepsilon} \log \frac{s}{\varepsilon}$ and $s_2 = \frac{s}{\varepsilon} \log \frac{s}{\varepsilon}$; such that $S = S_{CS}S_{GP}$. For the value of s we can choose an oversampling factor $\gamma > 1$, Set $s = \gamma d$ [15] Accordingly

$$D = S_{GP}^t S_{CS}^t A \in \mathbb{R}^{s_2 \times d} \text{ and } B = S_{GP}^t S_{CS}^t b \in \mathbb{R}^{s_2} \quad (13)$$

Finally, to compute the $\min_{x \in \mathbb{R}^d} \|Ax - b\|_2$ we suffice calculate

$$\min_{x \in \mathbb{R}^d} \|Dx - B\|_2 = R_D^\dagger Q_D^t B \quad (14)$$

Where R_D^\dagger, Q_D^t are respectively the pseudo inverse of R and the transpose of Q_D computed by the QR decomposition of $D = SA$.

Algorithm 2 Fast GP CS

Input: $A \in \mathbb{R}^{n \times d}$, A is full column rank and $b \in \mathbb{R}^n$.

Output: \tilde{x} , the solution of problem

1. Set $s = \gamma d$, $\gamma > 1$
2. Set the parameters as: $s_1 = \frac{s}{\varepsilon} \log \frac{s}{\varepsilon}$, and $s_2 = \frac{s}{\varepsilon} \log \frac{s}{\varepsilon}$;
3. Draw a $n \times s_1$ count sketch matrix S_{CS} and an $s_1 \times s_2$ matrix S_{GP} ;
4. Perform Sketching: $D = S_{GP}^t S_{CS}^t A \in \mathbb{R}^{s_2 \times d}$ and $B = S_{GP}^t S_{CS}^t b \in \mathbb{R}^{s_2}$;
5. QR Decomposition: $[Q_D, R_D] = qr(D)$;
6. $\tilde{x} = R_D^\dagger Q_D^t B$

So, here we give some advantages to combine Gaussian Projection with count-sketch [11]: the size of Gaussian Projection with count-sketch is as small as Gaussian projection, the time complexity is much lower than Gaussian projection when $d \ll n$, and the implementation is easier than SRHT with count-sketch.

4. Condition number of least squares problem

The standard approaches to solving an over-determined linear system $Ax = b$ find minimal corrections to the vector b and or the matrix A such that the corrected system is consistent, such as the least squares method.

The linear least squares problem [6,17] corresponds to

$$\min_{x \in \mathbb{R}^d} \|Ax - b\|_2 \quad (15)$$

With $A \in \mathbb{R}^{n \times d}$, $d \ll n$ and $b \in \mathbb{R}^n$. Assuming that A is full column rank, the problem consists of finding the solution $x \in \mathbb{R}^d$, such that

$$x = (A^t A)^{-1} A^t b = A^+ b \quad (16)$$

Where A^+ denotes the pseudo-inverse of the matrix A .

In this part of paper, we would like to find the condition number of randomized LLS solution.

For this, we consider

$$\begin{aligned} g: \mathbb{R}^{s_2 \times d} \times \mathbb{R}^{s_2} &\rightarrow \mathbb{R}^d \\ (D, B) &\mapsto g(D, B) = L^t x = L^t (D^t D)^{-1} D^t B \end{aligned} \quad (17)$$

Where $D = S_i^t S_{CS}^t A \in \mathbb{R}^{s_2 \times d}$ and $B = S_i^t S_{CS}^t b \in \mathbb{R}^{s_2}$; where $S_i \in \mathbb{R}^{s_1 \times s_2}$ a sketching matrix defined before in the two algorithms 1 and 2 given in the preceding section. For $i = 1$ we put the SRHT matrix and for $i = 2$ we make the Gaussian Projection. If the solution x for a given problem can be expressed as a function $g(y)$ of a data y , then if g is differentiable the condition number of g at y can be defined as [8]

$$k(y) = \max_{z \neq 0} \frac{\|g'(y).z\|}{\|z\|} \quad (18)$$

We consider the product norm defined on $\mathbb{R}^{s_2 \times d} \times \mathbb{R}^{s_2}$ by

$$\|(D, B)\|_F = \sqrt{\|D\|_F^2 + \|B\|_2^2} \quad (19)$$

Frobenius norm for the data space that has been mentioned in [10] to be appropriate for estimating the forward error obtained when the LLS problem is solved via normal equation, and for its flexibility and the Euclidian norm for the solution space \mathbb{R}^{s_2} . In this part of paper, we consider the problem of calculating the quantity $L^t x$, where x is the solution of the linear least squares problem. Since g is a Fréchet differentiable function of D and B , $(D^t D)^{-1}$ is also Fréchet differentiable in the neighborhood of (D, B) then g is a Fréchet derivative in the neighborhood of (D, B) .

According to [8] the condition number of g at the point (D, B) is given by

$$k(L, D, B) = \max_{(\Delta D, \Delta B) \neq 0} \frac{\|g'(D, B)(\Delta D, \Delta B)\|_2}{\|(\Delta D, \Delta B)\|_F} \quad (20)$$

$k(L, D, B)$ is sometimes called the absolute condition number of $L^t x$. The explicit formula for least squares condition number $k(D, B)$ (defined by (17) and according to [3]) is given by

$$k(L, D, B) = \sqrt{(1 + \|x\|_2^2) \|L^t(D^t D)^{-1} D^t\|_2^2 + \|r\|_2^2 \|L^t(D^t D)^{-1}\|_2^2} \quad (21)$$

Where $r = Dx - B$.

In the most cases, L is the identity matrix, this corresponds to the condition number of least squares solution, it is expressed by [3].

$$k(I, D, B) = \sqrt{(1 + \|x\|_2^2) \|(D^t D)^{-1} D^t\|_2^2 + \|r\|_2^2 \|(D^t D)^{-1}\|_2^2} \quad (22)$$

And it can be for instance a canonical vector of \mathbb{R}^d if we are interested to the conditioning of one component of x

$$k_i = \sqrt{(1 + \|x\|_2^2) \|e_i^t(D^t D)^{-1} D^t\|_2^2 + \|r\|_2^2 \|e_i^t(D^t D)^{-1}\|_2^2} \quad (23)$$

If we also assume that the LLS has already been solved with QR factorization when the R factor of the QR factorization of D is available.

With substitute: $\|L^t D^t\|_2 = \|R_D^{-t} L\|_2$, $\|L^t(D^t D)^{-1}\|_2 = \|R_D^{-1}(R_D^{-t} L)\|_2$ in equation (21) and considering $L = I$ identity matrix, we obtain the closed formula for the condition number of the randomized LLS solution

$$k_{LS} = \sqrt{\|R_D^{-1}\|_2 (\|R_D^{-1}\|_2^2 \|r\|_2^2 + \|x\|_2^2 + 1)} \quad (24)$$

5. Numerical investigations

An important aim in space geodesy is the computation of an accurate model of the Earth's gravity field and geoid that will have application in many scientific areas; it corresponds respectively to a particular surface of equal potential [4]. It is used to define physical altitude and to forecast water circulation that enables us to study ocean circulation.

This is the case in geodesy where the calculation of the gravity field requires the solution of large linear least-squares problems that are often solved using the normal equations approach. The objective of this application is the calculation as accurately as possible of normalized harmonic coefficients \bar{C}_{lm} and \bar{S}_{lm} for the determination of the gravitational potential V . [4]

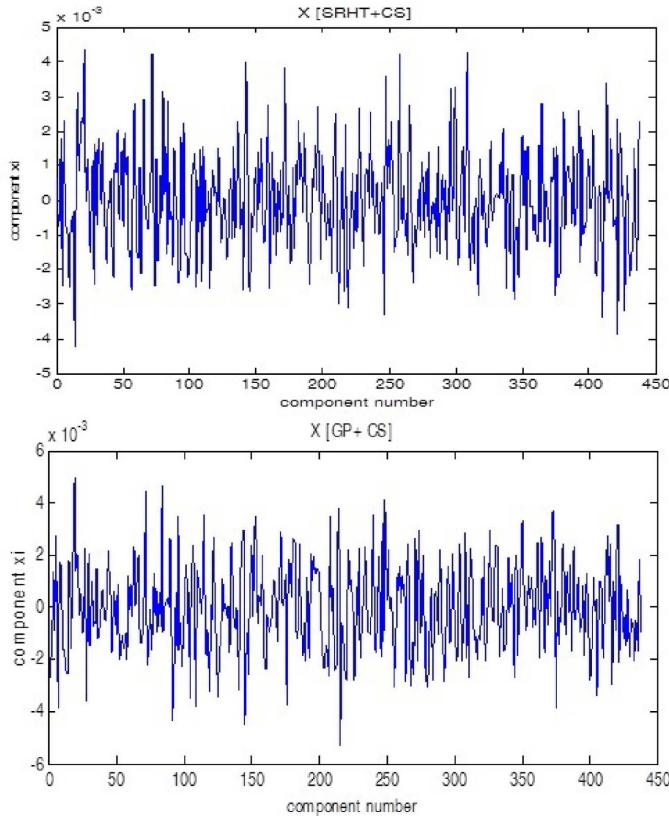
$$V(r, \theta, \lambda) = \frac{GM}{R} \sum_{i=0}^n \left(\frac{R}{r} \right)^{l+1} \sum_{m=0}^l \bar{P}_{lm}(\cos \theta) [\bar{C}_{lm} \cos m\lambda + \bar{S}_{lm} \sin m\lambda] \quad (25)$$

Where r , θ , λ geocentric radius, latitude and longitude respectively, GM is the product of the gravitational constant and Earth mass, R is the earth reference radius, \bar{P}_{lm} are the fully normalized Legendre functions of degree l and order m . For the study, we consider 30 days of observations simulated with a step of $10''$ from GOCE (Gravity field and steady state

Ocean Circulation Explorer) satellite mission; it provides a model of the Earth's gravity field and of the geoid with an unprecedented accuracy [1]. Measurements, which are simulated from a satellite position, velocity and acceleration, obtained via ground stations (Laser, Doppler) or other satellites (GPS).

The total number of observations is $m = 777603$; the number of spherical harmonic coefficients to be computed (which is also the number of unknown parameters n' in our problem) is a function of the degree of the geopotential model l [1], it is computed via the formula $n' = l^2 - 2l + 2$.

In this application, we consider $l = 20$ and therefore we have 438 unknowns. The first step in the computation of gravity field parameters is the rotation of the simulated data from Celestial to Terrestrial system, for this we have used the subroutine TERCEL from NOVAS [13]. In the second step, we have written a Matlab program which computes the configuration matrix A based on the theory of AUSTIN and REUBELT



Contd...

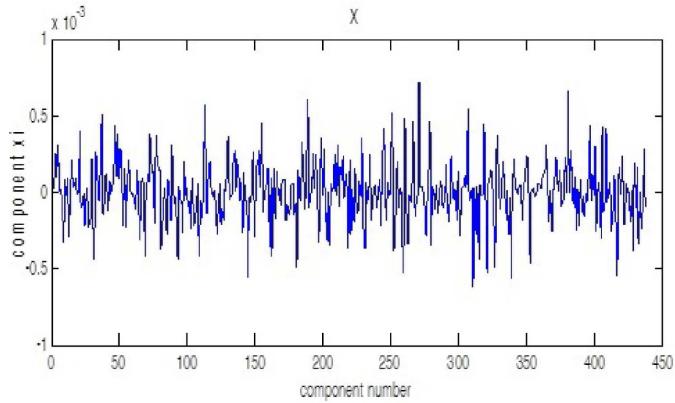


Figure 1

Variation of the components of the solution of LLS x_i in different cases

[1] and which goes on to apply the two different algorithms (1 and 2) proposed previously in section 3 to solve the problem.

In figure 1 we have the variation of component of the solution of linear least squares x . The first figure indicate x_i with the combination of count-sketch and SRHT, the second is count-sketch and Gaussian projection and the third x_i found with the normal-equation method without using randomization.

Thereafter, another Matlab program [12] computes different absolute condition number of the components x_i of LLS for the basic case where

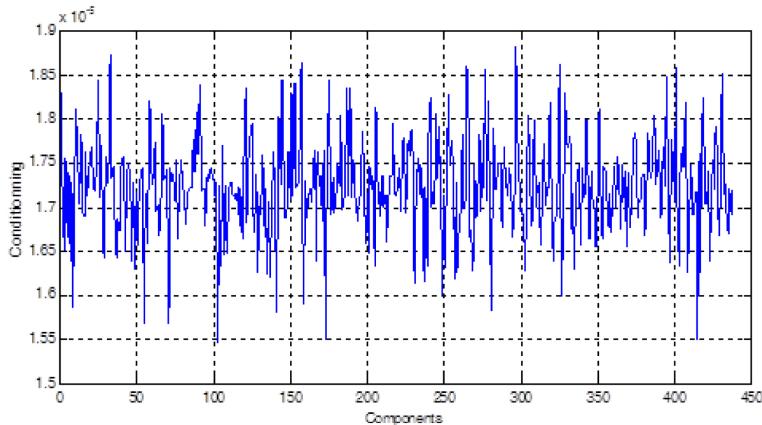


Figure 2

Condition number of the component x_i for linear least squares problem

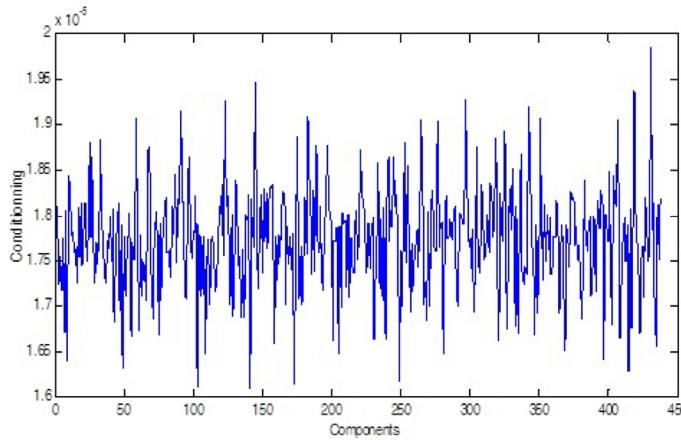


Figure 3

Variation of the condition number of the component x_i computed with the combination of Countsketch and SRHT

we don't use randomization, subsequently we include the combination of SRHT sampling and Gaussian Projection with count-sketch.

Figure 2 corresponds to the condition number of each coefficient x_i , for the case where randomization is not used.

Figure 3 correspond to the condition number of components x_i computed by the combination of count-sketch and SRHT (algorithm 1):

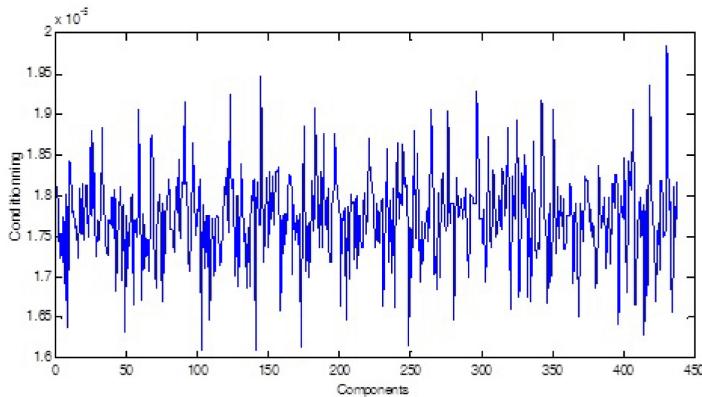


Figure 4

Variation of the condition number of the component x_i computed with the combination of count-sketch and Gaussian projection

we remark that the condition number varies in the order of 10^{-5} , noted that the absolute condition number in this case equal 2.59×10^{-5} .

Figure 4 define the condition number of components x_i computed by the combination of count-sketch and Gaussian projection (algorithm 2): we remark that the condition number again varies in the order of 10^{-5} , and note that the absolute condition number in this case equal 2.79×10^{-5} .

6. Conclusion

Linear least squares is a computational problem of primary importance, which originally arose from the need to fit a linear mathematical model to given observations. It minimizes the sum of squared residuals.

Application of least squares is important in many areas of applied and engineering research, such as statistic, geodetic, photogrammetric, signal processing and control.

The request for faster least squares solvers will continue to grow in view of new data applications and as problem scales become larger. A new challenges are exposed to modern large scale data analysis.

Recent years have seen great interest in randomized linear least squares problem that exploits randomization as a computational resource for the developpement of algorithms.

The use of subspace embedding, sketching and randomization theory drive us to faster and stable solution. The solution of a sketched LLS problem has been shown to be a good approximation to the solution of the LLS problem if we compare the different condition number of the components of the solution of LLS, in addition to the gain of running time.

Condition number is now considered as fundamental for sensitivity analysis. They have been applied to many problems of linear algebra such as linear least squares, linear systems. Condition number of linear least squares problem is considered as fundamental tools since they measure the effect on the solution of small changes in the data, it gives an indication of the accuracy of the results from matrix inversion and the solution of linear equation. This quantity depends on the norms chosen for measuring perturbations.

In this paper, we have proposed two algorithms in the goal to fast the compute of a solution of LLS when the matrix input is dense and large. There are faster than deterministic theories of resolution of large least-squares problems. In these two algorithms we combine SRHT with count-sketch and in the second the Gaussian projection with count-sketch. The choice between the different methods of randomization of LLS will depend

on finding a minimizer of a problem with high probabilities. We have focuses on very overdetermined very large-scale l_2 regression problems. Recall that the subspace embedding that is a crucial part of Randomized Linear Algebra algorithms can be data-aware. In the other words, after obtaining a subspace embedding matrix, one can obtain a low-precision solution by solving the resulting subproblem, or one can obtain a high-precision solution by invoking an iterative solver for l_2 regression, with a preconditioner constructed by the embedding. In the perspective, we envisage interesting to the study of the nonlinear least squares problem.

References

- [1] G. Austin, T. Reubelt, Raumliche Schwerfeldanalyse aus semi- kontinuierlichen Ephemeriden niedrigfliegender GPS- vermessener Satelliten vom Typ CHAMP, GRACE und GOCE, Diplomarbeit from university of Stuttgart, Geodatisches Institut, 2000.
- [2] M. Arioli, M. Baboulin, S. Gratton, Partial condition numbers for linear least squares problems, Technical report RAL-TR-2004-029, November 2004.
- [3] M. Baboulin, J. Dongarra, S. Gratton, J. Langou, Computing the conditioning of the components of a linear least squares solution, Numer. Linear Algebra Application, 16 (2009), 517-533.
- [4] W. Heiskanen, H. Moritz, Physical Geodesy, W.H. Freeman and Company San Francisco, 1976.
- [5] M. Baboulin, S. Gratton, R. Lacroix, A. Laub, Efficient computation of condition estimates for linear least squares problems, Research Report in INRIA N 8065, September 2012.
- [6] A. Björck, R. Lacroix, A. Laub, Numerical Methods for Least Squares Problems, SIAM, Philadelphia, 1996.
- [7] William B. Johnson and J. Lindenstrauss, Extensions of lipschitz mappings into a hilbert space. Contemporary mathematics, 26(189-206):1, 1984.
- [8] A.J. Geurts, A contibution to the theory of condition, Numerische Mathematik 39 (1982), 85-96
- [9] L. Kenneth, Clarkson, P. D. Woodruff, LOW rank approximation and regression in input sparsity time, april 8, 2013.
- [10] S. Gratton, On the condition number of linear least squares problems in a weighted frobenius norm, BIT 36 (1996), 523-530.

- [11] W. Shusen , A Practical Guide to Randomized Matrix Computations with MATLAB Implementations, November 4, 2015.
- [12] D. J. Higham, N. J. Higham , Matlab Guide, second edition, SIAM, 2000.
- [13] G. Kaplan, J. Bartlett, A. Monet, J. Bangert, W. Puatua , User's Guide to NOVAS Version F3.1–Naval Observatory Vector Astrometry Software, Fortran Edition, U.S. Naval Observatory, March 2011.
- [14] D. P. Woodruff, Sketching as a Tool for Numerical Linear Algebra, Foundations and Trends in Theoretical Computer Science. NOW Publishers, Boston, 2014.
- [15] Yang, Jiyang, Meng, Xiangrui, and Mahoney, Michael W , Implementing randomized matrix algorithms in parallel and distributed environments, To Appear in Proceedings of the IEEE, Available at arXiv:1502.03032, 2015.
- [16] W. B. Johnson and J. Lindenstrauss , Extensions of Lipschitz mappings into a Hilbert space, *Contemporary Mathematics*, 26(189-206):1, 1984.
- [17] A. Björck , Numerical Methods in Matrix Computations, Texts in Applied Mathematics volume 59. Springer International Publishing Switzerland, 2015.
- [18] X. Meng, Randomized algorithms for large scale strongly over determined linear regression problems, , a dissertation submitted to the institute for computational and mathematical engineering and the committee on graduate studies of Stanford University, June 2014.
- [19] M. W. Mahoney , Randomized algorithms for matrices and data, Version appearing as a monograph in Now Publishers “Foundations and Trends in Machine Learning” series. Cornell University Library, 15 November 2011.
- [20] P. Drineas, M W. Mahoney, S. Muthukrishnan, T. Sarlos , Faster Least Squares Approximation, Cornell University Library, 26 September 2010.

Received December, 2016

RESUME

Dans cette thèse, nous nous intéressons à la résolution des problèmes des moindres carrés de grande taille. Afin d'accélérer la résolution de ce problème on est conduit à utiliser les matrices aléatoires, cette approche a reçu une grande attention dans les années récentes. Les matrices aléatoires se divise en deux parties : matrices d'échantillonnage et de projection, dans chaque partie on trouve des matrices aléatoires avec différentes estimations du temps de calcul. La solution obtenue par différentes méthodes aléatoires permet un gain de temps et de mémoire dans le calcul. Dans cette thèse, différentes méthodes aléatoires sont analysées et trois algorithmes sont proposés qui démontre la performance de cette théorie. Par la suite, nous optons pour le calcul de la solution en utilisant une combinaison entre l'échantillonnage moyen et la matrice aléatoire d'Hadamard tout en respectant la moyenne statistique. Une application est introduite pour le calcul des coefficients harmoniques du modèle géopotentiel.

Mots Clés :

Moindres carrés; Matrices aléatoires; Johnson Lindenstrauss; Echantillonnage; Projection aléatoire; Conditionnement; Matrice large; Produit matriciel; Coefficients harmoniques; Modèle géopotentiel.