École doctorale Sciences pour l'Ingénieur de l'Université de Lille

Thèse de Doctorat

Discipline

Mathématiques appliquées

Numéro d'ordre : 42018

présentée par

Daoud Ounaissi

Méthodes quasi-Monte Carlo et Monte Carlo : Application aux calculs des estimateurs Lasso et Lasso bayésien

Soutenue publiquement le 02 juin 2016 devant le jury composé de

Directeur de thèse :	Pr. Azzouz Dermoune	Université de Lille 1, France
Rapporteurs :	Pr. Emmanuel Gobet	Ecole Polytechnique de Palaiseau, France
	Pr. Denis Talay	Inria Sophia-Antipolis, France
Examinateurs :	Pr. David Dereudre	Université de Lille 1, France

Remerciements

Je souhaite tout d'abord exprimer ma profonde gratitude au Professeur Azzouz DERMOUNE, mon directeur de thèse pour avoir dirigé ce travail. Sa rigueur, sa clairvoyance, sa patience ainsi que le soutien qu'il m'a toujours apporté, m'ont permis de mener à bien cette thèse. Je n'oublierai jamais ces qualités scientifiques et humaines qui ont contribué énormément à la progression de mes travaux de recherche.

Mes sincères remerciements sont également adressés à Nadji RAHMANIA, professeur et collaborateur dans notre groupe de travail, pour les conseils précieux qu'il m'a accordé tout au long de ces années de recherche.

Un grand merci aux professeurs Emmanuel GOBET (Ecole Polytechnique de Palaiseau) et Denis TALAY (Directeur de recherche Inria), qui ont accepté de rapporter cette thèse. Leurs lectures attentives m'ont permis d'améliorer mon travail. Je tiens également à remercier le membre du jury : professeur David DEREUDRE.

Je tiens à remercier les doctorants du bureau Landry LAVOINE, Andrea CESARO et Najib IDRISSI, pour les moments merveilleux que nous avons passé ensemble pendant ces années. Je n'ai pas oublié mes amis : Ayman KRAYEM, Ayman EL ROZ, Amélie RENDOUR et Sihem ZIANI.

Enfin, je pense beaucoup à ma mère et à mon père, qui m'ont soutenu inconditionnellement malgré la distance tout au long de mes études en France. Je les remercie du fond de mon cœur pour leur sacrifice énorme.

Table des matières

Liste des notations

a	: lettre minuscule signifie un scalaire réel
$\mathbf{a}$	: lettre miniscule gras signifie un vecteur réel
$\mathbf{A}$	: lettre majuscule gras signifie une matrice réelle
$\mathbf{A}^\top$	: la matrice transposée de A
$\mathbf{I}$	: la matrice d'identité
$\mathbb{R}^d$	: l'ensemble des vecteurs colonne de dimension d
$\mathbb{R}^{n\times m}$	: l'ensemble des matrices réelles à n lignes et p colonnes
$< a, b >$	: produit scalaire de a et b égal à $\sum_{i=1}^{p} a_i c_i$ pour tout a et b dans $\mathbb{R}^p$
$\mathbb{E}[\cdot]$	: l'opérateur d'espérance mathématique
$\|\cdot\|_2$	: la norme l^2
$\|\cdot\|_1$	: la norme l^1
$Dom(g)$	: domaine de la fonction $Dom(\partial\varphi) = \{\mathbf{x} : \quad \partial\varphi(\mathbf{x}) \neq \emptyset\}$
$\arg\min(g)$	: minimiseur de la fonction g
$\mathcal{N}(\mathbf{m}, \Sigma)$	: vecteur normal de moyenne $\mathbf{m}$ et de matrice de covariance Σ

Chapitre 1

Régression linéaire et pénalisation

Dans ce chapitre nous allons rappeler le modèle de la régression linéaire avec les pénalisation l^2 et l^1 et leur interprétation bayésienne. Nous concluons ce chapitre par le plan de la thèse.

1.1 Régression linéaire

L'expérience qui utilise p variables $a^1, \ldots, a^p \in \mathbb{R}$ (appelées les entrées) produira un résultat $y \in \mathbb{R}$ (appelé la sortie). Si l'expérimentateur choisit n entrées $(a_1^1, \ldots, a_1^p)$, $\ldots$, $(a_n^1, \ldots, a_n^p)$ des mêmes variables $(a^1, \ldots, a^p)$, alors il obtient n sorties $y_1, \ldots, y_n$. La matrice $\mathbf{A} = [a_i^j : i = 1, \ldots, n, j = 1, \ldots, p]$ d'ordre $n \times p$ est appelée plan d'expérience (ou design). La qualité de l'expérience réside dans le choix des variables et de la matrice $\mathbf{A}$ (voir pour des références [2] et [3]). Par la suite, nous supposerons que le choix de la matrice de design $\mathbf{A}$, a été préalablement effectué.

Les sorties $\mathbf{y} = (y_1, \ldots, y_n)^T$ sont représentées sous forme d'une matrice $n \times 1$.

Dans beaucoup de domaines (science de l'ingénieur, statistique, économie, psychologie, médecine, ...) le lien entre le design $\mathbf{A}$ et les sorties $\mathbf{y}$ est modélisé par la régression linéaire

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w} \tag{1.1.1}$$

$$= \sum_{j=1}^{p} x_j \mathbf{a}^j + \mathbf{w}. \tag{1.1.2}$$

Les vecteurs $\mathbf{a}^1, \ldots, \mathbf{a}^p$ sont les colonnes de la matrice de design $\mathbf{A}$. Le coefficient x^j mesure l'incidence de la variable a^j sur la sortie y. Si $x^j = 0$, alors la variable a^j n'a pas d'incidence sur la sortie y. Si $|x^j|$ est très grande, alors la variable a^j a une grande incidence sur la sortie y. Le vecteur $\mathbf{w}$ est la déviation par rapport au modèle linéaire. Le problème consiste alors à estimer le vecteur inconnu $\mathbf{x}$ et le bruit représenté par le vecteur $\mathbf{w}$ connaissant les sorties $\mathbf{y}$ et la matrice de design $\mathbf{A}$. Une première approche consiste à utiliser la méthode des moindres carrés.

La première publication de la méthode des moindres carrés (destinée à déterminer des quantités dans un système d'équations sur-déterminé) est due à Legendre qui l'a explicitée en annexe de son ouvrage intitulé : *Nouvelles méthodes pour la détermination des orbites des comètes (1805)*. Cette méthode est aussi attribuée à Gauss. Aujourd'hui la méthode des moindres carrés est utilisée dans tous les domaines scientifiques (voir par exemple [10] pour un travail récent dans le domaine des équations différentielles stochastiques rétrogrades et les mathématiques financières).

FIGURE 1.1 – De gauche à droite : Gauss et Legendre.

Elle consiste à choisir $\mathbf{x}$ qui minimise $\|\mathbf{w}\|_2^2$ (la somme des carrés des erreurs) :

$$\hat{\mathbf{x}}_{MCO} := \arg\min\{\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 : \mathbf{x} \in \mathbb{R}^p\}. \tag{1.1.3}$$

Les solutions sont données par le système suivant

$$\mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{y}, \tag{1.1.4}$$

(équation normale). On parle de problème sur-déterminé si le nombre de paramètres p est plus petit que le nombre d'observations n i.e. $p \leq n$. Dans ce

cas, on dispose de nombreuses mesures (ou résultats d'observations), donc de plus d'équations que d'inconnues. Si les colonnes de $\mathbf{A}$ sont linéairement indépendantes, alors $\mathbf{A}^T\mathbf{A}$ est inversible. Dans ce cas la solution des moindres carrés est unique. Elle est donnée par

$$\hat{\mathbf{x}}_{MCO} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y}. \tag{1.1.5}$$

Si la matrice $\mathbf{A}^T\mathbf{A}$ est non inversible (c'est le cas des problèmes sous-déterminés c'est-à-dire $n < p$), alors l'équation normale a une infinité de solutions.

Pour surmonter cette difficulté on peut choisir la solution de l'équation normale ayant la plus petite norme Euclidienne

$$\mathbf{x}^+ = \mathbf{A}^+\mathbf{y}, \tag{1.1.6}$$

où $\mathbf{A}^+$ est l'inverse de Moore-Penrose de $\mathbf{A}$. Nous pouvons aussi faire appel aux méthodes de pénalisation.

1.2 Pénalisation l^2

L'estimateur des moindres carrés pénalisés par la norme l^2 est défini par

$$\hat{\mathbf{x}}(\lambda) := \arg\min\{\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda\|x\|_2^2 : \quad \mathbf{x} \in \mathbb{R}^p\}. \tag{1.2.1}$$

Cet estimateur dépend du paramètre $\lambda > 0$ (connu sous le nom de paramètre de pénalisation ou de lissage). Lorsque $\lambda \to 0$ l'estimateur $\hat{x}(\lambda)$ converge vers l'estimateur $\mathbf{x}^+$ des moindres carrés. Si $\lambda \to \infty$ l'estimateur $\hat{\mathbf{x}}(\lambda) \to 0$.

Résoudre le problème (1.2.1), est équivalent à résoudre le problème de minimisation sous contrainte suivant,

$$\arg\min\{\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 : \quad \mathbf{x} \in \mathbb{R}^p,\ \|\mathbf{x}\|_2^2 \leq \delta\}. \tag{1.2.2}$$

Le paramètre δ est relié au paramètre λ par l'équation $\|\hat{x}(\lambda)\|^2 = \delta^2$ [4].

L'estimateur $\hat{x}(\lambda)$ est la solution de l'équation normale pénalisée (connue sous le nom de pénalisation de Tikhonov)

$$(\mathbf{A}^T\mathbf{A} + \lambda\mathbf{I}_p)\mathbf{x} = \mathbf{A}^T\mathbf{y}. \tag{1.2.3}$$

La matrice $\mathbf{A}^T\mathbf{A}$ est une matrice semi-définie positive, ses valeurs propres sont donc positives ou nulles. Par conséquent $\mathbf{A}^T\mathbf{A} + \lambda\mathbf{I}_p$ est inversible pour tout $\lambda > 0$. L'équation (1.2.3) a une unique solution donnée par

$$\hat{\mathbf{x}}(\lambda) = (\mathbf{A}^T\mathbf{A} + \lambda\mathbf{I}_p)^{-1}\mathbf{A}^T\mathbf{y}. \tag{1.2.4}$$

FIGURE 1.2 – Bayes

1.2.1 Interprétation bayésienne de la pénalisation l^2

Nous supposons que le vecteur des erreurs $\mathbf{w} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_n)$ et le vecteur $\mathbf{x} \sim \mathcal{N}(0, \sigma_s^2 \mathbf{I}_p)$ où $\mathbf{x}$ et $\mathbf{w}$ sont indépendants.

Dans ce cas, la loi de $\mathbf{y}|\mathbf{x} \sim \mathcal{N}(\mathbf{A}\mathbf{x}, \sigma^2 \mathbf{I}_n)$; par conséquent la densité conditionnelle de $\mathbf{y}|\mathbf{x}$ est

$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\Big(- \frac{\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2}{2\sigma^2}\Big). \tag{1.2.5}$$

Comme la densité de la loi a priori de $\mathbf{x}$ est

$$p(\mathbf{x}) = \frac{1}{(2\pi\sigma_s^2)^{\frac{p}{2}}} \exp\Big(- \frac{\|\mathbf{x}\|_2^2}{2\sigma_s^2}\Big),$$

alors la densité a posteriori de $\mathbf{x}|\mathbf{y}$ est donnée par,

$$\begin{aligned} p(\mathbf{x}|\mathbf{y}) &= \frac{p(\mathbf{x}, \mathbf{y})}{p(\mathbf{y})} \\ &= \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})} \\ &= \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{Z} \\ &= \frac{1}{Z} \exp\Big(- \frac{\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2}{2\sigma^2}\Big) \exp\Big(- \frac{\|\mathbf{x}\|_2^2}{2\sigma_s^2}\Big) \\ &= \frac{1}{Z} \exp\left[-\Big(\frac{\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2}{2\sigma^2} + \frac{\|\mathbf{x}\|_2^2}{2\sigma_s^2}\Big)\right] \end{aligned}$$

où $Z = p(\mathbf{y})$ appelée fonction de partition.

L'estimateur de maximum à posteriori (MAP)

$$\begin{aligned}\hat{\mathbf{x}}_{MAP} &:= \arg\max\{p(\mathbf{x}|\mathbf{y}) \ : \ \mathbf{x} \in \mathbb{R}^p\} \\ &= \arg\max\{p(\mathbf{y}|\mathbf{x})p(\mathbf{x}) : \quad \mathbf{x} \in \mathbb{R}^p\} \\ &= \arg\min\{\frac{\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2}{2\sigma^2} + \frac{\|\mathbf{x}\|_2^2}{2\sigma_s^2} : \quad \mathbf{x} \in \mathbb{R}^p\} \\ &= \arg\min\{\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 + \frac{\sigma^2}{\sigma_s^2}\|\mathbf{x}\|_2^2 : \quad \mathbf{x} \in \mathbb{R}^p\} \\ &= (\mathbf{A}^T\mathbf{A} + \frac{\sigma^2}{\sigma_s^2}\mathbf{I}_p)^{-1}\mathbf{A}^T\mathbf{y} \qquad (1.2.6)\end{aligned}$$

coincide avec l'estimateur de Bayes

$$\mathbb{E}[\mathbf{x}|\mathbf{y}] = \int \mathbf{x} \exp\Big(-\frac{\|\mathbf{A}\mathbf{x}-\mathbf{y}\|_2^2}{2\sigma^2} - \frac{\|\mathbf{x}\|_2}{\sigma_s^2}\Big)d\mathbf{x}. \qquad (1.2.7)$$

Nous concluons que, l'estimateur $\hat{x}(\lambda)$ (1.2.1) est donné par $\hat{\mathbf{x}}_{MAP}$ avec le choix de $\lambda = \frac{\sigma^2}{\sigma_s^2}$ (rapport bruit signal). Il est bien connu que dans le cas gaussien l'estimateur $\hat{x}_{MAP}$ coïncide avec l'estimateur de Bayes $\mathbb{E}[\mathbf{x}|\mathbf{y}]$. Finalement si $\lambda = \frac{\sigma^2}{\sigma_s^2}$, alors l'estimateur $\hat{x}(\lambda)$ des moindres carrés pénalisés par la norme l^2 est égal à $\mathbb{E}[\mathbf{x}|\mathbf{y}]$.

1.3 Pénalisation l^1

Dans cette section nous allons introduire la pénalisation l^1. Ce genre de problème de pénalisation s'impose lorsque n (le nombre des observations) est inférieur à p (le nombre de paramètres) i.e. $(n < p)$. Dans ce cas le nombre de colonnes de A linéairement indépendants est au plus égal à n. Par conséquent dans la combinaison linéaire

$$\mathbf{A}\mathbf{x} = \sum_{j=1}^{p} x_j \mathbf{a}^j$$

nous n'avons besoin au plus que de n paramètres $x_{j_1}, ..., x_{j_n}$ non nuls, Il faudra donc sélectionner au moins $p-n$ paramètres nuls parmi les p paramètres $x_1, ..., x_p$.

Pour résoudre ce genre de problème, la pénalisation naturelle est la pseudo norme l_0 définie par

$$\|\mathbf{x}\|_0 = card\{i : x_i \neq 0\}.$$

Nous obtenons la pénalisation des moindres carrés par la pseudo norme l_0 :

$$\arg\min\{\frac{\|\mathbf{Ax}-\mathbf{y}\|_2^2}{2}+\lambda\|\mathbf{x}\|_0: \quad \mathbf{x}\in\mathbb{R}^p\}. \tag{1.3.1}$$

L'optimisation (1.3.1) est non convexe et donc difficile à résoudre ([13],[14]). Plusieurs auteurs ([5], [6]) ont essayé de définir un lien entre (1.3.1) et la pénalisation l^1 (qui a l'avantage d'être convexe) :

$$\arg\min\{\frac{\|\mathbf{Ax}-\mathbf{y}\|_2^2}{2}+\lambda\|\mathbf{x}\|_1: \quad \mathbf{x}\in\mathbb{R}^p\}. \tag{1.3.2}$$

La pénalisation l^1 a aussi été présentée et popularisée par Tibshirani [16], et est aujourd'hui connue sous le nom de problème LASSO (Least Absolute Shrinkage Operator). Un grand nombre de résultats théoriques autour du LASSO sont donnés dans ([5], [17]).

Résoudre le problème LASSO est équivalent à résoudre

$$\hat{\mathbf{x}}(\lambda)=\arg\min\{\frac{\|\mathbf{Ax}-\mathbf{y}\|_2^2}{2}: \quad \mathbf{x}\in\mathbb{R}^p,\ \|\mathbf{x}\|_1\leq\delta\}, \tag{1.3.3}$$

où δ et λ sont liés par $\|\hat{x}(\lambda)\|_1=\delta$.

En général, LASSO ne peut pas être trouvé explicitement. Les algorithmes les plus populaires pour approcher LASSO sont : LARS [10], les algorithmes ISTA et FISTA [1] et [14].

1.3.1 Interprétation bayésienne du LASSO

À l'image de la pénalisation l^2, la forme de l'expression (2.1.2) suggère que LASSO peut être interprété par une approche bayésienne. Tibshirani [16] note que l'estimateur de LASSO peut être considéré comme le mode de la loi a posteriori de $\mathbf{x}$. Si on modélise la loi a priori de $\mathbf{x}$ par la loi Laplace de paramètre $\alpha>0$, i.e. la loi a priori sur $\mathbf{x}$ est de densité donnée par

$$p(\mathbf{x})=\frac{1}{(2\alpha)^p})\exp\Big(-\frac{\|\mathbf{x}\|_1}{\alpha}\Big). \tag{1.3.4}$$

Nous modélisons $\mathbf{w}\sim\mathcal{N}(0,\sigma^2\mathbf{I}_n)$ de facon à ce que $\mathbf{w}$ et $\mathbf{x}$ soient indépendants. Ainsi la densité a posteriori de $\mathbf{x}$ est donnée par,

$$
\begin{aligned}
p(\mathbf{x}|\mathbf{y}) &= \frac{p(\mathbf{x},\mathbf{y})}{p(\mathbf{y})} \\
&= \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})} \\
&= \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{Z} \\
&= \frac{1}{Z}\exp\Big(-\frac{\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2}{2\sigma^2}\Big)\exp\Big(-\frac{\|\mathbf{x}\|_1}{\alpha}\Big) \\
&= \frac{1}{Z}\exp\Big[-\Big(\frac{\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2}{2\sigma^2}+\frac{\|\mathbf{x}\|_1}{\alpha}\Big)\Big],
\end{aligned}
$$

où $Z = p(\mathbf{y})$ est appelé fonction de partition.

Dans ce cas, l'estimateur de MAP est égal à

$$
\begin{aligned}
\hat{\mathbf{x}}_{MAP} &= \arg\max\{p(\mathbf{x}|\mathbf{y}) \,:\, \mathbf{x}\in\mathbb{R}^p\} \\
&= \arg\max\{p(\mathbf{y}|\mathbf{x})p(\mathbf{x}) : \quad \mathbf{x}\in\mathbb{R}^p\} \\
&= \arg\min\{\frac{\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2}{2\sigma^2}+\frac{\|\mathbf{x}\|_1}{\alpha} : \quad \mathbf{x}\in\mathbb{R}^p\} \\
&= \arg\min\{\frac{\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2}{2}+\frac{\sigma^2}{\alpha}\|\mathbf{x}\|_1 : \quad \mathbf{x}\in\mathbb{R}^p\}. \qquad (1.3.5)
\end{aligned}
$$

Cette approche bayésienne nous présente l'estimateur LASSO (2.1.2) comme un estimateur de MAP $\hat{\mathbf{x}}_{MAP}$, avec le choix de $\lambda = \frac{\sigma^2}{\alpha}$. Contrairement au cas gaussien, l'estimateur $\hat{\mathbf{x}}_{MAP}$ n'est pas égal à l'estimateur de Bayes $\mathbb{E}[\mathbf{x}|\mathbf{y}]$.

1.4 Exemple

1.4.1 Exemple avec données réelles : Nombre d'observations n > Nombre de paramètres p

La pollution de l'air constitue actuellement une des préoccupations majeures de santé publique. De nombreuses études épidémiologiques ont permis de mettre en évidence l'influence sur la santé de certains composés chimiques comme le dioxyde de souffre ($SO2$), le dioxyde d'azote ($NO2$), l'ozone (O_3). L'influence de cette pollution est notable sur les personnes sensibles (nouveau-nés, asthmatiques, personnes âgées). Nous allons nous intéresser plus particulièrement à la concentration en ozone.

Afin de mieux comprendre ce phénomène, l'association Air Breizh (surveillance de la qualité de l'air en Bretagne) mesure depuis 1994 la concentration en O_3 (en $\mu g/ml$) toutes les 10 minutes et obtient donc le maximum journalier de la concentration en O_3, noté dorénavant O_3. Air Breizh collecte également à certaines heures de la journée des données météorologiques comme la température (T), la nébulosité (Ne), le vent (V)... le tableau 1.1. donne un extrait de ces données [12].

T	63.6	89.6	79	81.2	88	86.4	139	78.2	113.8	41.8
V	13.4	15	7.9	13.1	14.1	16.7	26.8	18.4	27.2	20.6
Ne	9.35	5.4	19.3	12.6	-20.3	-3.69	8.27	4.93	-4.93	-3.38
O_3	7	4	8	7	6	7	1	7	6	8

TABLE 1.1 – 10 données journalières.

Si nous voulons expliquer la concentration en $\mathbf{y} = O_3$ comme une fonction affine de ces trois variables météorologiques T, N_e, V, nous considérons le modèle : $O_3 = x_0 + x_1 T + x_2 N_e + x_3 V + w$. En se basant sur les données du tableau 1.1, le modèle précédent peut s'écrire : $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w}$, où

$$\mathbf{A} = \begin{bmatrix} 1 & 63.6 & 13.4 & 9.35 \\ 1 & 89.6 & 15 & 5.4 \\ 1 & 79 & 7.9 & 19.3 \\ 1 & 81.2 & 13.1 & 12.6 \\ 1 & 88 & 14.1 & -20.3 \\ 1 & 86.4 & 16.7 & -3.69 \\ 1 & 139 & 26.8 & 8.27 \\ 1 & 78.2 & 18.4 & 4.93 \\ 1 & 113.8 & 27.2 & -4.93 \\ 1 & 41.8 & 20.6 & -3.38 \end{bmatrix}, \mathbf{x} = \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{bmatrix} \text{ et } \mathbf{y} = \begin{bmatrix} 7 \\ 4 \\ 8 \\ 7 \\ 6 \\ 7 \\ 1 \\ 7 \\ 6 \\ 8 \end{bmatrix}.$$

La matrice $\mathbf{A}^T\mathbf{A}$ est inversible ; l'estimateur des moindres carrés est alors

$$\hat{x} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y} = \begin{bmatrix} 84.5843 \\ 1.3150 \\ 0.4864 \\ -4.8935 \end{bmatrix}$$

1.5 Plan de la thèse

Dans cette thèse nous nous intéressons uniquement au cas où le nombre de paramètres p est supérieur au nombre des observations n.

1. La matrice $\mathbf{A}$ sera toujours une réalisation d'une matrice aléatoire dont les entrées sont i.i.d. de loi de Bernoulli $\mathcal{B}(\pm\frac{1}{\sqrt{n}})$ ou bien des gaussiennes $\mathcal{N}(0, \frac{1}{n})$.
2. Nous générons le vecteur bruit $\mathbf{w}$ selon la loi gaussienne $\mathcal{N}(0, \sigma^2\mathbf{I}_n)$, et le vecteur $\mathbf{x}$ selon la loi de Laplace de paramètre α (1.3.4).
3. Nous obtenons la sortie

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w}.$$

Sachant $\mathbf{y}$, $\mathbf{A}$ et les lois a priori de $\mathbf{x}$ et $\mathbf{w}$ nous nous proposons de retrouver le vecteur $\mathbf{x}$.

Dans le chapitre 2 nous rappelons l'algorithme FISTA et étudions statistiquement sa convergence (la partie entière et les trois premières décimales). Nous utiliserons à cet effet l'outil de l'entropie et l'algorithme de Pitman-Yor.

Dans le chapitre 3 nous nous intéressons à l'estimateur bayésien de $\mathbf{x}$, c'est-à-dire la moyenne de la loi a posteriori. Nous comparons les méthodes de quasi-Monte Carlo et Monte-Carlo dans les calculs de l'estimateur bayésien et de sa fonction de partition Z.

Le chapitre 4 est extrait de la prépublication [7]. Dans ce chapitre, nous étudions l'estimateur bayésien en utilisant les coordonnées polaires. Ceci nous permet de donner une interprétation géométrique au mode de la loi a posteriori (LASSO) et à la fonction de partition Z. De plus nous déduisons une inégalité de concertation pour la loi a posteriori. Nous utilisons cette inégalité comme critère de convergence de la méthode MCMC.

Le chapitre 5 est extrait de la prépublication [8]. Dans ce chapitre, nous proposons d'estimer la loi a posteriori en utilisant un système d'équations différentielles stochastique (EDS) dont la dérive est singulière. Nous rappelons les résultats théoriques et numériques associés à cette (EDS). Nous interprétons la loi a posteriori comme la loi limite de l'EDS, et nous comparons les schémas d'Euler semi-implicite et explicite en utilisant les méthodes de Monte Carlo, Monte Carlo à plusieurs couches et MCMC.

Enfin dans le chapitre 6, qui est une publication parue dans Mathematics and Computers in Simulation [9], nous étudions le comportement de la loi a posteriori lorsque les paramètres $\sigma^2 \to 0$ et $\frac{\sigma^2}{\alpha} = 1$. Nous montrons que la loi a posteriori converge vers LASSO et nous précisons sa vitesse de convergence. Cette vitesse de convergence est utilisée pour donner des critères de convergence des méthodes MCMC ayant la loi bayesienne comme loi cible.

Bibliographie

[1] A. Beck, M. Teboulle, A Fast Iterative Shrinkage-thresholding algorithm for linear inverse problem, SIAM J. Imaging Sci. 183–202 (2009).

[2] G.E.P. Box and N.R. Drapper, Empirical model–building and response surfaces, John Wiley and Sons, New York, (1987).

[3] G.E.P. Box, W.G. Hunter and J.S. Hunter, Statistics for experimenters : An introduction to designs, data analysis and model building, Wiley, New York, (1978).

[4] S. Boyd, L. Vandenberghe, Convex Optimization, Cambridge University Press (2004).

[5] E.J. Candès, The restricted isometry property and its implications for compressed sensing, Applied and Computational Mathematics, California Institute of Technology, Pasadena, CA91125-5000 (2008).

[6] S. Chen, D.L. Donoho, M. Saunders, Atomic decomposition by basis pursuit, SIAM J. Sci. Comput. Vol. 20, no. 1 33–61 (1998).

[7] A. Dermoune, D. Ounaissi, N. Rahmania, MCMC convergence diagnosis using geometry of Bayesian LASSO, arXiv :1512.01366v1 [math.ST] (2015).

[8] A. Dermoune, D. Ounaissi, N. Rahmania, Multilevel Monte Carlo simulation of a diffusion with non-smooth drift, arXiv :1504.06441v1 [math.ST] (2015).

[9] A. Dermoune, D. Ounaissi, N. Rahmania, Oscillation of Metropolis-Hastings and simulated annealing algorithms around LASSO estimator, Math. Comput. Simulation (2015).

[10] B. Efron, T. Hastie, I. Johnstone, R. Tibshirani, Least angle regression, Ann. Statist. vol. 32 no. 2 407–499 (2004).

[11] E. Gobet, P.Turkedjiev, Linear regression MDP scheme for discrete backward stochastic differential equations under general conditions, hal-00855760v2 (2014).

[12] A. Guyader, Cours régression linéaire, Université de Rennes 2, master de statistique 2 1–82 (2009).

[13] S.G. Mallat and Z. Zhang, Matching pursuits with time-frequency dictionaries, IEEE Trans. Signal Process. vol. 41 no. 12 3397–3415, (1993).

[14] B. K. Natarajan, Sparse approximate solutions to linear systems, SIAM J. Comput. vol. 24 no.2 227–234 (1995).

[15] N. Parikh, S. Boyed, Proximal algorithms, Found. Trends Optim. vol. 1 no. 3 123–231 (2003).

[16] R. Tibshirani, Regression shrinkage and selection via LASSO, J. R. Stat. Soc. Ser. B Stat. Methodol. vol. 58 no. 1 267–288 (1996).

[17] R. Tibshirani, The LASSO problem and uniqueness, Electron, J. Stat. vol. 7 1456–1490 (2013).

Chapitre 2

Résolution du problème LASSO par l'algorithme FISTA : Application à l'estimateur de Pitman-Yor

Dans ce chapitre nous rappelons certains algorithmes classiques dans les problèmes d'optimisation et décrivons l'algorithme FISTA pour résoudre numériquement le problème LASSO. Nous terminons ce chapitre par l'étude statistique de la convergence de FISTA en utilisant l'estimateur de Pitman-Yor [1].

2.1 L'optimisation convexe

Nous allons tout d'abord rappeler les algorithmes permettant de résoudre numériquement le problème d'optimisation suivant

$$\min\{F(\mathbf{x}) = f(\mathbf{x}) + g(\mathbf{x}) : \quad \mathbf{x} \in \mathbb{R}^p\}. \tag{2.1.1}$$

Dans ce chapitre $f : \mathbb{R}^p \to \mathbb{R}$ est une fonction convexe de classe $C^{1,1}$ i.e. une fonction continue, différentiable, ayant un gradient continu et L_f-Lipschitzien :

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\| \leq L_f \|\mathbf{x} - \mathbf{y}\|, \quad \forall \mathbf{x} \,\text{et}\, \mathbf{y} \in \mathbb{R}^p.$$

Tout au long de ce chapitre L_f désigne la constante de Lipschitz de f. La fonction $g : \mathbb{R}^p \to \mathbb{R}$ est une fonction convexe mais pas forcement différentiable.

Bien souvent le problème (2.1.1) n'admet pas de solution analytique. Sa résolution recourt à des algorithmes numériques.

Exemple

Le problème LASSO est évidemment un cas particulier du problème (2.1.1) avec $f(\mathbf{x}) = \frac{\|\mathbf{A}\mathbf{x}-\mathbf{y}\|_2^2}{2}$, $g(\mathbf{x}) = \|\mathbf{x}\|_1$. La constante de Lipschitz de ∇f est donnée par $L_f = \lambda_{max}(\mathbf{A}^T\mathbf{A})$, où $\lambda_{max}(\mathbf{A}^T\mathbf{A})$ est la plus grande valeur propre de la matrice $\mathbf{A}^T\mathbf{A}$.

L'optimisation (2.1.1) devient

$$\min\{\frac{\|\mathbf{A}\mathbf{x}-\mathbf{y}\|_2^2}{2} + \|\mathbf{x}\|_1 : \quad \mathbf{x} \in \mathbb{R}^p\}. \tag{2.1.2}$$

2.2 Le cas régulier : $g \equiv 0$

2.2.1 Algorithme de descente du gradient

Dans le cas où la fonction $g \equiv 0$, le problème (2.1.1) devient,

$$\min\{f(\mathbf{x}) : \quad \mathbf{x} \in \mathbb{R}^p\} \tag{2.2.1}$$

où f est une fonction convexe de classe $C^{1,1}$. L'algorithme le plus connu est la méthode de descente du gradient

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \lambda_k \nabla f(\mathbf{x}_k),$$

où $\lambda_k > 0$ est choisi convenablement. Dans la suite nous allons rappeler d'autres algorithmes plus performants.

2.2.2 Algorithme de la boule lourde (Polyak)

L'algorithme de la boule lourde (heavy-ball) a été introduit par Polyak [5]. Cette méthode est une discrétisation du système dynamique (mouvement d'une boule lourde) défini par l'équation différentielle suivante,

$$\frac{d^2\mathbf{x}(t)}{d^2t} + \alpha_1\frac{d\mathbf{x}(t)}{dt} + \alpha_2\nabla f(\mathbf{x}(t)) = 0, \quad t > 0$$

avec α_1 et $\alpha_2 > 0$ sont les paramètres du système.

Cet algorithme (à deux mémoires) est donné par,

Etape 0. Nous Choisissons un $\mathbf{x}_0$ et $\mathbf{x}_1 \in \mathbb{R}^p$, $\alpha_1 \in (0,1)$
Etape k. $(k \geq 1)$ nous calculons

$$\mathbf{z}_k = \mathbf{x}_k + \alpha_k(\mathbf{x}_k - \mathbf{x}_{k-1}), \tag{2.2.2}$$

$$\mathbf{x}_{k+1} = \mathbf{z}_k - \lambda_k \nabla f(\mathbf{x}_k) \tag{2.2.3}$$

où $\alpha_k \in (0,1)$ et λ_k choisi convenablement. De nombreux choix existent pour α_k ([2], [6]). Clairement le calcul de $\mathbf{x}_{k+1}$ se fait à l'aide de $\mathbf{x}_k$ et $\mathbf{x}_{k-1}$.

Remarque 2.2.1. Le choix $\alpha_k = 0$ dans (2.3.1) nous conduit à la méthode du gradient.

2.2.3 Algorithme de Nesterov

Nesterov a proposé un algorithme qui améliore la performance de l'algorithme de la boule lourde [4]. Son algorithme est donné par,

Etape 0. Nous choisissons $\mathbf{x}_0$ et $\mathbf{x}_1 \in \mathbb{R}^p$, $\alpha_1 \in (0,1)$
Etape k. $(k \geq 1)$ nous calculons

$$\mathbf{z}_k = \mathbf{x}_k + \alpha_k(\mathbf{x}_k - \mathbf{x}_{k-1}), \tag{2.2.4}$$

$$\mathbf{x}_{k+1} = \mathbf{z}_k - \frac{1}{L_f}\nabla f(\mathbf{z}_k) \tag{2.2.5}$$

où L_f est la constante de Lipschitz de ∇f. Contrairement à l'algorithme de la boule lourde, $\mathbf{x}_{k+1}$ se calcule uniquement en fonction de $\mathbf{z}_k$.

2.2.4 Algorithme de Beck et Teboulle

Récemment, Beck et Teboulle [2] ont proposé un algorithme qui améliore la performance de Nesterov, et qui est aussi une méthode du gradient à deux

mémoires. Cet algorithme est défini comme ci dessous,

Nous donnons : L_f la constante de Lipschitz de ∇f
Etape 0. Nous choisissons $\mathbf{z}_1 = \mathbf{x}_0 \in \mathbb{R}^p$ et $t_1 = 1$
Etape k. $(k \geq 1)$ nous calculons

$$\mathbf{x}_k = \mathbf{z}_k + \frac{1}{L_f}\nabla f(\mathbf{z}_k) \tag{2.2.6}$$

$$t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2} \tag{2.2.7}$$

$$\mathbf{z}_{k+1} = \mathbf{x}_k + \Big(\frac{t_k - 1}{t_{k+1}}\Big)(\mathbf{x}_k - \mathbf{x}_{k-1}) \tag{2.2.8}$$

Remarque 2.2.2. La relation (2.2.7) est équivalente à

$$\begin{aligned}(2t_{k+1} - 1)^2 &= 1 + 4t_k^2\\ t_{k+1}^2 - t_{k+1} &= t_k^2\\ t_{k+1}^2 - t_{k+1} &= t_k^2.\end{aligned}$$

En parcourant la preuve de la vitesse de convergence de cet algorithme, nous nous apercevons que nous n'avons besoin que de l'inégalité

$$t_{k+1}^2 - t_{k+1} \leq t_k^2. \tag{2.2.9}$$

Soit (t_k) une suite qui vérifie (2.2.9). **Etape 0.** Nous choisissons $\mathbf{x}_0 = \mathbf{v}_0$
Etape k. $(k \geq 1)$ nous calculons

$$\mathbf{z}_{k-1} = (1 - \frac{1}{t_k})\mathbf{x}_{k-1} + \frac{1}{t_k}\mathbf{v}_{k-1}, \tag{2.2.10}$$

$$\mathbf{x}_k = \mathbf{z}_{k-1} - \frac{1}{L_f}\nabla f(\mathbf{z}_{k-1}) \tag{2.2.11}$$

$$\mathbf{v}_k = \mathbf{x}_{k-1} + t_k(\mathbf{x}_k - \mathbf{x}_{k-1}). \tag{2.2.12}$$

Rappelons que l'algorithme proposé par Beck et Teboulle est le plus performant pour résoudre le problème (2.2.1). Sa vitesse de convergence est donnée par,

$$f(\mathbf{x}_k) - f(\mathbf{x}^*) \leq \frac{L_f}{2t_k^2}\|x_0 - \mathbf{x}^*\|^2, \quad \forall k \geq 1. \tag{2.2.13}$$

L'inégalité (2.2.13) nous donne la vitesse de convergence de la fonction f, qui est de l'ordre $O(\frac{1}{k^2})$, mais elle ne nous donne pas d'information ni sur la vitesse de convergence ni de la convergence de la suite $\{\mathbf{x}_k\}_k$.

Si la fonction f est m-strictement convexe au voisinage de $\mathbf{x}^*$ i.e.

$$\frac{m}{2}\|\mathbf{x} - \mathbf{x}^*\|^2 \leq f(\mathbf{x}) - f(\mathbf{x}^*),$$

alors la vitesse de convergence de la suite $\{\mathbf{x}_k\}_k$ est donnée par l'inégalité suivante,

$$\frac{m}{2}\|\mathbf{x} - \mathbf{x}^*\|^2 \leq f(\mathbf{x}_k) - f(\mathbf{x}^*) \leq \frac{2L_f}{(k+1)^2}\|x_0 - \mathbf{x}^*\|^2, \quad \forall k \geq 1.$$

Récemment Chambolle et Dossal [3] ont montré le résultat suivant. Le choix $a > 2$ et $t_k = \frac{k+a-1}{a}$ entraîne la convergence de la suite $\{\mathbf{x}_k\}_k$ vers un minimisateur de f.

2.3 Algorithme FISTA

Revenons au problème général (2.1.1) que nous rappelons ci dessous,

$$\min\{F(\mathbf{x}) = f(\mathbf{x}) + g(\mathbf{x}) : \quad \mathbf{x} \in \mathbb{R}^p\}$$

où f est une fonction de classe $C^{1,1}$ convexe, et g est seulement convexe.

L'algorithme de Beck et Teboulle associé à cette optimisation repose sur la fonction proximale de la fonction g définie dans l'exemple ci-dessous.

Introduit par Beck-Teboulle, l'algorithme FISTA est défini par :

Étape 0. Nous choisissons $\mathbf{x}_0 = \mathbf{v}_0$ et soit L_f - la constante Lipschitz de ∇f.
Étape k. $(k \geq 1)$ nous calculons

$$\mathbf{z}_{k-1} = (1 - \frac{1}{t_k})\mathbf{x}_{k-1} + \frac{1}{t_k}\mathbf{v}_{k-1}, \tag{2.3.1}$$

$$\begin{aligned}\mathbf{x}_k &= prox_{\frac{g}{L_f}}(\mathbf{z}_{k-1} - \frac{1}{L_f}\nabla f(\mathbf{z}_{k-1})) \\ &= p_{L_f}(\mathbf{z}_{k-1})\end{aligned} \tag{2.3.2}$$

$$\mathbf{v}_k = \mathbf{x}_{k-1} + t_k(\mathbf{x}_k - \mathbf{x}_{k-1}). \tag{2.3.3}$$

La vitesse de convergence de cet algorithme est la même que dans le cas $g = 0$. Le résultat de Chambolle et Dossal [3] concernant la convergence de la suite $\{\mathbf{x}_k\}_k$ reste aussi valable.

Une des applications de l'algorithme FISTA est la résolution du problème LASSO. L'opérateur *prox* de la fonction $g(\mathbf{x}) = \lambda\|\mathbf{x}\|_1$ est égal à

$$prox_{\lambda\|.\|_1}(\mathbf{x}) = \big(sgn(x_i)\max(|x_i| - \lambda, 0); \quad i = 1, ..., p\big).$$

Cet opérateur attribue aux composantes $|x_i| \leq \lambda$ la valeur 0. Ceci explique le nom FISTA (Fast Iterative Shrinkage Threshold Algorithm) donné à l'algorithme de Beck-Teboulle.

2.4 Étude statistique de la convergence de la suite générée par FISTA

L'algorithme FISTA ne donne pas d'information sur la vitesse de convergence de la suite générée $\{\mathbf{x}_k\}_k$. Dans le reste de ce chapitre, nous allons étudier statistiquement la convergence des quatre premières décimales de la suite $\{\mathbf{x}_k\}_k$ dans le cas du problème LASSO :

$$\arg\min_{\mathbf{x}\in\mathbb{R}^p}\{\frac{\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2}{2} + \|\mathbf{x}\|_1\}.$$

La matrice $\mathbf{A}$ et le vecteur $\mathbf{y}$ sont générés selon les trois procédures de la section (1.5). Nous éxécutons FISTA avec la suite $\frac{1}{t_k} := \theta_k = \frac{2}{k+1}$. Ici $n = 7$ et $p = 10$.

Nous allons faire cette étude dans deux cas. Le cas où les entrées de la matrice $\mathbf{A}$ sont i.i.d. de loi de Bernoulli $\mathcal{B}(\pm\frac{1}{\sqrt{n}})$, et le cas où les entrées de la matrice $\mathbf{A}$ sont i.i.d. de loi normale $\mathcal{N}(0, \frac{1}{n})$.

Nous exécutons l'algorithme FISTA et obtenons les itérations $(\mathbf{x}_k : k = 0, \ldots, N)$. Pour chaque k, $(x_k(i) : i = 1, ..., p)$ désigne les composantes du vecteur $\mathbf{x}_k$. Pour chaque i, nous représentons $x_k(i)$ à l'aide de sa partie entière "$\lfloor . \rfloor$" et du développement décimal :

$$\begin{cases} x_k(i) & := & \lfloor x_k(i) \rfloor + 10^{-1} \times d_k^1(i) + 10^{-2} \times d_k^2(i) + ... + 10^{-m} \times d_k^m(i) + ... \\ i & = & 1, ..., p \end{cases}$$

Les variables entières qui nous intéressent sont : le nombre minimal d'itérations pour stabiliser la partie entière et les trois premières décimales de la suite $(\mathbf{x}_k)$, c'est-à-dire

$$\begin{cases} s_0 & := & \min\{k : \lfloor x_k(i) \rfloor = \lfloor x_{k+1}(i) \rfloor, \forall i = 1, ..., p\}, \\ s_l & := & \min\{k : d_k^l(i) = d_{k+1}^l(i), \forall i = 1, ..., p\}, \quad l = 1, 2, 3, \end{cases}$$

Nous obtenons un N-échantillon $[(s_l^m, l = 0, \ldots, 3) : m = 1, \ldots, N]$ de $(s_l : l = 0, \ldots, 3)$, en exécutant l'algorithme FISTA N-fois (en générant $\mathbf{y}$ à chaque fois).

Nous allons estimer l'entropie de la loi des variables aléatoires discrètes s_0, s_1, s_2, s_3 en utilisant l'estimateur empirique (l'estimateur Plugin) et et l'estimateur de Pitman-Yor .

2.5 Rappels

2.5.1 Entropie

Soit $\mathcal{A}$ un nombre entier positif, X une variable aléatoire à valeurs dans $\{1, \ldots, \mathcal{A}\}$, et $(\pi_1, \ldots, \pi_{\mathcal{A}})$ sa distribution de probabilité. L'entropie de X est définie par

$$H(X) = -\sum_{i=1}^{\mathcal{A}} \pi_i \ln(\pi_i). \tag{2.5.1}$$

Nous rappelons que l'entropie est minimale lorsque X est constante, et elle est maximale lorsque X est uniforme. Ainsi

$$0 \leq H(X) \leq \ln(\mathcal{A})$$

est une mesure de l'aléa contenu dans X.

Définition : Estimateur Plug-in de l'entropie

Soit $x_1, \ldots, x_N$ un N-échantillon de X. L'entropie empirique de X s'obtient en estimant la probabilité π_i par la fréquence empirique $\hat{\pi}_i := \frac{n_i}{N}$, où $n_i = \sum_{k=1}^{N} \mathbf{1}_{[x_k=i]}$:

$$\hat{H}_{Plugin} = -\sum_{i=1}^{\mathcal{A}} \hat{\pi}_i \ln(\hat{\pi}_i). \tag{2.5.2}$$

Remarque :

Si $N < \mathcal{A}$, alors certains éléments i ne sont pas observés, c'est-à-dire $n_i = 0$, alors l'estimateur $\hat{\pi} = 0$. Pour contourner cette difficulté, nous utilisons l'approche Bayésienne. Nous avons besoin de tirer le loi de X selon une loi a priori. Les lois de Dirichlet sont les plus populaires.

2.5.2 Loi de Dirichlet

La loi de Dirichlet $\mathcal{P}_{n,\beta}$ de paramètres $n \geq 2$ (entier), $\beta > 0$, est une loi de probabilité sur l'ensemble

$$\Pi = \{(\pi_1, \ldots, \pi_n) \in (0,1)^n : \quad \sum_{i=1}^{n} \pi_i = 1\} \tag{2.5.3}$$

des probabilités sur l'ensemble $\{1, \ldots, n\}$.

Comment simuler selon la loi de Dirichlet $\mathcal{P}_{n,\beta}$? Nous rappelons que la loi beta $B(a,b)$ de paramètres $a > 0, b > 0$, a pour densité

$$\frac{\Gamma(a+b)x^{a-1}(1-x)^{b-1}}{\Gamma(a)\Gamma(b)}.$$

Nous générons une suite $V_1, V_2, \ldots, V_{n-1}$ mutuellement indépendantes de lois respectives $B(a,(n-1)a)$, $B(a,(n-2)a)$, $\ldots$, $B(a,a)$ et posons $V_n = 1$. Puis nous construisons la suite

$$\pi_1 = V_1,\ \pi_i = (1-V_1)\ldots(1-V_{i-1})V_i, \quad i = 2, \ldots, n.$$

Alors $(\pi_1, \ldots, \pi_n)$ est une probabilité sur $\{1, \ldots, n\}$ tirée selon la loi de Dirichlet $\mathcal{P}_{n,a}$.

2.5.3 Loi de Dirichlet généralisée

Ceci nous permet de définir une classe plus large de loi à priori à partir de deux suites $\mathbf{a} = (a_1, \ldots, a_{n-1})$ et $\mathbf{b} = (b_1, \ldots, b_{n-1})$ de nombres positifs.

Nous générons une suite $V_1, V_2, \ldots, V_{n-1}$ mutuellement indépendantes de lois respectives $B(a_1, b_1)$, $B(a_2, b_2)$, $\ldots$, $B(a_{n-1}, b_{n-1})$ et posons $V_n = 1$. Nous construisons ainsi la suite décroissante

$$\pi_1 = V_1,\ \pi_i = (1 - V_1)\ldots(1 - V_{i-1})V_i, \quad i = 2, \ldots, n.$$

Alors $(\pi_1, \ldots, \pi_n)$ est une probabilité sur $\{1, \ldots, n\}$ tirée selon la loi de Dirichlet $\mathcal{P}_{n,(\mathbf{a},\mathbf{b})}$.

2.5.4 Estimateur de Dirichlet

Nous considérons l'ensemble

$$\Pi = \{(\pi_1, \ldots, \pi_{\mathcal{A}}) \in (0,1)^{\mathcal{A}} : \quad \sum_{i=1}^{\mathcal{A}} \pi_i = 1\} \tag{2.5.4}$$

des probabilités sur l'ensemble $\{1, \ldots, \mathcal{A}\}$. Soit $\mathcal{P}$ une loi a priori de π (2.5.3). Ayant $\mathcal{P}$, et ayant l'échantillon $x_1, \ldots, x_N$ de X la loi a posteriori de π

$$P(\pi \,|\, \mathbf{x}) = \frac{P(\mathbf{x} \,|\, \pi)\mathcal{P}(\pi)}{P(\mathbf{x})}. \tag{2.5.5}$$

Ici $\mathbf{x} = (x_1, \ldots, x_N)$,

$$P(\mathbf{x}) = \int_{[0,1]^{\mathcal{A}}} P(\mathbf{x} \,|\, \pi)\mathcal{P}(\pi)\delta(1 - \sum_{i=1}^{\mathcal{A}} \pi_i)d\pi.$$

Sachant π, la loi de $\mathbf{x}$ est donnée par :

$$P(\mathbf{x} \,|\, \pi) = \pi_1^{n_1} \ldots \pi_{\mathcal{A}}^{n_{\mathcal{A}}}.$$

Par conséquent

$$P(\pi \,|\, \mathbf{x}) = \frac{\pi_1^{n_1} \ldots \pi_{\mathcal{A}}^{n_{\mathcal{A}}}\mathcal{P}(\pi)}{\int_\Pi \pi_1^{n_1} \ldots \pi_{\mathcal{A}}^{n_{\mathcal{A}}}\mathcal{P}(\pi)d\pi}. \tag{2.5.6}$$

La loi de Dirichlet de paramètre $\beta > 0$ sur Π est très populaire. Elle est définie par

$$\mathcal{P}_{\mathcal{A},\beta}(\pi) = \frac{1}{Z_\beta}\delta(1 - \sum_{i=1}^{\mathcal{A}} \pi_i) \prod_{i=1}^{\mathcal{A}} \pi_i^{\beta-1} \tag{2.5.7}$$

La constante

$$Z_\beta = \int_{[0,1]^{\mathcal{A}}} \prod_{i=1}^{\mathcal{A}} \pi_i^{\beta-1} \delta(1 - \sum_{i=1}^{\mathcal{A}} \pi_i) d\pi = \frac{\Gamma(\beta)^{\mathcal{A}}}{\Gamma(\mathcal{A}\beta)}.$$

Si $\beta = 1$, $\mathcal{P}_\beta(\pi) = \frac{1}{Z_1}$ est la loi "uniforme" sur Π.

Dans ce cas l'estimateur de Bayes de π_i sachant $\mathbf{n} := (n_1, \ldots, n_{\mathcal{A}})$ est égal à

$$\langle q_i \rangle_\beta := \int_{[0,1]^{\mathcal{A}}} \pi_i P_\beta(\pi \,|\, \mathbf{n}) d\pi = \frac{n_i + \beta}{N + \mathcal{A}\beta}.$$

Lorsque $\beta \to 0$, nous retrouvons l'estimateur empirique

$$\langle q_i \rangle_0 = \frac{n_i}{N}.$$

Lorsque $\beta = 1$, nous retrouvons l'estimateur de Laplace

$$\langle q_i \rangle_1 = \frac{n_i + 1}{N + \mathcal{A}}.$$

En combinant (2.5.6) et (2.5.7) nous avons que,

$$P(\pi|\mathbf{x}) \approx \prod_{i=1}^{\mathcal{A}} \pi_i^{(n_i+\beta)-1} \tag{2.5.8}$$

et nous obtenons donc que la loi a posteriori de π est aussi une loi de Dirichlet de paramètre $n_1 + \beta, \ldots, n_{\mathcal{A}} + \beta$.

2.6 Estimateur de Pitman-Yor

Nous prenons $\mathcal{A} = +\infty$ lorsque $\mathcal{A}$ est très grand. Dans ce cas la loi à priori $\mathcal{P}_{(\mathbf{a},\mathbf{b})}$ dépend de deux suites infinies $\mathbf{a} = (a_1, \ldots)$, $\mathbf{b} = (b_1, \ldots)$ choisie de la manière suivante :

$$\sum_{i=1}^{+\infty} \ln(1 + \frac{a_i}{b_i}) = +\infty.$$

La loi de Pitman-Yor $\mathcal{P}_{(\mu,\alpha)}$ de paramètres $\mu \in (0,1)$, $\alpha > -\mu$ est une loi de Dirichlet généralisée de paramètres

$$a_k = 1 - \mu, \quad b_k = \alpha + k\mu, \quad k = 1, 2, \ldots.$$

L'estimateur de (PYM) est donné par,

$$\hat{H}_{PYM} = \mathbf{E}[H|\mathbf{x}] = \int \mathbf{E}[H|\mathbf{x},\mu,\alpha]\frac{p(\mathbf{x}|\mu,\alpha)p(\mu,\alpha)}{p(\mathbf{x})}d(\mu,\alpha) \tag{2.6.1}$$

avec

$$\mathbf{E}[H|\mathbf{x},\alpha,\mu] = \psi_0(\alpha+N+1) - \frac{\alpha+K\mu}{\alpha+N}\psi_0(1-\mu) - \frac{1}{N+1}[\sum_{i=1}^{K}(n_i-\mu)\psi_0(n_i-\mu+1)] \tag{2.6.2}$$

où ψ_0 est la fonction digamma (la dérivée logarithmique de la fonction gamma i.e. $\psi_0(z) := \frac{\Gamma'(z)}{\Gamma(z)}$), et

$$p(\mathbf{x}|\mu,\alpha) = \frac{(\prod_{l=1}^{K-1}(\alpha+l\mu))(\prod_{i=1}^{K}\Gamma(n_i-\mu))\Gamma(1+\alpha)}{\Gamma(1-\mu)^K\Gamma(\alpha+N)}. \tag{2.6.3}$$

où K est le nombre de valeurs distinctes. L'expression de la loi a priori $p(\mu,\alpha)$ et le code Matlab permettant de calculer l'entropie de pitman-Yor sont données dans [1] où $(\mu,\alpha) \in [0,1]\times[0,+\infty[$.

2.7 Illustration numérique

Les paramètres statistiques des variables $(s_l, l = 0,1,2,3)$ sont donnés dans les tableaux 2.1 et 2.2. Le tableau 2.1 correspond à la matrice **A** tirée selon la loi de Bernoulli. Le tableau 2.2 correspond à la matrice **A** tirée selon la loi normale.

	Moyenne	Variance	Skewness	Kurtosis	min	max	H_{plug}	H_{PYM}
s_0	55.5120	348.8147	32.3049	-3.1517	7	206	2.6237	2.6237
s_1	235.4740	18476	1.7536	4.9380	130	584	2.3670	2.3764
s_2	264.0400	16172	1.4344	5.4718	27	701	3.5709	3.6126
s_3	512.8590	46098	0.4631	2.4956	40	982	3.8488	3.9032

TABLE 2.1 – Cas Bernoulli : $N = 1000$, $\alpha = 1$, $\sigma^2 = 1$, H_{plugin} := Estimateur de Plug-in et H_{PYM} := Estimateur de Pitman-Yor.

	Moyenne	Variance	Skewness	Kurtosis	min	max	H_{plug}	H_{PYM}
s_0	21.0960	78.9257	2.3821	9.2188	3	51	1.2428	1.2513
s_1	38.8180	658.7516	1.2621	3.1751	3	95	2.3389	2.3509
s_2	64.1730	11527	1.8240	5.3509	3	151	2.3265	2.337
s_3	91.3610	10710	0.2220	2.0832	3	149	2.5089	2.5220

TABLE 2.2 – Cas Gaussien : $N = 1000$, $\alpha = 1$, $\sigma^2 = 1$, H_{plugin} := Estimateur de Plug-in et H_{PYM} := Estimateur de Pitman-Yor.

Bilan :

En analysant les résultats des tableaux (2.1) et (2.2), et en particulier les moyennes de la suite $(s_l : l = 0, 1, 2, 3, ...)$, nous constatons que FISTA est plus rapide dans le cas gaussien que dans le cas de Bernoulli. Les variances dans le cas gaussien sont plus petites que dans le cas de Bernoulli. Le skewness dans les deux cas sont penchés vers la droite. Le cas de Bernoulli est plus penché que le cas gaussien. Les étendues sont plus grandes dans le cas de Bernoulli que dans le cas gaussien. Finalement, les entropies dans le cas de Bernoulli sont supérieures aux entropies dans le cas gaussien. Ceci implique que FISTA contient plus d'aléa dans le cas de Bernoulli que le cas gaussien. Ceci est peut être dû au fait que la matrice $\mathbf{A}$ tirée selon la loi de Bernoulli contient plus d'aléa que dans le cas gaussien.

Bibliographie

[1] E. Archer, I.M. Park and J.W. Pillow, Bayesian entropy estimation for countable discrete distributions, Journal of Machine Learning Research vol. 15 833–2868 (2014).

[2] A. Beck, M. Teboulle, A Fast Iterative Shrinkage-thresholding algorithm for linear inverse problem, SIAM J. Imaging Sci. 183–202 (2009).

[3] A. Chambolle, C. Dossal, On the convergence of the iterates of FISTA. 2014 <hal-01060130v3>.

[4] Y. Nesterov, A method for solving the convex programming problem with convergence rate $O(1/k^2)$, Dokl. Akad. Nauk SSSR vol. 269 no. 3 543–547 (1983).

[5] B.T. Polyak, Some methods of speeding up the convergence of iteration methods, USSR. Comput. Math. Math. Phys. vol. 4 no. 5 1–17 (1964).

[6] P. Tseng, On accelerated proximal gradient methods for convex-concave optimization, Technical report (2008).

Chapitre 3

Intégration numérique du LASSO bayésien

3.1 LASSO bayésien

Dans ce chapitre nous proposons de calculer numériquement l'estimateur de Bayes

$$\int_{\mathbb{R}^p} \mathbf{x} \exp\Big(- \frac{\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2}{2} - \|\mathbf{x}\|_1 \Big) \frac{d\mathbf{x}}{Z}$$

en utilisant les méthodes QMC (quasi Monte Carlo), MC (Monte Carlo).

3.2 Méthode Quasi-Monte Carlo sur $[0,1]^p$: Rappels

La méthode de Quasi-Monte Carlo consiste à approximer

$$\int_{[0,1]^p} F(\mathbf{u}) d\mathbf{u}$$

par la moyenne empirique

$$\frac{\sum_{k=1}^N F(\mathbf{u}_k)}{N}$$

où $\mathbf{u}_1, \ldots, \mathbf{u}_N$ est une suite de points déterministes uniformément distribués sur $[0,1]^p$. Nous rappelons qu'une suite $(u_1, \ldots, u_N)$ est dite uniformément distribuée sur $[0,1]^p$ si sa discrépance

$$D^*(\mathbf{u}_1, \ldots, \mathbf{u}_N) = \sup_{\mathbf{t} \in [0,1]^p} |\frac{\sum_{k=1}^N \mathbf{1}_{[\mathbf{u}_k \leq \mathbf{t}]}}{N} - \prod_{i=1}^p t_i|$$

tend vers 0 lorsque le nombre N de points tend vers $+\infty$.

Si F est absolument continue au sens de Hardy et Krause alors l'erreur d'approximation (voir par exemple [4] pour plus de détails)

$$|\int_{[0,1]^p} F(\mathbf{u})d\mathbf{u} - \frac{\sum_{k=1}^N F(\mathbf{u}_k)}{N}| \leq V(F)D^*(u_1,\ldots,u_N),$$

où $V(F)$ désigne la variation totale de F au sens de Hardy et Krause. Le cas

$$F(\mathbf{u}) = \int_{\mathbf{v}\leq u} \partial^p_{v_1\ldots v_p} F(\mathbf{v})d\mathbf{v}, \quad \forall\, \mathbf{u} \in [0,1]^p,$$

implique

$$V(F) = \int_{[0,1]^p} |\partial^p_{v_1\ldots v_p} F(\mathbf{v})|d\mathbf{v}.$$

Dans le cas de la dimension $p = 1$, si F est absolument continue, alors l'erreur d'approximation [3]

$$|\int_0^1 F(u)du - \frac{\sum_{k=1}^N F(u_k)}{N}| \leq \int_0^1 |F'(u)|duD^*(u_1,\ldots,u_N),$$

de plus la discrépance

$$D^*(u_1,\ldots,u_N) = \sup_{t\in[0,1]} |\frac{\sum_{k=1}^N \mathbf{1}_{[u_k\leq t]}}{N} - t|,$$

de toute suite $u_1, \ldots, u_N \in [0,1]$, vérifie

$$D^*(u_1,\ldots,u_N) \geq \frac{1}{2N}.$$

La borne minimale est atteinte pour les points

$$u_i = \frac{2i-1}{2N}, \quad i = 1,\ldots,N.$$

Dans le cas de la dimension $p \geq 2$, il semble que les points $\boldsymbol{h}_1, \ldots, \boldsymbol{h}_N$ de Hammersley [4] ont la plus petite discrépance

$$D^*(\boldsymbol{h}_1,\ldots,\boldsymbol{h}_N) \leq C_p \frac{\ln(N)^{p-1}}{N}$$

parmi les points déterministes. Observons que la suite $\dfrac{\ln(N)^{p-1}}{N}$ est décroissante seulement pour $N \geq exp(p-1)$.

Heinrich et al. [1] ont montré

$$\inf\{D^*(\mathbf{u}_1,\dots,\mathbf{u}_N): \quad \mathbf{u}_1,\dots,\mathbf{u}_N \in [0,1]^p\} \leq C\sqrt{\frac{p}{N}},$$

où C est une constante ne dépendant ni de p ni de N. Les suites minimisantes ne sont pas déterministes. Elles sont des échantillons de la loi uniforme sur $[0,1]^p$. Nous avons pour une telle suite l'erreur d'approximation

$$|\int_{[0,1]^p} F(\mathbf{u})d\mathbf{u} - \frac{\sum_{k=1}^N F(\mathbf{u}_k)}{N}| \leq CV(F)\sqrt{\frac{p}{N}}. \tag{3.2.1}$$

3.3 Méthode Quasi-Monte Carlo sur $\mathbb{R}^p$

3.3.1 Le cas $p = 1$

Soient f une fonction appartenant à un bon espace de fonctions et $\rho > 0$ une densité de probabilité continue sur $\mathbb{R}$ et différentiable par morceaux. Nous voulons ramener l'intégrale

$$\int_{\mathbb{R}} f(x)\rho(x)dx$$

à une intégrale sur $[0,1]$. Nous posons le changement de variable

$$u := G(x) = \int_{-\infty}^{x} \rho(t)dt,$$

et son inverse $x = G^{-1}(u)$. Nous posons

$$F(u) = f(G^{-1}(u)),$$

et alors

$$\int_{\mathbb{R}} f(x)\rho(x)dx = \int_0^1 F(u)du.$$

Si F est absolument continue, alors l'erreur d'approximation

$$|\int_0^1 F(u)du - \frac{\sum_{k=1}^N F(u_k)}{N}| \leq \int_0^1 |F'(u)|duD^*(u_1,\dots,u_N).$$

Si f est continue, dérivable par morceaux avec f' intégrable, alors F est aussi absolument continue, avec F' intégrable sur $[0,1]$. En effet

$$u \in (0,1) \to F(u) = f(G^{-1}(u))$$

est continue, dérivable par morceaux avec

$$F'(u) = \frac{f'(G^{-1}(u))}{G'(G^{-1}(u))}, \quad \forall\, u \in (0,1). \tag{3.3.1}$$

De plus

$$\int_0^1 |F'(u)|du = \int_0^1 |\frac{f'(G^{-1}(u))}{G'(G^{-1}(u))}|du \tag{3.3.2}$$

$$= \int_{-\infty}^{+\infty} |\frac{f'(x)}{G'(x)}|\rho(x)dx \tag{3.3.3}$$

$$= \int_{-\infty}^{+\infty} |f'(x)|dx < +\infty. \tag{3.3.4}$$

Nous résumons cette estimation dans la proposition suivante.

Proposition 3.3.1. *Si f' est intégrable, alors*

$$|\int_{\mathbb{R}} f(x)\rho(x)dx - \frac{\sum_{k=1}^N f(x_k)}{N}| \leq \int_{\mathbb{R}} |f'(x)|dx D^*(G(x_1), \ldots, G(x_N)).$$

3.3.2 Le cas $p \geq 2$

Si $p \geq 2$ et si la densité

$$\rho(\mathbf{x}) = \prod_{i=1}^p \rho_i(x_i)$$

alors

$$\int_{\mathbb{R}^p} f(\mathbf{x}) \prod_{i=1}^p \rho_i(x_i)dx_i = \int_{[0,1]^p} F(\mathbf{u})d\mathbf{u},$$

où

$$F(\mathbf{u}) = f(G_1^{-1}(u_1), \ldots, G_p^{-1}(u_p)), \quad \forall\, u_1, \ldots, u_p \in (0,1)^p,$$

et $G(\mathbf{x}_i) = (G_1(x_{i1}, \ldots, G_p(x_{ip}))$ et $G_1, \ldots, G_p$ sont les fonctions de répartitions des densités $\rho_1, \ldots, \rho_p$.

L'hypothèse

$$F(\mathbf{u}) = \int_{[0 \leq v_1 \leq u_1, \ldots, 0 \leq v_p \leq u_p]} \partial^p_{v_1 \ldots v_p} F(\mathbf{v})d\mathbf{v},$$

le lemme de Fubini, l'égalité (3.3.1) et (3.3.4) impliquent

$$f(\mathbf{x}) = \int_{[z_1 \leq x_1, \ldots, z_p \leq x_p]} \partial^p_{z_1 \ldots z_p} f(\mathbf{z}) d\mathbf{z}, \tag{3.3.5}$$

$$\int_{[0,1]^p} |\partial^p_{u_1 \ldots u_p} F(u_1, \ldots, u_p)| d\mathbf{u} = \int_{\mathbb{R}^p} |\partial^p_{x_1 \ldots x_p} f(\mathbf{x})| d\mathbf{x}.$$

Nous résumons cette estimation dans la proposition suivante.

Proposition 3.3.2. *Sous (3.3.5) et la condition*

$$\int_{\mathbb{R}^p} |\partial^p_{x_1 \ldots x_p} f(\mathbf{x})| dx < +\infty, \tag{3.3.6}$$

nous avons pour toute suite $\mathbf{x}_1 = (x_{11}, \ldots, x_{1p}), \ldots, \mathbf{x}_N = (x_{N1}, \ldots, x_{Np})$ *de points de* $\mathbb{R}^p$*, l'estimation suivante*

$$\Big| \int_{\mathbb{R}^p} f(\mathbf{x}) \rho(\mathbf{x}) d\mathbf{x} - \frac{\sum_{k=1}^N f(\mathbf{x}_k)}{N} \Big| \leq \int_{\mathbb{R}^p} |\partial^p_{x_1 \ldots x_p} f(\mathbf{x})| d\mathbf{x} D^*(G(\mathbf{x}_1), \ldots, G(\mathbf{x}_N)), \tag{3.3.7}$$

où $G(\mathbf{x}_i) = (G_1(x_{i1}, \ldots, G_p(x_{ip}))$ *et* $G_1, \ldots, G_p$ *sont les fonctions de répartition des densités* $\rho_1, \ldots, \rho_p$*.*

Voir [2] pour une étude générale.

3.4 Applications

En utilisant les méthodes numériques des sections 2 et 3 nous pouvons calculer notre estimateur de Bayes en calculant séparement les numérateurs

$$\int_{\mathbb{R}^p} x_i \exp\Big(- \frac{\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2}{2} - \|\mathbf{x}\|_1 \Big) d\mathbf{x}, \quad i = 1, \ldots, p,$$

et le dénominateur

$$Z = \int_{\mathbb{R}^p} \exp\Big(- \frac{\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2}{2} - \|\mathbf{x}\|_1 \Big) d\mathbf{x}.$$

Nous allons comparer ces méthodes d'intégration numérique lorsque les numérateurs et le dénominateur sont connus.

En prenant $\mathbf{y} = 0$, alors les numérateurs

$$b_i := \int_{\mathbb{R}^p} x_i \exp\Big(- \frac{\|\mathbf{A}\mathbf{x}\|_2^2}{2} - \|\mathbf{x}\|_1 \Big) d\mathbf{x} = 0,$$

avec $i = 0, 1, \ldots, p$.

Nous remarquons que pour chaque i et $\lambda \in (0, 1]$

$$b_i = \frac{2^p}{(\lambda)^p} \int_{\mathbb{R}^p} x_i \exp\Big(- \frac{\|\mathbf{A}\mathbf{x}\|_2^2}{2} - (1-\lambda)\|\mathbf{x}\|_1\Big)\rho_\lambda(\mathbf{x})d\mathbf{x},$$

où la densité

$$\rho_\lambda(\mathbf{x}) := \frac{(\lambda)^p}{2^p} \exp(-\lambda\|\mathbf{x}\|_1).$$

Nous proposons pour chaque $\lambda \in (0, 1]$ et pour chaque suite $(\mathbf{x}_k)$

$$\frac{\sum_{k=1}^N f_i(\lambda, \mathbf{x}_k)}{N} = \hat{b}_i$$

comme estimateur de b_i, où

$$f_i(\lambda, \mathbf{x}) = \frac{2^p}{(\lambda)^p} x_i \exp\Big(- \frac{\|\mathbf{A}\mathbf{x}\|_2^2}{2\sigma^2} - (1-\lambda)\|\mathbf{x}\|_1\Big),$$

et les points $\mathbf{x}_k = G^{-1}(\mathbf{u}_k)$ avec G désignant la fonction de répartition de ρ_λ.

L'erreur d'estimation pour $\lambda \in (0, 1)$ est contrôlée par

$$|b_i - \frac{\sum_{k=1}^N f_i(\lambda, \mathbf{x}_k)}{N}| \leq \int_{\mathbb{R}^p} |\partial^p_{x_1 \ldots x_p} f_i(\lambda, \mathbf{x})| d\mathbf{x} D_N^*(G(\mathbf{x}_1), \ldots, G(\mathbf{x}_N)),$$

Pour $\lambda = 1$ la densité ρ_1 ne garantit pas le contrôle de l'erreur

$$|b_i - \frac{\sum_{k=1}^N f_i(1, \mathbf{x}_k)}{N}|,$$

car la fonction $\mathbf{x} \to f_i(1, \mathbf{x})$ ne vérifie pas l'hypothèse (3.3.6).

En calculant la somme des carrés des erreurs

$$Err(\lambda) := \sum_{i=1}^p |b_i - \hat{b}_i|^2,$$

les résultats numériques obtenus, en utilisant une matrice $\mathbf{A}$ tirées selon la loi de Bernoulli, montrent que le meilleur choix est $\lambda = 1$ parmi les choix $\lambda = 0.1$, $\lambda = 0.5$, $\lambda = 0.8$ et $\lambda = 1$. Plus précisément, nous avons

$$\arg \min_{\lambda = 0.1, 0.5, 0.8, 1} Err(\lambda) = 1.$$

Les résultats sont reportés dans les tableaux 3.1, 3.2 et 3.3.

Il ressort de ces tableaux que les points de Hammersely sont les meilleurs.

	$\hat{b}_1$	$\hat{b}_2$	$\hat{b}_3$	$\hat{b}_4$	$\hat{b}_5$	$\hat{b}_6$	$\hat{b}_7$	$Err(\lambda)$
$\lambda = 0.1$	-1.7827	-0.0075	0.7460	-0.7850	0.0995	-1.7005	2.4916	3.6689
$\lambda = 0.5$	-0.1780	-0.2329	0.3085	-0.0972	-0.1137	-0.1904	0.3249	0.5877
$\lambda = 0.8$	0.0461	0.0387	-0.0419	0.0023	-0.0338	-0.0149	0.0827	0.1166
$\lambda = 1$	0.0752	0.0599	-0.0812	0.0165	-0.0435	-0.0248	0.0519	0.1160

TABLE 3.1 – Uniforme : $N = 10^5$, $p = 7$, $n = 4$.

	$\hat{b}_1$	$\hat{b}_2$	$\hat{b}_3$	$\hat{b}_4$	$\hat{b}_5$	$\hat{b}_6$	$\hat{b}_7$	$Err(\lambda)$
$\lambda = 0.1$	-0.1355	0.0335	-0.2021	0.1084	0.2540	0.0152	-0.2324	0.4369
$\lambda = 0.5$	-0.3267	-0.0225	0.0757	0.0216	0.0072	-0.2294	0.0421	0.4097
$\lambda = 0.8$	-0.0574	0.0428	-0.0382	-0.0067	-0.0087	-0.0524	0.0402	0.1052
$\lambda = 1$	-0.0222	0.0274	-0.0449	0.0163	-0.0280	-0.0364	0.0215	0.0781

TABLE 3.2 – Halton : $N = 10^5$, $p = 7$, $n = 4$.

	$\hat{b}_1$	$\hat{b}_2$	$\hat{b}_3$	$\hat{b}_4$	$\hat{b}_5$	$\hat{b}_6$	$\hat{b}_7$	$Err(\lambda)$
$\lambda = 0.1$	0.1030	-0.0029	0.0726	-0.0877	-0.1492	0.1532	0.2320	0.3509
$\lambda = 0.5$	0.0159	-0.0491	0.0463	0.0743	-0.0351	-0.0483	0.0524	0.1290
$\lambda = 0.8$	0.0119	0.0070	-0.0016	0.0198	-0.0023	0.0026	0.0226	0.0333
$\lambda = 1$	0.0139	0.0039	-0.0052	0.0150	-0.0001	0.0125	0.0072	0.0259

TABLE 3.3 – Hammersley : $N = 10^5$, $p = 7$, $n = 4$.

Nous terminons ce chapitre par la comparaison de ces méthodes pour le calcul des numérateurs et du dénominateur (fonction de partition Z). Le choix $\mathbf{y} = 0$ et $\mathbf{A}$ gaussienne ou bien Bernoulli nous ne permet pas d'avoir une valeur explicite de Z. En revanche le choix $\mathbf{y} = 0$ et $\mathbf{A}$ ayant un seul coefficient non nul ($a_{11} = 1$) donne

$$Z = 2^p \sqrt{2e\pi}(1 - F(1)) = 83.9270,$$

où F désigne la fonction de répartition de la loi $\mathcal{N}(0, 1)$.

Le tableau 3.4 montre comment chaque méthode converge vers la vraie valeur de Z. Nous avons besoin de $N = 10^6$ points de Halton et Hammer-

sely pour la convergence. Les points de Halton et de Hammersely sortent gagnants contre les points uniformes.

	$Z_{Uniforme}$	Z_{Halton}	$Z_{Hammersley}$
$N = 10^2$	88.9405	84.8299	84.7579
$N = 10^4$	83.6148	83.9318	83.9354
$N = 10^6$	83.9309	83.9270	83.9270

TABLE 3.4 – $p = 7$, $n = 4$, $\mathbf{y} = 0$, $\lambda = 1$.

Le tableau 3.5 représente les calculs numériques des numérateurs. Il confirme que les points de Hammersely sont les meilleurs.

	b_1	b_2	b_3	b_4	b_5	b_6	b_7	$Err(\lambda)$
Uniforme	0.0090	-0.3085	-0.0067	0.0651	-0.1099	0.0447	-0.0929	0.3496
Halton	0.0000	-0.0040	-0.0042	-0.0061	-0.0091	-0.0082	-0.0118	0.0190
Hammer	-0.0000	-0.0016	-0.0018	-0.0006	-0.0036	-0.0028	-0.0046	0.0069

TABLE 3.5 – $N = 10^6$, $p = 7$, $n = 4$.

Bibliographie

[1] S. Heinrich, E. Novak, G.W. Wasilkowski and H. Wozniakowski, The inverse of the star-discrepancy depends linearly on the dimension, Acta Arithmetica (2000).

[2] F.J. Hickernell, I.H. Shoan and G.W. Wasilkowski, On tractability of weighted integration over bounded and unbounded regions in $\mathbb{R}^s$, Math. Comput. vol. 73 no. 248 1885–1901 (2004).

[3] H. Morohosi and M. Fushimi, A Practical approach to the error estimation of quasi Monte Carlo integrations, Department ofbMathematical Engineering and Information Physics, Graduate School of Engineering, University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo, 113–8656, Japan.

[4] H. Niederreiter, Random Number generation and quasi-Monte Carlo methods, Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania 1992.

Chapitre 4

Les fonctions cylindre parabolique, gamma incomplète et lasso

Dans ce chapitre nous étudions, pour $\mathbf{A}$ et $\mathbf{y}$ fixés, la simulation exacte de la densité du lasso bayésien

$$c(\mathbf{x})d\mathbf{x} = \frac{1}{Z_p}\exp\{-f(\mathbf{x})\}d\mathbf{x} \tag{4.0.1}$$

où

$$f(\mathbf{x}) := \frac{\|\mathbf{A}\mathbf{x}-\mathbf{y}\|_2^2}{2} + \|\mathbf{x}\|_1, \quad \mathbf{x}\in\mathbb{R}^p. \tag{4.0.2}$$

Plus précisément, si $\mathbf{x}$ est tiré selon la densité $c(\mathbf{x})d\mathbf{x}$, nous allons simuler les coordonnées polaires

$$\mathbf{s} = \frac{\mathbf{x}}{\|\mathbf{x}\|}\in\mathcal{S}, \quad r = \|\mathbf{x}\|,$$

où $\|\cdot\|$ désigne l'une des normes l^2 ou l^1 sur $\mathbb{R}^p$, $\mathcal{S}$ désigne la sphère unité de la norme $\|\cdot\|$ et $d\mathbf{s}$ sa mesure de Lebesgue. Nous exprimons $c(\mathbf{x})d\mathbf{x}$ à l'aide de la formule de changement de variables

$$c(\mathbf{x})d\mathbf{x} = \frac{1}{Z_p(\mathbf{s})}\exp\{-f(r\mathbf{s})\}r^{p-1}drd\mathbf{s}\frac{Z_p(\mathbf{s})}{Z_p}, \tag{4.0.3}$$

où $|\mathcal{S}|$ désigne la mesure de Lebesgue de $\mathcal{S}$, et la fonction de partition

$$Z_p(\mathbf{s}) = \int_0^{+\infty}\exp\{-f(r\mathbf{s})\}r^{p-1}dr. \tag{4.0.4}$$

Nous exprimons la fonction de partition (4.0.4) de la loi

$$c(r,\mathbf{s})dr = \frac{1}{Z_p(\mathbf{s})}\exp\{-f(r\mathbf{s})\}r^{p-1}dr \tag{4.0.5}$$

en utilisant la fonction cylindre parabolique. Puis nous calculons pour chaque angle $\mathbf{s}$ fixé la fonction de répartition de la densité (4.0.5). Nous donnons aussi une inégalité de concentration et une interprétation géométrique de la fonction de partition Z_p (4.0.1).

Remarque 4.0.1. Le calcul de Z_p est donné par la relation suivante :

$$Z_p = \int_{\mathcal{S}} Z_p(\mathbf{s})d\mathbf{s}. \tag{4.0.6}$$

4.1 Simulation de l'angle

Les notations $\mathcal{S}_{p-1,2}$, $\mathcal{S}_{p-1,1}$ désignent les sphères unités de l'espace $\mathbb{R}^p$ pour les normes l^2 et l^1 respectivement. Un élément générique de $\mathcal{S}_{p-1,2}$ est noté par θ et $d\theta$ désigne la mesure uniforme sur $\mathcal{S}_{p-1,2}$. Un élément générique de $\mathcal{S}_{p-1,1}$ est noté par ω et $d\omega$ désigne la mesure uniforme sur $\mathcal{S}_{p-1,1}$.

Pour tirer uniformément un point $\theta \in \mathcal{S}_{p-1,2}$ nous utilisons la loi gaussienne. Nous tirons un vecteur $\mathbf{x}$ selon la loi gaussienne $\mathcal{N}(0,\mathbf{I}_{p\times p})$ puis nous formons le point

$$\theta = \frac{\mathbf{x}}{\|\mathbf{x}\|_2}.$$

Pour tirer uniformément un point $\omega \in \mathcal{S}_{p-1,1}$ nous utilisons la loi de Bernoulli et la loi de Dirichlet. D'abord nous remarquons qu'il y a 2^p faces :

$$\sum_{i=1}^{p}\varepsilon_i\omega_i = 1, \quad \varepsilon_i = sgn(\omega_i), \quad i = 1,\dots,p.$$

Nous tirons un p-échantillons $\varepsilon_1,\dots,\varepsilon_p$ i.i.d. des signes $\pm$. Puis nous tirons une distribution de probabilités

$$\sum_{i=1}^{p}\pi_i = 1, \quad \pi_1,\dots,\pi_p \geq 0$$

en utilisant la loi de Dirichlet uniforme. Nous obtenons ainsi la réalisation

$$(\varepsilon_1\pi_1,\dots,\varepsilon_p\pi_p)$$

d'un élément de $\mathcal{S}_{p-1,1}$.

4.2 Fonction cylindre parabolique et la fonction de partition

Nous prolongeons la fonction $\mathbf{s} \in \mathcal{S} \to Z_p(\mathbf{s})$ à

$$\mathbf{x} \in \mathbb{R}^p \to Z_p(\mathbf{x}) = \int_0^{+\infty} \exp\{-f(r\mathbf{x})\} r^{p-1} dr. \tag{4.2.1}$$

Ce prolongement est homogène d'ordre $-p$, c'est-à-dire

$$Z_p(\lambda \mathbf{x}) = \lambda^{-p} Z_p(\mathbf{x}), \quad \forall\, \lambda > 0, \mathbf{x} \in \mathbb{R}^p. \tag{4.2.2}$$

Si $\mathbf{A}\mathbf{x} = 0$, alors

$$f(r\mathbf{x}) = \frac{\|\mathbf{y}\|_2^2}{2} + r\|\mathbf{x}\|_1,$$

et si en plus $\mathbf{x} \neq 0$, alors

$$Z_p(\mathbf{x}) = (p-1)! \|\mathbf{x}\|_1^{-p} \exp(-\frac{\|\mathbf{y}\|_2^2}{2}).$$

Si $\mathbf{A}\mathbf{x} \neq 0$, alors nous allons exprimer $Z_p(\mathbf{x})$ à l'aide de la fonction cylindre parabolique.

Nous rappelons [18] que pour $\nu \in \mathbb{R}$ la fonction $U(\nu, z)$ solution du problème

$$\frac{d^2 W(z)}{d^2 z} = (\nu + \frac{z^2}{4}) W(z), \tag{4.2.3}$$

$$U(\nu, z) = z^{-\nu - \frac{1}{2}} \exp(-\frac{z^2}{4})[1 + O(z^{-2})], \quad \text{lorsque} \quad z \to +\infty, \tag{4.2.4}$$

est la fonction cylindre parabolique. Une autre notation que nous trouvons dans la littérature est

$$D_b(z) = U(-b - \frac{1}{2}, z), \quad \forall\, b \in \mathbb{R}, z \geq 0. \tag{4.2.5}$$

Nous avons à partir de (4.2.4)

$$D_b(z) = z^b \exp(-\frac{z^2}{4})[1 + O(z^{-2})], \tag{4.2.6}$$

lorsque $z \to +\infty$.

Nous rappelons aussi la représentation intégrale de Erdélyi [4] pour la fonction cylindre parabolique

$$\exp(\frac{z^2}{4}) \Gamma(\nu) D_{-\nu}(z) = \int_0^{+\infty} \exp(-\frac{1}{2} r^2 - zr) r^{\nu - 1} dr, \quad \nu > 0,$$

où $\Gamma(\nu) = \int_0^{+\infty} \exp(-t) t^{\nu-1} dt$ désigne la fonction Γ.

Proposition 4.2.1. *La variable*

$$\omega_{lasso} := \frac{\|\mathbf{x}\|_1 - \langle \mathbf{A}\mathbf{x}, \mathbf{y}\rangle}{\|\mathbf{A}\mathbf{x}\|_2} \tag{4.2.7}$$

va jouer un rôle important. Elle ne dépend que de $\omega = \frac{\mathbf{x}}{\|\mathbf{x}\|_1} \in \mathcal{S}_{p-1,1}$ *et la fonction*

$$\omega \in \mathcal{S}_{p-1,1} \to \omega_{lasso}(\omega)$$

est minorée par

$$\lambda_{1,2} := \min\{\frac{1}{\|\mathbf{A}\omega\|_2} - \frac{\langle \mathbf{A}\omega, \mathbf{y}\rangle}{\|\mathbf{A}\omega\|_2} : \quad \omega \in \mathcal{S}_{p-1,1}\}.$$

Nous sommes maintenant prêt à exprimer $Z_p(\mathbf{x})$ à l'aide de D_{-p}. Si $\mathbf{A}\mathbf{x} \neq 0$, alors

$$\begin{aligned} f(r\mathbf{x}) &= \frac{\|\mathbf{y}\|_2^2}{2} + r(\|\mathbf{x}\|_1 - \langle \mathbf{A}\mathbf{x}, \mathbf{y}\rangle) + \frac{\|\mathbf{A}\mathbf{x}\|_2^2}{2} r^2 \\ &= \frac{\|\mathbf{y}\|_2^2}{2} + r\|\mathbf{A}\mathbf{x}\|_2 \omega_{lasso} + \frac{\|\mathbf{A}\mathbf{x}\|_2^2}{2} r^2. \end{aligned} \tag{4.2.8}$$

Par conséquent

$$\begin{aligned} Z_p(\mathbf{x}) &= \exp(-\frac{\|\mathbf{y}\|_2^2}{2}) \int_0^{+\infty} \exp\{-\frac{\|\mathbf{A}\mathbf{x}\|_2^2}{2} r^2 - r\|\mathbf{A}\mathbf{x}\|_2 \omega_{lasso}\} r^{p-1} dr \\ &= \exp(-\frac{\|\mathbf{y}\|_2^2}{2}) \|\mathbf{A}\mathbf{x}\|_2^{-p} \int_0^{+\infty} \exp\{-\frac{1}{2} r^2 - r\omega_{lasso}\} r^{p-1} dr. \end{aligned}$$

Nous résumons ces résultats dans la proposition suivante.

Proposition 4.2.2. *Nous avons pour* $\mathbf{A}\mathbf{x} \neq 0$*,*

$$Z_p(\mathbf{x}) = (p-1)! \exp(-\frac{\|\mathbf{y}\|_2^2}{2}) \|\mathbf{A}\mathbf{x}\|_2^{-p} \exp(\frac{\omega_{lasso}^2}{4}) D_{-p}(\omega_{lasso}).$$

Si $\mathbf{A}\mathbf{x} \to 0$*, alors* $\omega_{lasso} \to +\infty$ *et*

$$Z_p(\mathbf{x}) = (p-1)! \exp(-\frac{\|\mathbf{y}\|_2^2}{2}) \|\mathbf{x}\|_1^{-p} [1 + O(\omega_{lasso}^{-2})].$$

N.B. Pour des valeurs modérées de z le calcul numérique de $U(\nu, z)$ utilise l'approximation suivante [2] :

$$U(\nu, z) = D_{-\nu-\frac{1}{2}}(z) = Y_1 \cos(\beta) - Y_2 \sin(\beta),$$

où

$$\beta = \pi(\frac{\nu}{2} + \frac{1}{4}),$$

$$\mathbf{Y}_1 = \frac{y_1 \Gamma(\frac{1}{4} - \frac{\nu}{2})}{\sqrt{\pi} 2^{\frac{1}{4}+\frac{\nu}{2}}}, \qquad \mathbf{Y}_2 = \frac{y_2 \Gamma(\frac{3}{4} - \frac{\nu}{2})}{\sqrt{\pi} 2^{\frac{\nu}{2}-\frac{1}{4}}}$$

avec,

$$y_1 = 1 + \nu\frac{z^2}{2!} + (\nu^2 + \frac{1}{2})\frac{z^4}{4!} + (\nu^3 + \frac{7\nu}{2})\frac{z^6}{6!} + (\nu^4 + 11\nu^2 + \frac{15}{4})\frac{z^8}{8!} + \ldots,$$

$$y_2 = 1 + \nu\frac{z^3}{3!} + (\nu^2 + \frac{3}{2})\frac{z^5}{5!} + (\nu^3 + \frac{13\nu}{2})\frac{z^7}{7!} + (\nu^4 + 17\nu^2 + \frac{63}{4})\frac{z^9}{9!} + \ldots.$$

Pour les grandes valeurs de z le calcul de $U(\nu, z)$ utilise l'approximation donnée par (4.2.6).

Corollaire 4.2.3. *Si* $\mathbf{y} = 0$, *alors* $\omega_{lasso} = \frac{1}{\|\mathbf{A}\omega\|_2}$ *est minorée par* $\frac{1}{\lambda_{1,2}}$, *où* $\lambda_{1,2} = \max(\|\mathbf{A}\omega\|_2 : \omega \in \mathcal{S}_{p-1,1})$ *est la norme de l'opérateur* $\mathbf{A} : (\mathbb{R}^p, \|\cdot\|_1) \to (\mathbb{R}^n, \|\cdot\|_2)$. *La fonction de partition*

$$Z_p(\omega) = (p-1)!\omega_{lasso}^p \exp(\frac{\omega_{lasso}^2}{4}) D_{-p}(\omega_{lasso})$$

est une fonction de $\|\mathbf{A}\omega\|_2^2$ *convexe et décroissante.*

Preuve 4.2.4. *Il suffit de remarquer que*

$$Z_p(\omega) = \int_0^{+\infty} \exp\{-\frac{\|\mathbf{A}\omega\|^2}{2} r^2 - r\} r^{p-1} dr.$$

4.3 Calcul de la fonction de répartition

Si $\mathbf{x}$ est tiré selon la densité $c(\mathbf{x})d\mathbf{x}$, alors sachant l'angle $\mathbf{s} = \frac{\mathbf{x}}{\|\mathbf{x}\|}$, la loi de $\|\mathbf{x}\|$ a pour densité $c(r, \mathbf{s})dr$ (4.0.5).

1) Si $\mathbf{As} = 0$, alors $c(r, \mathbf{s})dr$ est la densité de la loi gamma de paramètres $(p, \|\mathbf{s}\|_1)$.

2) Le cas $\mathbf{As} \neq 0$. Nous allons calculer la fonction de répartition de $c(r, \mathbf{s})dr$ en utilisant la fonction gamma incomplète. Les calculs donnent

$$\begin{aligned} c(r, \mathbf{s})dr &= \frac{1}{Z_p(\mathbf{s})} \exp\{-f(r\mathbf{s})\} r^{p-1} dr \\ &= \frac{1}{z_p(\mathbf{s})} \exp\{-\frac{1}{2}(\|\mathbf{As}\|_2 r + \omega_{lasso})^2\} r^{p-1} dr, \end{aligned}$$

où

$$z_p(\mathbf{s}) = \int_0^{+\infty} \exp\{-\frac{1}{2}(\|\mathbf{As}\|_2 r + \omega_{lasso})^2\} r^{p-1} dr.$$

Cette nouvelle fonction de partition ne dépend que des paramètres $a = \|\mathbf{As}\|_2$ et $b = \frac{\|\mathbf{s}\|_1 - \langle \mathbf{As}, \mathbf{y} \rangle}{\|\mathbf{As}\|_2} = \omega_{lasso}$. Si nous introduisons la notation

$$z_p(r, a, b) = \int_0^r \exp\{-\frac{(a\tau + b)^2}{2}\} \tau^{p-1} d\tau, \quad r \geq 0, \tag{4.3.1}$$

alors

$$\int_0^r c(\tau, \mathbf{s}) d\tau = \frac{z_p(r, a, b)}{z_p(a, b)},$$

où

$$z_p(a, b) := z_p(+\infty, a, b). \tag{4.3.2}$$

Remarque 4.3.1. Nous réécrivons pour chaque couple $a > 0$ et $b \in \mathbb{R}$, la fonction

$$z_p(a, b) = \int_0^{+\infty} \exp\{-\frac{1}{2} g_{a,b}(r)\} r^{p-1} dr, \tag{4.3.3}$$

où la fonction

$$g_{a,b}(r) := \frac{(ar + b)^2}{2}, \quad r \geq 0. \tag{4.3.4}$$

Le lien entre $Z_p(\omega)$ (4.0.4) et $z_p(a, b)$ est donné par l'équation

$$Z_p(\omega) = \exp(-\frac{\|\mathbf{y}\|_2^2}{2} + \frac{b^2}{2}) z_p(a, b).$$

Nous en déduisons la relation

$$z_p(a, b) = (p-1)! a^{-p} \exp(-\frac{b^2}{4}) D_{-p}(b), \quad \forall\, a > 0, b \in \mathbb{R}.$$

Maintenant nous allons calculer $z_p(r, a, b)$ en utilisant la fonction gamma incomplète. Le changement de variables $a\tau + b \to \tau$ implique

$$\begin{aligned} z_p(r, a, b) &= a^{-p} \int_b^{ar+b} \exp(-\frac{\tau^2}{2}) (\tau - b)^{p-1} d\tau, \quad r \geq 0, \\ &= a^{-p} \sum_{k=0}^{p-1} \binom{k}{p-1} (-b)^{p-1-k} \int_b^{ar+b} \exp(-\frac{\tau^2}{2}) \tau^k d\tau. \end{aligned}$$

Il nous reste à calculer pour chaque entier k l'intégrale

$$I_k(\alpha) := \int_0^\alpha \exp(-\frac{\tau^2}{2})\tau^k d\tau, \quad \alpha \in \mathbb{R}.$$

Si $\alpha < 0$ alors

$$\begin{aligned} I_k(\alpha) &= -\int_\alpha^0 \exp(-\frac{\tau^2}{2})\tau^k d\tau \\ &= (-1)^{k+1}2^{\frac{k-1}{2}} \int_0^{-\alpha} \exp(-\frac{\tau^2}{2})(\frac{\tau^2}{2})^{\frac{k-1}{2}} d\frac{\tau^2}{2} \\ &= (-1)^{k+1}2^{\frac{k-1}{2}} \int_0^{-\alpha} \exp(-\tau)\tau^{\frac{k-1}{2}} d\tau \\ &= (-1)^{k+1}2^{\frac{k-1}{2}} \gamma(\frac{k+1}{2}, -\alpha), \end{aligned}$$

où $\gamma(a,x) = \int_0^x \exp(-t)t^{a-1}dt$ désigne la fonction gamma incomplète inférieure. Si $\alpha > 0$, alors

$$I_k(\alpha) = 2^{\frac{k-1}{2}} \gamma(\frac{k+1}{2}, \alpha).$$

Finalement

$$z_p(r,a,b) = \sum_{k=0}^{p-1} \binom{k}{p-1} (-b)^{p-1-k} [I_k(ar+b) - I_k(b)]. \tag{4.3.5}$$

Proposition 4.3.2. *1) Si $a > 0$ et $b \geq 0$, alors*

$$\begin{aligned} z_p(r,a,b) &= a^{-p} \sum_{k=0}^{p-1} \binom{k}{p-1} (-b)^{p-1-k} 2^{\frac{k-1}{2}} \\ &\quad \{\gamma(\frac{k+1}{2}, ar+b) - \gamma(\frac{k+1}{2}, b)\} \qquad (4.3.6) \\ &= a^{-p} \sum_{k=0}^{p-1} \binom{k}{p-1} (-b)^{p-1-k} 2^{\frac{k-1}{2}} \\ &\quad \{\Gamma(\frac{k+1}{2}, b) - \Gamma(\frac{k+1}{2}, ar+b)\}, \qquad (4.3.7) \end{aligned}$$

où

$$\Gamma(a,x) = \int_x^{+\infty} \exp(-t)t^{a-1}dt$$

est la fonction gamma incomplète supérieure.

2) Si $r = +\infty$, $a > 0$ et $b \geq 0$, alors

$$z_p(+\infty, a, b) = a^{-p} \sum_{k=0}^{p-1} \binom{k}{p-1} (-b)^{p-1-k} 2^{\frac{k-1}{2}} \Gamma(\frac{k+1}{2}, b) \quad (4.3.8)$$

$$:= a^{-p} z_p(b). \quad (4.3.9)$$

Maintenant nous donnons une autre formule pour calculer $z_p(r, a, b)$ en utilisant la récurrence sur la dimension p.

Proposition 4.3.3. *Soit F la fonction de répartition de la loi normale. Nous avons*

$$z_1(r, a, b) = \sqrt{\frac{2\pi}{a^2}} \{F(ar + b) - F(b)\},$$

$$z_2(r, a, b) = \frac{1}{a^2} \{\exp(-\frac{b^2}{2}) - \exp(-\frac{(ar + b)^2}{2})\} - z_1(r, a, b)\frac{b}{a},$$

$$z_p(r, a, b) = -z_{p-1}(r, a, b)\frac{b}{a} + z_{p-2}(r, a, b)\frac{p-2}{a^2} - \frac{r^{p-2}}{a^2} \exp\{-(ar + b)^2/2\}, \quad p \geq 3.$$

Des calculs numériques sont en cours pour comparer les trois méthodes de calculs : Le calcul numérique de la fonction cylindre parabolique, le calcul numérique de la fonction gamma incomplète et le calcul numérique en utilisant la récurrence sur la dimension p.

	gamma incomplète	cylindre parabolique	récurrence
$Z_p(s)$	0.3174	0.3593	0.3826

TABLE 4.1 – $p = 2$, $n = 1$, $\mathbf{y} = 0$ et $\mathbf{A} = (1 \quad 1)$.

4.4 Interprétation géométrique de la fonction de partition

D'abord nous représentons $f(r\mathbf{x})$ pour $\mathbf{Ax} \neq 0$ sous la forme

$$f(r\mathbf{x}) = \frac{\|\mathbf{y}\|_2^2}{2} - \frac{\omega_{lasso}^2}{2} + \frac{(r\|A\mathbf{x}\|_2 + \omega_{lasso})^2}{2}. \quad (4.4.1)$$

La fonction

$$\exp\{-f(\mathbf{x})\}, \quad \forall\, \mathbf{x} \in \mathbb{R}^p$$

est log-concave et intégrable sur $\mathbb{R}^p$. La fonction de partition

$$Z_p(\mathbf{x}) = \int_0^{+\infty} \exp\{-f(r\mathbf{x})\} r^{p-1} dr$$

définie pour chaque $\mathbf{x} \in \mathbb{R}^p$ est homogène d'ordre $-p$ (4.2.2).

Si $\mathbf{x}$ est non nul et $\mathbf{A}\mathbf{x} = 0$, alors

$$\mathbf{x} \in \mathbb{R}^p \to Z_p^{-\frac{1}{p}}(\mathbf{x}) := \frac{\|\mathbf{x}\|_1}{(p-1)!^{\frac{1}{p}}} \exp(\frac{\|\mathbf{y}\|_2^2}{2p}).$$

Par conséquent

$$\mathbf{x} \in Ker(\mathbf{A}) \to Z_p^{-\frac{1}{p}}(\mathbf{x})$$

est une norme. Un résultat général [1] nous dit que

$$\mathbf{x} \in \mathbb{R}^p \to Z_p^{-\frac{1}{p}}(\mathbf{x}) := \|\mathbf{x}\|_c$$

est une quasi-norme sur $\mathbb{R}^p$ (seule la symétrie par rapport à l'origine est manquante).

En exprimant $Z_p(\mathbf{x})$ à l'aide de la fonction cylindre parabolique, la boule unité de $\|\cdot\|_c$ est définie par

$$\begin{aligned}
\mathcal{B}(\mathbf{A}, \mathbf{y}) &:= \{\mathbf{x} \in \mathbb{R}^p : \|\mathbf{x}\|_c \leq 1\} \\
&= \{\mathbf{x} \in \mathbb{R}^p : Z_p(\mathbf{x}) \geq 1\} \\
&= \{\mathbf{x} = r\omega \in \mathbb{R}^p : Z_p(\omega) \geq r^p\} \\
&= \{\mathbf{x} = r\omega \in \mathbb{R}^p : \exp(-\frac{\|\mathbf{y}\|_2^2}{2} + \frac{\omega_{lasso}^2}{2}) z_p(\omega) \geq r^p\} \\
&= \{\mathbf{x} = r\omega \in \mathbb{R}^p : (p-1)! \exp(-\frac{\|\mathbf{y}\|_2^2}{2}) \|\mathbf{A}\omega\|_2^{-p} \exp(\frac{\omega_{lasso}^2}{4}) D_{-p}(\omega_{lasso}) \geq r^p\}.
\end{aligned}$$

Son contour est égal à

$$\begin{aligned}
C(\mathbf{A}, \mathbf{y}) &:= \{\mathbf{x} \in \mathbb{R}^p : \|\mathbf{x}\|_c = 1\} \\
&= \{\mathbf{x} \in \mathbb{R}^p : Z_p(\mathbf{x}) = 1\} \\
&= \{\mathbf{x} = r\omega \in \mathbb{R}^p : Z_p(\omega) = r^p\} \\
&= \{\mathbf{x} = r\omega \in \mathbb{R}^p : \exp(-\frac{\|\mathbf{y}\|_2^2}{2} + \frac{\omega_{lasso}^2}{2}) z_p(\omega) = r^p\} \\
&= \{\mathbf{x} = r\omega \in \mathbb{R}^p : (p-1)! \exp(-\frac{\|\mathbf{y}\|_2^2}{2}) \|\mathbf{A}\omega\|_2^{-p} \exp(\frac{\omega_{lasso}^2}{4}) D_{-p}(\omega_{lasso}) = r^p\}.
\end{aligned}$$

Nous résumons nos résultats dans la proposition suivante.

Proposition 4.4.1. *1) Pour chaque $\omega \in \mathcal{S}_{p-1,1}$, le plus long segment $[0,r]\omega$ contenu dans $\mathcal{B}(\mathbf{A},\mathbf{y})$ a lieu pour $r = r_{max}(\omega)$ solution de l'équation*

$$r^p = (p-1)!\exp(-\frac{\|\mathbf{y}\|_2^2}{2})\|\mathbf{A}\omega\|_2^{-p}\exp(\frac{\omega_{lasso}^2}{4})D_{-p}(\omega_{lasso}).$$

2) La boule

$$\mathcal{B}(\mathbf{A},\mathbf{y}) = \bigcup_{\omega\in\mathcal{S}_{p-1,1}} [0, r_{max}(\omega)]\omega,$$

et son contour est égal à

$$C(\mathbf{A},\mathbf{y}) = \{r_{max}(\omega)\omega : \quad \omega \in \mathcal{S}_{p-1,1}\}.$$

3)Le volume de $\mathcal{B}(\mathbf{A},\mathbf{y})$ est

$$\int_{\mathcal{S}_{p-1,1}} \frac{r_{max}^p(\omega)}{p} d\omega = \frac{Z_p}{p}.$$

4.5 Condition nécessaire et suffisante pour que lasso soit nul

Maintenant nous pouvons donner la condition nécessaire et suffisante pour que lasso soit nul.

Proposition 4.5.1. *Les assertions suivantes sont équivalentes.*
1) $0 = lasso$.
2) $\omega_{lasso}(\omega) \geq 0$ *pour tout* $\omega \in \mathcal{S}_{p-1,1}$.
3) $1 \geq |\langle \mathbf{A}\omega, \mathbf{y}\rangle|$ *pour tout* $\omega \in \mathcal{S}_{p-1,1}$.
4) $|\mathbf{A}_i^\top \mathbf{y}| \leq 1$ *pour tout* $i = 1,\ldots,p$, *où* $\mathbf{A}_i^\top$ *désigne la* i*-ème ligne de* $\mathbf{A}$.

4.6 Concentration autour du lasso

4.6.1 Le cas lasso nul

La formule de changement de variables (4.0.3) nous dit que nous pouvons tirer un vecteur $\mathbf{x}$ selon $c(\mathbf{x})d\mathbf{x}$ en tirant son angle $\mathbf{s}$ uniformément, puis tirer sa distance à l'origine selon $c(r,\mathbf{s})dr$ (4.0.5).

Maintenant nous allons estimer pour $r > 0$ la probabilité

$$\mathbf{P}(\|\mathbf{x}\| > r) = \int_{\mathcal{S}}\int_r^{+\infty} c(r,\mathbf{s})dr\frac{d\mathbf{s}}{|\mathcal{S}|},$$

où $|\mathcal{S}|$ désigne la mesure de Lebesgue de $\mathcal{S}$. Nous introduisons pour chaque couple $a \geq 0$, $b \in \mathbb{R}^p$ la fonction

$$g_{a,b,p}(r) := g_{a,b}(r) - (p-1)\ln(r), \quad r > 0. \tag{4.6.1}$$

Dans ce qui suit

$$a = \|\mathbf{As}\|_2, \quad b = \omega_{lasso}.$$

La fonction $r \geq 0 \to g_{a,b}(r)$ est croissante (car $b := \omega_{lasso} \geq 0$). La fonction $r \to g_{a,b,p}(r)$ est convexe et atteint son minimum au point $r(\mathbf{s})$ solution de l'équation

$$\|\mathbf{As}\|_2(r\|\mathbf{As}\|_2 + \omega_{lasso}) = \frac{p-1}{r}.$$

La racine positive est donnée par

$$r(\mathbf{s}) = \frac{-\omega_{lasso} + \sqrt{\omega_{lasso}^2 + 4(p-1)}}{2\|\mathbf{As}\|_2} \tag{4.6.2}$$

D'une part

$$\int_0^{+\infty} \exp\{-g_{a,b,p}(r)\}dr \geq \exp\{-g_{a,b}(r(\mathbf{s}))\} \int_0^{r(\mathbf{s})} r^{p-1}dr = \exp\{-g_{a,b,p}(r(s))\}\frac{r(\mathbf{s})}{p}.$$

D'autre part en utilisant la convexité de $r \to g_{a,b}(r)$, nous avons pour tout $r > 0$,

$$g_{a,b}(r) \geq g_{a,b}(r(\mathbf{s})) + \frac{(p-1)(r - r(\mathbf{s}))}{r(\mathbf{s})},$$

car $\partial_r g_{a,b}(r(\mathbf{s})) = \frac{p-1}{r(\mathbf{s})}$. Nous en déduisons pour $q > 0$,

$$\begin{aligned}
\int_{qr(\mathbf{s})}^{+\infty} \exp\{-g_{a,b,p}(r)\}dr &\leq \exp\{-g_{a,b}(r) + p - 1\} \int_{qr(\mathbf{s})}^{+\infty} \exp\{-\frac{p-1}{r(\mathbf{s})}r\}r^{p-1}dr \\
&\leq \exp\{-g_{a,b}(r(\mathbf{s})) + p - 1\} \int_{q(p-1)}^{+\infty} \exp(-r)r^{p-1}dr \frac{r(\mathbf{s})}{(p-1)^p} \\
&\leq \exp\{-g_{a,b}(r(\mathbf{s})) + p - 1\}\frac{r(\mathbf{s})}{(p-1)^p}\Gamma(p, q(p-1)),
\end{aligned}$$

où $\Gamma(\nu, x) = \int_x^{+\infty} \exp(-t)t^{\nu-1}dt$ connue sous le nom de la fonction gamma incomplète supérieure. Finalement nous obtenons le résultat suivant.

Proposition 4.6.1. *Nous avons pour tout $q > 0$,*

$$\mathbf{P}(\|\mathbf{x}\| \geq qr(\mathbf{s})) \leq \frac{p\exp(p-1)}{(p-1)^p}\Gamma(p, q(p-1)) := P(q,p). \qquad (4.6.3)$$

En utilisant l'estimation suivante [14]

$$x^{a-1}\exp(-x) < \Gamma(a,x) < Bx^{a-1}\exp(-x), \quad \forall\, a > 1,\ B > 1,\ x > \frac{B}{B-1}(a-1),$$

nous obtenons pour $q > 1$,

$$\Gamma(p, q(p-1)) \leq 2q^{p-1}(p-1)^{p-1}\exp(-q(p-1)).$$

Par conséquent la quantité

$$P(q,p) \leq \frac{2pq^{p-1}}{(p-1)}\exp\{-(q-1)(p-1)\}.$$

Bilan. Si $\mathbf{x}$ est tiré selon la densité c, alors $\frac{\mathbf{x}}{r(\theta)} \in \mathcal{B}_2(0,q)$ avec une probabilité au moins égale à $1 - P(q,p)$.

Dans la figure ci-dessous nous tracons pour $p = 2$, $n = 1$, $\mathbf{A} = \begin{pmatrix} 1 & 1 \end{pmatrix}$ et $\mathbf{y} = 0$ la densité $c(r, \mathbf{s})dr$ pour une valeur de $\mathbf{s}$ fixée.

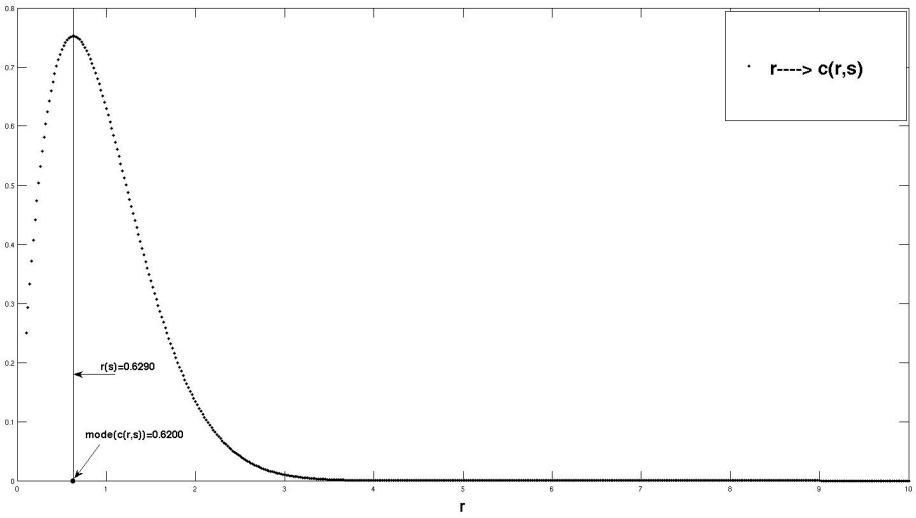

FIGURE 4.1 – pour $r \in [0.1; 10] \to c(r, \mathbf{s})$.

Nous remarquons que le mode de $c(r, \mathbf{s}) = 0.6200$ est très proche de la valeur $r(\mathbf{s})$ (4.6.2) qui vaut 0.6290 pour le même $\mathbf{s}$ fixé.

4.6.2 Le cas général

Nous prenons un vecteur $\mathbf{l} \in lasso$. Nous allons étudier la concentration de c autour de $\mathbf{l}$. La variable d'intérêt est $\mathbf{u} = \mathbf{x} - \mathbf{l}$. La loi de $\mathbf{u}$ a pour densité

$$c(\mathbf{u}+\mathbf{l})d\mathbf{u} = \frac{1}{Z_p}\exp\{-f(\mathbf{u}+\mathbf{l},\mathbf{A},\mathbf{y})\}d\mathbf{u}.$$

La formule de changement de variables donne pour chaque norme $\|\cdot\|$

$$c(\mathbf{u}+\mathbf{l})d\mathbf{u} = \frac{1}{Z_p}\exp\{-f(r\theta+\mathbf{l})\}r^{p-1}drd\theta, \quad r>0,\ \theta\in\mathcal{S}.$$

Par définition pour tout vecteur $\mathbf{x}$, la fonction convexe $r \geq 0 \to f(r\mathbf{s}+\mathbf{l})$ atteint son minimum au point $r=0$. Par conséquent $r \geq 0 \to f(r\mathbf{s}+\mathbf{l})$ est croissante. La fonction

$$f(r\mathbf{s}+\mathbf{l},p) := f(r\mathbf{s}+\mathbf{l}) - (p-1)\ln(r), \quad r>0, \tag{4.6.4}$$

est strictement convexe. Son point critique $r_{\mathbf{l}}(\mathbf{s})$ est solution de l'équation

$$\partial_r f(r\mathbf{s}+\mathbf{l}) = \frac{p-1}{r}.$$

Par une preuve similaire à la proposition (4.6.1) nous avons le résultat suivant ;

Proposition 4.6.2. *Si* $\mathbf{x}$ *est tiré selon la densité* c*, et* $\mathbf{s} = \frac{\mathbf{x}-\mathbf{l}}{\|\mathbf{x}-\mathbf{l}\|}$*, alors pour tout* $q>0$*,*

$$\mathbf{P}(\|\mathbf{x}-\mathbf{l}\| \geq q r_{\mathbf{l}}(\mathbf{s})) \leq \frac{p\exp(p-1)}{(p-1)^p}\Gamma(p,q(p-1)) := P(q,p). \tag{4.6.5}$$

4.7 Applications

4.7.1 Le controur dans le cas $p=2$, $n=1$

Soit $\mathbf{A} := (a_1,a_2)$ une matrice d'ordre 1×2. Son noyau $Ker(\mathbf{A}) = \{(x_1,x_2) : \quad a_1x_1+a_2x_2=0\}$. Nous savons que $\mathcal{B}(a_1,a_2,y)$ contient

$$Ker(\mathbf{A})\bigcap \mathcal{B}_{2,1}.$$

Cette intersection est un segment symétrique noté $[(x_1(a_1,a_2),x_2(a_1,a_2)),-(x_1(a_1,a_2),x_2(a_1,a_2))]$.

Pour déterminer les autres points de l'ensemble $\mathcal{B}(a_1, a_2, y)$, nous allons calculer directement $Z_2(\omega)$. Un simple calcul donne

$$Z_2(\omega) = \exp(-\frac{y^2}{2} + \frac{\omega_{lasso}^2}{2}) \int_0^{+\infty} \exp\{-\frac{(|\mathbf{A}\omega|r + \omega_{lasso})^2}{2}\} r dr,$$

et

$$|\mathbf{A}\omega| \int_0^{+\infty} \exp\{-\frac{(|\mathbf{A}\omega|r + \omega_{lasso})^2}{2}\} r dr + \omega_{lasso} \int_0^{+\infty} \exp\{-\frac{(|\mathbf{A}\omega|r + \omega_{lasso})^2}{2}\} dr = 1.$$

Finalement nous avons la proposition suivante.

Proposition 4.7.1. *1) Si* $\mathbf{A}\omega \neq 0$*, alors*

$$Z_2(\omega) = \exp(-\frac{y^2}{2} + \frac{\omega_{lasso}^2}{2}) |\mathbf{A}\omega|^{-1} \{1 - \frac{\omega_{lasso}}{|\mathbf{A}\omega|} \sqrt{2\pi}(1 - F(\omega_{lasso}))\},$$

où F *est la fonction de répartition de la loi normale.*

2) Si $\mathbf{A}\omega \neq 0$ *et* $y = 0$*, alors*

$$Z_2(\omega) = \omega_{lasso} \exp(\frac{\omega_{lasso}^2}{2}) \{1 - \omega_{lasso}^2 \sqrt{2\pi}(1 - F(\omega_{lasso}))\}.$$

3) Si $\omega \in \mathcal{S}_{1,1}$*,* $\mathbf{A}\omega \neq 0$ *et* $y = 0$*, alors la fonction* z_2

$$z_2(b^2) = \frac{1}{b} \exp(\frac{1}{2b^2}) \{1 - \frac{1}{b^2} \sqrt{2\pi}(1 - F(\frac{1}{b}))\}$$

définie sur $(0, \lambda_{1,2}^2]$ *est convexe et décroissante, où* $\lambda_{1,2} = \max_{\omega \in \mathcal{S}_{1,1}} |\mathbf{A}\omega|$.

4) Nous avons pour $\omega \in \mathcal{S}_{1,1}$

$$Z_2(\omega) = z_2(\frac{1}{\omega_{lasso}^2}), \quad \forall\, \omega \in \mathcal{S}_{1,1}.$$

La boule

$$\mathcal{B}(\mathbf{A}, 0) = \{r\omega : \quad Z_2(\omega) \geq r^2\}.$$

Il est contenu dans le disque unité $\|\mathbf{x}\|_1 \leq 1$ pour la norme l^1. Le contour est défini par l'équation

$$Z_2(\omega) = r^2.$$

La norme de l'opérateur linéaire $\mathbf{A} : (\mathbb{R}^2, \|\cdot\|_1) \to (\mathbb{R}, \|\cdot\|_2)$ est définie par

$$\lambda_{1,2} = \sup_{\omega : \|\omega\|_1 = 1} \|\mathbf{A}\omega\|_2.$$

La fonction $\omega \to Z_2(\omega) = z_2(\lambda_{1,2}^2)$ est constante sur

$$\Omega_{1,2} = \{\omega : \quad \|\omega\|_1 = 1, \quad \|\mathbf{A}\omega\| = \lambda_{1,2}\}.$$

Si $\mathbf{A} = (1,1)$ alors

$$\begin{aligned}\Omega_{1,2} &= \{\omega : \quad \|\omega\|_1 = 1, \quad \|\mathbf{A}\omega\| = \lambda_{1,2}\} \\ &= [(1,0),(0,1)] \bigcup [(-1,0),(0,-1)].\end{aligned}$$

Si $\mathbf{A} = (a_1, a_2)$ avec $|a_1| < |a_2|$, alors

$$\begin{aligned}\Omega_{1,2} &= \{\omega : \quad \|\omega\|_1 = 1, \quad \|\mathbf{A}\omega\|_2 = \lambda_{1,2}\} \\ &= \{(0, sgn(a_2)), (0, -sgn(a_2))\}.\end{aligned}$$

Dans les deux cas

$$\{z_2(\lambda_{1,2}^2)\}^{\frac{1}{2}}\Omega_{1,2}$$

est une partie du contour. Les autres points du contour se déduisent de l'équation

$$z_2(b^2) = a^2, \quad b \in (0, \lambda_{1,2}).$$

Chaque couple (a,b) donne naissance à quatre points de $\mathcal{B}((a_1,a_2),0)$ de la forme $a\omega$ où

$$|\omega_1| + |\omega_2| = 1, \quad |a_1\omega_1 + a_2\omega_2| = b.$$

Nous avons tracé le contour de $\mathcal{B}(a_1, a_2, 0)$ pour différents choix de la matrice a_1, a_2. Nous remarquons que la surface de $\mathcal{B}(a_1, a_2, 0)$ est décroissante en fonction de la norme $\lambda_{1,2}$ de la matrice $\mathbf{A}$, voir figure 4.1.

Remarque 4.7.2. Les calculs numériques montrent que $Z(\omega_{lasso})$ explose pour les grandes valeurs de ω_{lasso}, c'est-à-dire lorsque ω est proche du noyau de $\mathbf{A}$. Pour éliminer les explosions il faut utiliser l'estimation de la queue de la densité gaussienne. En utilisant l'estimation de Gordon [7]

$$\frac{\exp(-\frac{x^2}{2})}{x + \frac{1}{x}} \leq \sqrt{2\pi}(1 - F(x)) \leq \frac{\exp(-\frac{x^2}{2})}{x}, \quad x > 0,$$

nous obtenons l'approximation suivante

$$\frac{1}{b^2} - \frac{1}{b} \leq \psi(b^2) \leq \frac{1}{b^2} - \frac{1}{1+b^3} \tag{4.7.1}$$

au voisinage de 0.

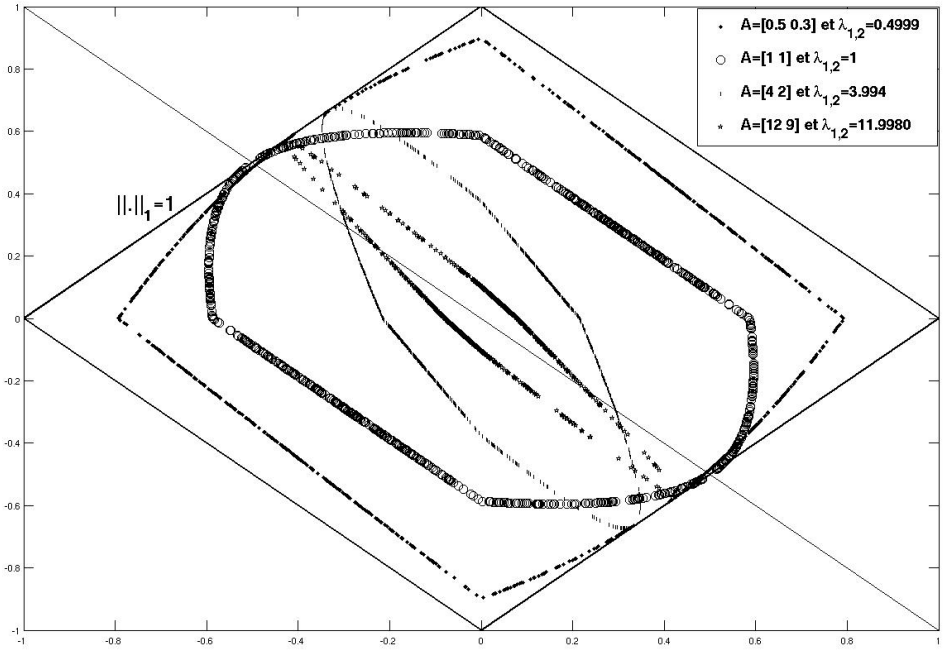

FIGURE 4.2 – Contours de $\mathcal{B}(a_1, a_2, 0)$ pour différentes matrices A, $p = 2$ et $n = 1$.

4.7.2 Application au diagnostic de convergence de l'algorithme de Metropolis-Hastings

Nous considérons le cas $\mathbf{y} = 0$. Nous donnons une illustration de l'inégalité de concentration en comparant les algorithmes de Metropolis à marche aléatoire et de Hastings indépendant en utilisant notre inégalité de concentration. Nous rappelons que si $\mathbf{x}$ est généré selon la loi cible c, alors $\|\mathbf{x}\|_2 \leq qr(\theta)$ avec une probabilité au moins égale à $P(q, p)$ (4.6.5). Le tableau 4.1 donne les valeurs de probabilités $P(q, 7)$ pour différentes valeurs de q. Notons que pour $q \geq 2.5$ le critère $\|\mathbf{x}\|_2 \leq qr(\theta)$ est satisfait avec une probabilité au moins égale à 0.9446.

Étant donnée une chaîne $(\mathbf{x}^{(t)})$ de l'algorithme de Metropolis-Hastings, si $\|\mathbf{x}^{(t)}\|_2 \leq qr(\theta^{(t)})$ alors le vecteur $\mathbf{x}^{(t)}$ peut être considéré comme une réalisation de la loi cible c. Ici

$$\theta^{(t)} := \frac{\mathbf{x}^{(t)}}{\|\mathbf{x}^{(t)}\|_2}.$$

Nous prenons $p = 7$, $n = 4$, $\mathbf{A} \sim \mathcal{B}(\pm\frac{1}{\sqrt{n}})$ et nous prenons le cas où $\mathbf{y} = 0$. Nous simulons selon la loi c (4.0.1) en utilisant l'algorithme de Metropolis-Hastings $(\mathbf{x}^{(t)})$ et nous proposons le test

$$\|\mathbf{x}^{(t)}\|_2 \leq qr(\theta^{(t)})$$

comme critère de convergence.

q	2	2.5	3	3.5	4	4.5	5
$P(q,7)$	0.6672	0.9446	0.9924	0.9991	0.9999	1.0000	1.0000

TABLE 4.2 – Les valeurs de probabilités de $P(q, 7)$.

Les figures 4.2 (a) et (b) représentent respectivement la courbe de $t \to 5r(\theta^{(t)})$ et $t \to \|\mathbf{x}^{(t)}\|_2$ pour l'algorithme de Hastings indépendant et de Metropolis à marche aléatoire. D'après les figures 4.2(a) et 4.2(b), il semble que l'algorithme Metropolis à marche aléatoire converge plus vite que l'algorithme de Hastings indépendant. Plus précisément :

1) D'après les figures 4.2 (a) et (b), l'algorithme de Metropolis à marche aléatoire commence à satisfaire le critère $\|\mathbf{x}^{(t)}\|_2 \leq 3.5r(\theta^{(t)})$ à partir de l'itération $t = 866372$, mais l'algorithme de Hastings indépendant ne vérifie pas le critère de convergence qu'à partir de l'itération $t = 992412$.
2) D'après les figures 4.3 (a) et (b), l'algorithme de Hastings indépendant commence à satisfaire le critère $\|\mathbf{x}^{(t)}\|_2 \leq 5r(\theta^{(t)})$ à partir de l'itération $t = 932996$. Mais L'algorithme de Metropolis à marche aléatoire satisfait le critère $\|\mathbf{x}^{(t)}\| \leq 5r(\theta^{(t)})$ pour $t = 1, \ldots, 10^6$.

Nous proposons un autre critère de comparaison basé sur le fait que

$$\int_{\boldsymbol{R}^p} \mathbf{x}c(\mathbf{x})d\mathbf{x} = 0, \ (car\ ici\ \mathbf{y} = 0). \tag{4.7.2}$$

Le meilleur algorithme parmi $a = HI$ (Hastings indépendant) et $a = MA$ (Metropolis à marche aléatoire) est celui qui donne la meilleure approximation de l'intégrale

$$\int_{\boldsymbol{R}^p} \mathbf{x}c(\mathbf{x})d\mathbf{x} = \frac{1}{N}\sum_{t=1}^{N} \mathbf{x}_a^{(t)} := \hat{\mathbf{x}}_a.$$

Les résultats numériques sont donnés dans le tableau 4.2.

	x_1	x_2	x_3	x_4	x_5	x_6	x_7
$\hat{\mathbf{x}}_{HI}$	-0.0005	-0.0037	0.0016	0.0164	0.0050	0.0021	-0.0058
$\hat{\mathbf{x}}_{MA}$	0.0005	-0.0019	-0.0002	0.0012	-0.0005	0.0031	-0.0011

TABLE 4.3 – les estimateurs $\hat{\mathbf{x}}_{HI}$ et $\hat{\mathbf{x}}_{MA}$ pour $N = 10^6$ itérations.

Il en résulte que

$$\|\frac{1}{N}\sum_{t=1}^{N}\mathbf{x}_{HI}^{(t)}\| = 0.0187$$

et

$$\|\frac{1}{N}\sum_{t=1}^{N}\mathbf{x}_{MA}^{(t)}\| = 0.0041.$$

De nouveau l'algorithme de Metropolis à marche aléatoire gagne contre l'algorithme de Hastings indépendant.

4.8 Appendice : Centrage autour du lasso

Soit $\mathbf{l}$ un élément de lasso et $\mathbf{x}$ un vecteur tiré selon la densité c du lasso bayésien. Nous nous proposons de calculer la loi de $\mathbf{x}-\mathbf{l}$ sachant l'angle $\mathbf{s} := \frac{\mathbf{x}-\mathbf{l}}{\|\mathbf{x}-\mathbf{l}\|}$. Il s'agit de calculer la fonction de répartition de la densité $c(r,\mathbf{s},\mathbf{l})$ proportionnelle à

$$c_{\mathbf{l}}(r,\mathbf{s})dr = \frac{1}{Z_{\mathbf{l}}(\mathbf{s})}\exp\{-f(r\mathbf{s}+\mathbf{l})\}\}r^{p-1}, \tag{4.8.1}$$

où la fonction de partition

$$Z_{\mathbf{l}}(\mathbf{s}) = \int_0^{+\infty}\exp\{-f(r\mathbf{s}+\mathbf{l})\}\}r^{p-1}dr.$$

Nous introduisons les notations suivantes

$$\begin{aligned} S_0 &= \{i=1,\dots,p: \quad s_i = 0\},\\ S_+ &= \{i=1,\dots,p: \quad s_i \neq 0,\ s_i l_i \geq 0\},\\ S_- &= \{i=1,\dots,p: \quad s_i \neq 0,\ s_i l_i < 0\}, \end{aligned}$$

où $l_1, \ldots, l_p$ désignent les composantes du vecteur $\mathbf{l}$. Nous notons par $|S|$ le cardinal de l'ensemble S. Nous avons aussi besoin d'ordonner la suite $(\frac{|l_i|}{|s_i|} : \quad i \in S_-)$:

$$ls(0) := 0 \leq ls(1) := \frac{|l|}{|s|}(1) \leq \ldots \leq ls(|S_-|) \leq ls(|S_-|+1) := +\infty.$$

En utilisant ces notations, nous obtenons

$$\|r\mathbf{s}+\mathbf{l}\|_1 = \sum_{i\in S_0} |l_i| + \sum_{i\in S_+} |s_i|(r + \frac{l_i}{s_i}) + \sum_{i\in S_-} |s_i||r - \frac{|l_i|}{|s_i|}|.$$

Si $ls(k) \leq r < ls(k+1)$, alors

$$\|r\mathbf{s}+\mathbf{l}\|_1 = \|\mathbf{s}\|_{1,k} r + c_k,$$

où

$$\begin{aligned} a_k &= \sum_{i\in S_0} |l_i| + \sum_{i\in S_+} |l_i| - \sum_{i=0}^{k} |l(i)| + \sum_{i=k+1}^{|S_-|} |l(i)|, \\ \|\mathbf{s}\|_{1,k} &= \sum_{i\in S_+} |s_i| + \sum_{i=0}^{k} |s(i)| - \sum_{i=k+1}^{k} |s(i)|. \end{aligned}$$

Ici (i) désigne un indice $k \in S_-$ qui vérifie $\frac{l}{s}(i) = \frac{l_k}{s_k}$. Observer que $\|\mathbf{s}\|_{1,|S_-|} = \|\mathbf{s}\|_1$. Si $\mathbf{As} = 0$, alors la fonction

$$f(r\mathbf{s}+\mathbf{l}) = \|\mathbf{s}\|_{1,k} r + \frac{\|\mathbf{y_l}\|_2^2}{2}, \quad ls(k) \leq r < ls(k+1),$$

où

$$\mathbf{y_l} = \mathbf{y} - \mathbf{Al}.$$

Pour traiter le cas $\mathbf{As} \neq 0$, nous avons besoin des notations suivantes :

$$\begin{aligned} \omega_{lasso}(k) &= \frac{\|\mathbf{s}\|_{1,k}}{\|\mathbf{As}\|_2} - \|\mathbf{y_l}\|_2 \cos(\mathbf{As}, \mathbf{y_l}), \\ \alpha_k &= \frac{\|\mathbf{y_l}\|_2^2}{2} - \frac{\omega_{lasso}^2(k)}{2} + a_k. \end{aligned}$$

Nous obtenons

$$f(r\mathbf{s}+\mathbf{l}) = \alpha_k + \frac{(r\|\mathbf{As}\|_2 + \omega_{lasso}(k))^2}{2}, \quad ls(k) \leq r < ls(k+1).$$

La densité

$$c_{\mathbf{l}}(r,\mathbf{s}) = \frac{1}{Z_{\mathbf{l}}(\mathbf{s})} \sum_{k=0}^{|S_-|+1} \exp(-\alpha_k) \exp\{-\frac{(r\|\mathbf{As}\|_2 + \omega_{lasso}(k))^2}{2}\} \mathbf{1}_{[ls(k)\leq r<ls(k+1)]}.$$

Si nous introduisons

$$Z_{\mathbf{l}}(r,\mathbf{s}) = \int_0^r \exp\{-f(r\mathbf{s}+\mathbf{l})\} r^{p-1} dr,$$

alors pour $ls(k) \leq r < ls(k+1)$

$$Z_{\mathbf{l}}(r,\mathbf{s}) = \sum_{j=0}^{k} \exp(-\alpha_j)\{z(\min(ls(j+1),r), \|\mathbf{As}\|_2, \omega_{lasso}(j)) - z(ls(j), \|\mathbf{As}\|_2, \omega_{lasso}(j))\}.$$

La fonction de répartition de $c_{\mathbf{l}}(r,\mathbf{s})dr$ est égale à

$$r \geq 0 \to \frac{Z_{\mathbf{l}}(r,\mathbf{s})}{Z_{\mathbf{l}}(+\infty,\mathbf{s})}.$$

Exemple. $n = 1, p = 2$, $\mathbf{A} = (1,1)$, $y = 2$. Nous pouvons montrer que $lasso = [(1,0),(0,1)]$. Prenons $\mathbf{l} = (1,0)$ et la norme $\|\cdot\|_1$. Dans ce cas

$$\begin{aligned}
\|r\omega + \mathbf{l}\|_1 &= |r\omega_1 + 1| + r|\omega_2| \\
&= r + 1, \quad \text{si} \quad \omega_1 \geq 0, \\
&= r(\omega_1 + |\omega_2|) + 1, \quad \text{si} \quad \omega_1 < 0, \quad r \leq \frac{1}{|\omega_1|}, \\
&= r - 1, \quad \text{si} \quad \omega_1 < 0, \quad r > \frac{1}{|\omega_1|}.
\end{aligned}$$

Si $\omega_1 \geq 0$, $\omega_1 + \omega_2 \neq 0$ alors

$$Z_{\mathbf{l}}(\omega) = \exp(-1)\exp(-\frac{1}{2} + \frac{\omega_{lasso}^2}{2}) z(\|\mathbf{A}\omega\|_2, \omega_{lasso}),$$

où $\omega_{lasso} = \frac{1}{\|\mathbf{A}\omega\|_2} - 1$. Si $\omega_1 < 0$, alors

$$Z_{\mathbf{l}}(\omega) = \exp(-1)\exp(-\frac{1}{2} + \frac{\omega_{lasso}^2(-)}{2}) z(\frac{1}{|\omega_1|}, \|\mathbf{A}\omega\|_2, \omega_{lasso}(-)) +$$

$$\exp(1)\exp(-\frac{1}{2} + \frac{\omega_{lasso}^2}{2})\{z(\|\mathbf{A}\omega\|_2, \omega_{lasso}) - z(\frac{1}{|\omega_1|}, \|\mathbf{A}\omega\|_2, \omega_{lasso})\}$$

où

$$\omega_{lasso}(-) = \frac{\omega_1 + |\omega_2|}{\|\mathbf{A}\omega\|_2} - 1.$$

4.9 Appendice : Algorithme de Metropolis-Hastings

Il permet de simuler selon notre loi cible c, de manière approchée, en interprétant c comme la loi stationnaire d'une chaîne de Markov (X_n) représentant les étapes successives de l'algorithme. En attendant suffisamment longtemps, X_n a une distribution proche de c. Voir [6], [8], [10], [12].

Nous nous donnons un noyau de transition (appelée aussi loi de proposition) $q(\mathbf{x}_1, \mathbf{x}_2) > 0$, c'est-à-dire $\mathbf{x}_2 \to q(\mathbf{x}_1, \mathbf{x}_2)$ une densité de probabilité pour chaque $\mathbf{x}_1$ fixé.

L'algorithme de Hastings se décrit ainsi :

k, N : **int** ;
$\mathbf{x}_0, \ldots, \mathbf{x}_N$: **configuration** ;
$\mathbf{x}_0 \leftarrow$ configuration initiale
For k **from** 1 **to** N **do**
  $\mathbf{x}^* \leftarrow$ simulation suivant la transition $q(n_{k-1}, .)$ en partant de $\mathbf{x}_{n-1}$;
  If $\left(rand \leq \frac{c(\mathbf{x}^*)q(\mathbf{x}^*, \mathbf{x}_{n-1})}{c(\mathbf{x}_{n-1})q(\mathbf{x}_{n-1}, \mathbf{x}^*)} \right)$ **then**
    $\mathbf{x}_n \leftarrow \mathbf{x}^*$; [acceptation de la proposition de $\mathbf{x}^*$]
  else
    $\mathbf{x}_n \leftarrow \mathbf{x}_{n-1}$; [rejet]
  end If
end for
return $\mathbf{x}_1, \ldots, \mathbf{x}_N$; [la mesure empirique de $(\mathbf{x}_1, \ldots, \mathbf{x}_N)$ est proche de c].

L'algorithme de Metropolis est un cas particulier de l'algorithme de Hastings. Il correspond aux lois de proposition symétrique $q(x_2, x_1) = q(x_1, x_2)$. L'algorithme de Metropolis à marche aléatoire correspond au choix $q(x_1, x_2) = q(|x_2 - x_1|)$.

La chaîne $(\mathbf{x}^{(t)})$ a pour probabilité de transition [5]

$$p(\mathbf{x}_1, \mathbf{x}_2) = q(\mathbf{x}_1, \mathbf{x}_2)\alpha(\mathbf{x}_1, \mathbf{x}_2), \quad \mathbf{x}_1 \neq \mathbf{x}_2,$$
$$p(\mathbf{x}, \mathbf{x}) = \int_{\mathbb{R}^p} q(\mathbf{x}, \mathbf{x}_2)[1 - \alpha(\mathbf{x}, \mathbf{x}_2)]d\mathbf{x},$$

où $\alpha(\mathbf{x}_1, \mathbf{x}_2) := \dfrac{c(\mathbf{x}_2)q(\mathbf{x}_2, \mathbf{x}_1)}{c(\mathbf{x}_1)q(\mathbf{x}_1, \mathbf{x}_2)}$.

La densité c est invariante pour cette chaîne, c'est-à-dire

$$\int_A c(\mathbf{x})d\mathbf{x} = \int_{\mathbb{R}^p} c(\mathbf{x})p(\mathbf{x}, A)d\mathbf{x}$$

pour tout borélien A.

Nous rappelons certains résultats ergodiques. Voir par exemple [11], [13], [15], [16]. Si la densité cible c et la loi de proposition q sont continues et strictement positives, alors nous avons le théorème ergodique

$$\frac{1}{N}\sum_{t=1}^{N} f(\mathbf{x}^{(t)}) \overset{N\to+\infty}{\to} \int_{\mathbb{R}^p} f(\mathbf{x})c(\mathbf{x})d\mathbf{x} \tag{4.9.1}$$

vérifié pour toute fonction f, $cd\mathbf{x}$ intégrable. Ce résultat est beaucoup plus fort que la loi forte des grands nombres classique. Contrairement à la loi forte des grands nombres la convergence a lieu pour chaque trajectoire de la chaîne.

La convergence de $\mathbf{x}^{(t)}$ vers la densité cible c s'énonce comme suit [17]. Sauf sur un ensemble $\mathcal{N}$ tel que $\int_{\mathcal{N}} c(\mathbf{x})d\mathbf{x} = 0$,

$$\|P^t(\mathbf{x}^{(0)},\cdot) - cd\mathbf{x}\|_{VT} := \frac{1}{2}\sup_{A\subset\mathcal{B}(\mathbb{R}^p)} |\mathbb{P}(\mathbf{x}^{(t)}\in A)\,|\,\mathbf{x}^{(0)}) - \int_A c(\mathbf{x})d\mathbf{x}| \to 0$$

pour tout $\mathbf{x}^{(0)} \notin \mathcal{N}$. Cette convergence est appelée : théorème ergodique faible.

L'ergodicité uniforme est définie par

$$\sup_{\mathbf{x}^{(0)}} \|\mathbb{P}^t(\mathbf{x}^{(0)},\cdot) - cd\mathbf{x}\|_{VT} = \frac{1}{2}\sup_{\mathbf{x}^{(0)}}\sup_{A\subset\mathcal{B}(\mathbb{R}^p)} |\mathbb{P}(\mathbf{x}^{(t)}\in A)\,|\,\mathbf{x}^{(0)}) - \int_A c(\mathbf{x})d\mathbf{x}| \to r(t),$$

avec la vitesse de convergence $r(t) \to 0$ lorsque $t \to +\infty$.

Metropolis-Hastings indépendant

L'algorithme de Hastings indépendant correspond au choix de la loi de proposition

$$q(\mathbf{x}_1,\mathbf{x}_2) = p(\mathbf{x}_2), \quad \forall\, \mathbf{x}_1,\mathbf{x}_2.$$

S'il existe $M \geq 1$ tel que

$$\frac{c(\mathbf{x})}{p(\mathbf{x})} \leq M, \quad \forall\, \mathbf{x},$$

alors la chaîne de Markov $(\mathbf{x}^{(t)})$ construite par l'algorithme de Hastings est uniformément ergodique [11] pour tout $\mathbf{x}^{(0)}$ et pour tout entier $t \geq 1$,

$$\sup_{A\subset\mathcal{B}(\mathbb{R}^p)} |\mathbb{P}(\mathbf{x}^{(t)}\in A\,|\,\mathbf{x}^{(0)}) - \int_A c(\mathbf{x})d\mathbf{x}| \leq (1-\frac{1}{M})^t.$$

Dans le cas de notre densité cible

$$c(\mathbf{x}) = \frac{1}{Z_p} \exp(-\frac{\|A\mathbf{x} - \mathbf{y}\|^2}{2} - \|\mathbf{x}\|_1)$$

et le choix

$$p(\mathbf{x}) = \frac{1}{2^p} \exp(-\|\mathbf{x}\|_1)$$

la constante

$$M = \frac{2^p}{Z_p}.$$

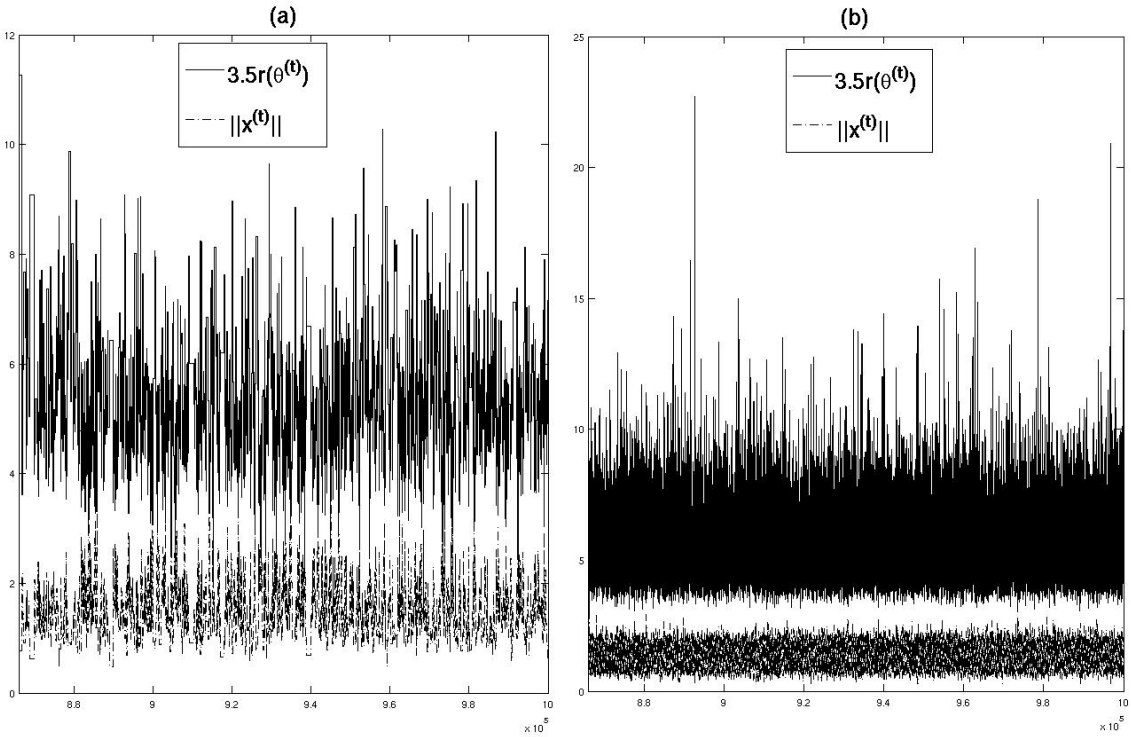

FIGURE 4.3 – (a) : Test de convergence de l'algorithme de Hastings indépendant avec la loi de proposition de Laplace. (b) : Test de convergence de l'algorithme de Metropolis à marche aléatoire avec la loi de proposition $\mathcal{N}(0, 0.5\mathbf{I}_p)$. $N = 10^6$ itérations.

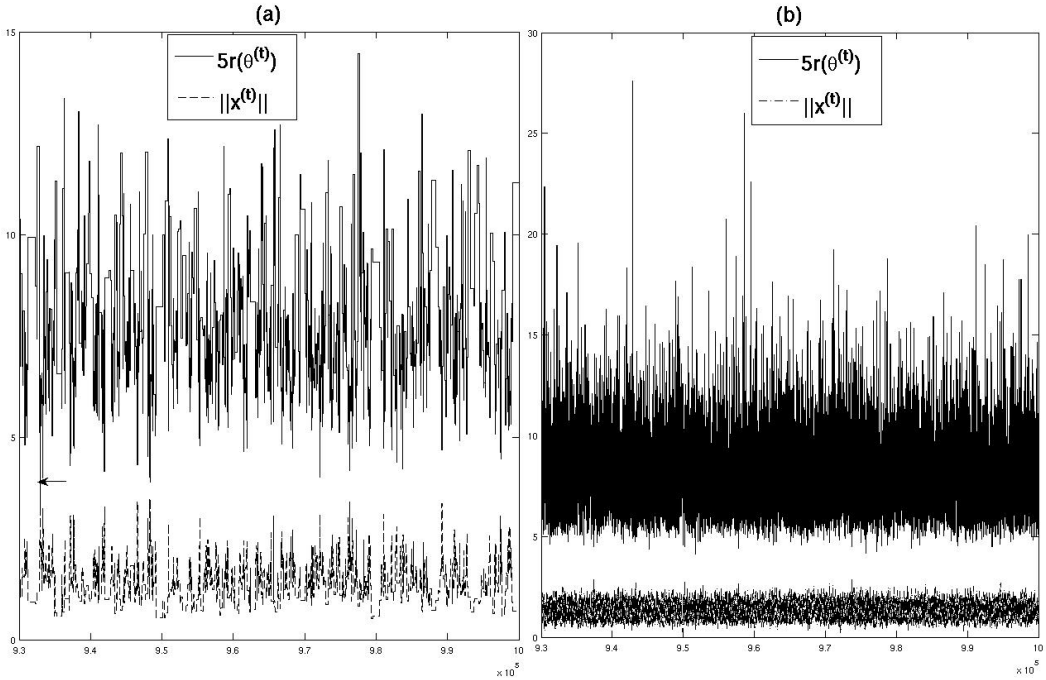

FIGURE 4.4 – (a) : Test de convergence de l'algorithme de Hastings indépendant avec la loi de proposition de Laplace. (b) : Test de convergence de l'algorithme de Metropolis à marche aléatoire avec la loi de proposition $\mathcal{N}(0, 0.5\mathbf{I}_p)$. $N = 10^6$ itérations.

Bibliographie

[1] K. Ball, Logarithmically concave functions and sections of convex sets in R^n, Stud. Math. T. LXXXVIII 69–84 (1988).

[2] E.Cojocaru, Parabolic cylinder functions implemented in Matlab, arXiv :0901.2220 [math-ph] (2009).

[3] A. Dermoune, D. Ounaissi and N. Rahmania, MCMC convergence diagnosis using geometry of Bayesian LASSO, arXiv :1512.01366v1 [math.ST] (2015).

[4] A. Erdélyi, Higher transcendental functions, California institute of technology, bateman manuscript project, vol. 2 p. 119 (1953).

[5] L. Evers, A. M. Johansen : Monte-Carlo Methods, Lecture Notes, University of Bristol, 1–128 (2007).

[6] E. Gobet, Méthodes de Monte-Carlo et processus stochastiques : du linéaire au non-linéaire, Editions de l'Ecole Polytechnique (2013).

[7] R.D. Gordon, Values of Mills' ratio of area to bounding ordinate of the normal probability integral for large values of the argument, Annals of Mathematical Statistics, vol. 12 364–366 (1941)

[8] W.K. Hastings, Multilevel Monte Carlo methods, In LSSC'01 Proceedings of the third international conference on large-scale scientific computing, vol. 2179 of lecture notes in computer science pages 58–67, Springer-Verlag (2001).

[9] B. Klartag and V.D. Milman, Geometry of log-concave functions and measures, Geom. Dedicata vol. 112 no. 1 169–182 (2005).

[10] J.S. Liu, Monte Carlo strategies in scientific computing, Springer Series in Statistics, Springer-Verlag, New-York (2001).

[11] K. Mengersen and R.L Tweedie, Rates of convergence of the Hastings and Metropolis algorithms, Ann. Statist. vol. 24 101–121 (1996).

[12] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller and E. Teller, Equations of state calculations by fast computing machines, journal of Chemical Physics, vol. 21 no. 6 1087–1092 (1953).

[13] S.P. Meyn, R.L.Tweedie, Markov chains and stochastic stability, Springer-Verlag, London. Available at : probability.ca/MT (1993).

[14] P. Natalini, B. Palumbo, Inequalities for the incomplete gamma function, Mathematical inequalities and Applications, vol 3 no. 1 69–77 (2000).

[15] C.P. Robert and G. Casella, Monte Carlo statistical methods, Springer science (2004).

[16] G.O. Roberts, A.F.M. Smith, Simple conditions for the convergence of the Gibbs sampler and Metropolis-Hastings algorithms, Stochastic Processes and their Applications vol. 49 207–216 (1994).

[17] G.0. Roberts and A.L. Tweedie, Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms, Biometrika vol. 83 no. 1 95–110 (1996).

[18] N.M. Temme, Numerical and asymptotic aspects of parabolic cylinder functions, ournal of Computational and Applied Mathematics 121 221–246 (2000).

Chapitre 5

Simulation du lasso bayésien par une équation différentielle stochastique avec dérive discontinue

Nous pouvons voir la densité $c(\mathbf{x})d\mathbf{x}$ comme la densité invariante de l'équation différentielle stochastique

$$d\mathbf{x}(t) = -\frac{1}{2}\partial f(\mathbf{x}(t))dt + d\mathbf{w}(t), \tag{5.0.1}$$

où $f(\mathbf{x}) = -\ln(c(\mathbf{x}))$ (4.0.1) et $\mathbf{w}$ désigne le mouvement brownien sur $\mathbb{R}^p$.

Nous pouvons aussi fixé la direction $\omega = \frac{\mathbf{x}}{\|\mathbf{x}\|_1}$ et voir la densité de $r = \|\mathbf{x}\|_1$ (voir (4.0.5 chapitre 4) comme la densité invariante pour l'EDS sur $(0, +\infty)$ définie par

$$dr(t) = \frac{1}{2}\{\frac{(p-1)}{r(t)} - \|\mathbf{A}\omega\|_2(\|\mathbf{A}\omega\|_2 r(t) + \omega_{lasso})\}dt + dB_t, \tag{5.0.2}$$

où B est un mouvement brownien standard sur $\mathbb{R}$. Si $\omega_{lasso} = 0$, alors $(r^2(t))$ est le processus de Cox-Ingersoll-Ross (CIR) connu aussi comme le processus de Bessel généralisé de paramètres $(p-1, -\|\mathbf{A}\omega\|_2^2)$ [12].

La densité $c(\mathbf{x})d\mathbf{x}$ est le candidat pour être la densité invariante de notre EDS (5.0.1), mais nous n'avons pas pu le montrer. En revanche nous pouvons montrer en utilisant [18] que la densité de r est invariante pour la diffusion (5.0.2).

Dans ce chapitre nous allons approcher la loi cible $c(\mathbf{x})d\mathbf{x}$ (4.0.1) par la loi de $\mathbf{x}(T)$ lorsque T est grand, en utilisant les schémas d'Euler semi-implicite et explicite en utilisant les méthodes de Monte Carlo (MC), Monte Carlo à plusieurs niveaux (MLMC). La même étude est en cours concernant la diffusion (5.0.2).

Nous rappelons d'abord les résultats théoriques concernant les équations différentielles stochastiques multivariées (EDSM).

5.1 Équation différentielle stochastique multivariée (EDSM)

Soit $\mathbf{w}$ un mouvement brownien standard $\mathbb{R}^p$ et $\mathbf{b} : \boldsymbol{R}^p \to \boldsymbol{R}^p$ une application régulière. Une solution à l'EDSM

$$d\mathbf{x}_t = -[\partial\varphi(\mathbf{x}_t)dt + \mathbf{b}(\mathbf{x}_t)dt] + d\mathbf{w}_t \tag{5.1.1}$$

est un couple $t \in [0, +\infty) \to (\mathbf{x}(t), \mathbf{l}(t)) \in \mathbb{R}^p \times \mathbb{R}^p$ de processus adaptés à la filtration brownienne et continues avec $\mathbf{l}(0) = 0$. L'application $t \to \mathbf{l}(t)$ est à variation localement bornée et

$$d\mathbf{x}_t = -d\mathbf{l}_t - \mathbf{b}(\mathbf{x}_t)dt + d\mathbf{w}_t.$$

La dérivée "$\dfrac{d\mathbf{l}(t)}{dt} \in \partial\varphi(\mathbf{x}(t))$" dans le sens suivant :

$$\langle \mathbf{x}_t - \alpha_t, d\mathbf{l}_t - \beta_t dt \rangle$$

est une mesure positive pour tout couple de trajectoires $t \to (\alpha_t, \beta_t)$ continues et telles que $\beta_t \in \partial\varphi(\alpha_t)$. Nous observons que si $d\mathbf{l}_t = \mathbf{l}'_t dt$, alors $\mathbf{l}'_t \in \partial\varphi(\mathbf{x}_t)$.

Il est bien connu que si

$$\begin{aligned} \|\mathbf{b}(\mathbf{x}_1) - \mathbf{b}(\mathbf{x}_2)\|_2 &\leq C\|\mathbf{x}_1 - \mathbf{x}_2\|_2, \quad \forall\, \mathbf{x}_1, \mathbf{x}_2, \\ \|\mathbf{b}(\mathbf{x})\|_2 &\leq C(1 + \|\mathbf{x}\|_2), \quad \forall\, \mathbf{x}, \end{aligned}$$

alors il existe une unique solution $(\mathbf{x}, \mathbf{l})$. See e.g. [2], [3], [5], [6], [14], [21]. Par conséquent l'équation (5.0.1) a une unique solution $(\mathbf{x}, \mathbf{l})$. En général la mesure $d\mathbf{l}_t$ n'est pas absolument continue par rapport à la mesure de Lebesgue dt. Dans notre cas, nous allons montrer que $d\mathbf{l}_t$ est absolument continue.

Nous rappelons deux méthodes de construction de la solution $\mathbf{x}$ of (5.0.1).

1) En choisissant $\varphi(\mathbf{x}) = \|\mathbf{x}\|_1$, $\mathbf{b}(\mathbf{x}) = \mathbf{A}^*(\mathbf{A}\mathbf{x} - \mathbf{y})$, alors la solution de (5.0.1) est l'unique couple $(\mathbf{x}, \mathbf{l})$ de processus adaptés et continus avec $\mathbf{l}(0) = 0$, $t \to \mathbf{l}(t)$ à variation localement bornée et

$$d\mathbf{x}(t) = -[d\mathbf{l}_t + \mathbf{A}^*(\mathbf{A}\mathbf{x}(t) - \mathbf{y})dt] + d\mathbf{w}_t, \quad \frac{d\mathbf{l}(t)}{dt} \in \partial\|\mathbf{x}(t)\|_1. \tag{5.1.2}$$

2) En choisissant $\varphi(\mathbf{x}) = \|\mathbf{x}\|_1 + \dfrac{\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2}{2}$, $\mathbf{b}(\mathbf{x}) = 0$, alors la solution de (5.0.1) est l'unique couple $(\mathbf{x}(t), \mathbf{k}(t))$ continu et adpaté qui vérifie

$$d\mathbf{x}(t) = -d\mathbf{k}(t) + d\mathbf{w}(t), \quad \mathbf{k}(t) \in \partial\varphi(\mathbf{x}(t)).$$

L'unicité de la solution de (5.0.1) implique

$$d\mathbf{k}(t) = d\mathbf{l}_t + \mathbf{A}^*(\mathbf{A}\mathbf{x}(t) - \mathbf{y})dw.$$

Maintenant, nous allons montrer que $\mathbf{l}$ est absolument continu.

5.1.1 Problème de Skorokhod

Soit $f : [0, T] \to \mathbb{R}^d$ une fonction continue et $\psi : \mathbb{R}^p \to \mathbb{R}$ une fonction convexe. Ils existe un unique couple $(\mathbf{x}, \mathbf{k})$ de fonctions continues avec $\mathbf{k}(0) = 0$, $t \to \mathbf{k}(t)$ à variation localement bornée tels que

$$\mathbf{x}(t) = \boldsymbol{f}(t) - \mathbf{k}(t), \quad \forall\, t \geq 0, \tag{5.1.3}$$

$$\langle \mathbf{x}(t) - \alpha(t), d\mathbf{k}(t) - \beta(t)dt \geq 0, \tag{5.1.4}$$

pour tout couple de trajectoires continues $t \to (\alpha(t), \beta(t))$ et telle que $\beta(t) \in \partial\psi(\alpha(t))$. Voir par exemple [6].

Maintenant nous pouvons énoncer le résultat suivant.

Proposition 5.1.1. *Nous supposons que*

$$m = \sup\{\|\mathbf{v}\|_2 : \quad \mathbf{v} \in \bigcup_{\mathbf{x}\in\mathbb{R}^p} \partial\psi(\mathbf{x})\} \tag{5.1.5}$$

est fini. Alors la fonction $\mathbf{l}$ *solution du problème de Skorokhod (5.1.4) est absolument continue.*

Preuve 5.1.2. *Soit* $\mathbf{e} \in \mathbb{R}^p$ *tel que* $\|\mathbf{e}\|_2 = 1$, $\gamma > 0$ *et* $v \in \partial\psi(\gamma\mathbf{e})$ *ayant la plus petite norme euclidienne. Comme* $(\mathbf{x}, \mathbf{k})$ *est la solution du problème de Skorokhod, alors*

$$\begin{aligned}\langle \mathbf{x}(t) - \gamma\mathbf{e}, d\mathbf{l}(t)\rangle &\geq \langle \mathbf{x}(t) - \gamma\mathbf{e}, \mathbf{v}dt\rangle \\ &\geq -m\left(\|\mathbf{x}(t)\|_2 + \gamma\right)dt.\end{aligned}$$

Pour chaque $0 \leq s < t$, *nous avons*

$$\begin{aligned}\langle \mathbf{l}(t) - \mathbf{l}(s), \mathbf{e} \rangle &= \int_s^t \langle \mathbf{e}, d\mathbf{l}(u) \rangle \\ &= \gamma^{-1} \int_s^t \langle \mathbf{x}(u), d\mathbf{l}(u) \rangle - \gamma^{-1} \int_s^t \langle \mathbf{x}(u) - \gamma \mathbf{e}, d\mathbf{l}(u) \rangle \\ &\leq \gamma^{-1} \int_s^t \langle \mathbf{x}(u), d\mathbf{l}(u) \rangle + m\gamma^{-1} \int_{t_i}^{t_{i+1}} \|\mathbf{x}(u)\| du + m(t-s).\end{aligned}$$

En utilisant la dernière inégalité,

$$\|\mathbf{l}(t) - \mathbf{l}(s)\|_2 = \sup\{\langle \mathbf{l}(t) - \mathbf{l}(s), \mathbf{e} \rangle : \quad \mathbf{e} \in \mathbb{R}^p, \quad \|\mathbf{e}\|_2 = 1\},$$

et en faisant $\gamma \to +\infty$, *nous obtenons*

$$\|\mathbf{l}(t) - \mathbf{l}(s)\|_2 \leq m(t-s).$$

Ceci achève la preuve.

En prenant

$$\boldsymbol{f}(t) = \mathbf{x}_0 - \int_0^t \mathbf{A}^*(\mathbf{A}\mathbf{x}(s) - \mathbf{y}) ds + \mathbf{w}_t,$$

nous déduisons que la solution $(\mathbf{x}, \mathbf{l})$ (5.1.2) est la solution au problème de Skorokhod. Par hypothèse l'équation (5.1.5) est satisfaite pour $\psi(\mathbf{x}) = \|\mathbf{x}\|_1$, avec $m = 1$, et alors $\mathbf{l}$ est absolument continu. Finalement la solution (5.0.1) satisfait

$$\mathbf{x}(t) = \mathbf{x}(0) - \int_0^t [\mathbf{v}(s) + \mathbf{A}^*(\mathbf{A}\mathbf{x}(s) - \mathbf{y})] ds + \mathbf{w}(t), \tag{5.1.6}$$

et $\mathbf{v}(t) \in \partial \|\mathbf{x}(t)\|_1$, $\|\mathbf{v}(t)\|_2 \leq 1$, dt p.p.

De plus nous montrons p.s. pour $i = 1, \ldots, p$ que $x_i(t) \neq 0$ et $v_i(t) = sgn(x_i(t))$, dt p.p.

5.2 Simulation du LASSO bayésien en utilisant l'EDSM

Pour approcher $\mathbb{E}[\mathbf{x}(T)]$, nous allons utiliser des schémas numériques pour approximer la solution de (5.0.1) en utilisant les pas de discrétisation du temps

$$\Delta t_l = 2^{-l} T, \tag{5.2.1}$$

avec les niveaux $l = l_s, l_s + 1, \ldots$. Dans la suite le plus petit niveau

$$l_s := \frac{\ln(T)}{\ln(2)} + 1.$$

Ayant un schéma numérique $(\mathbf{x}_L(sc, k) : \quad k = 1, \ldots, 2^L)$ tel que

$$\mathbb{E}[\mathbf{x}_L(sc, 2^L)] \to \mathbb{E}[\mathbf{x}(T)]$$

lorsque $L \to +\infty$. Nous allons calculer $\mathbb{E}[\mathbf{x}_L(sc, 2^L)]$ pour L grand. Pour atteindre ce but, nous proposons la méthode de Monte Carlo (MC) [10] et celle de Monte Carlo multivariée (MLMC) [8]. Nous allons discuter l'efficacité des estimateurs de $\mathbb{E}[\mathbf{x}(T)]$ en utilisant MC et MLMC.

Nous allons nous inspirer des résultats obtenus dans [20] pour les collisions de Coulomb, et nous proposons une nouvelle méthode pour calculer le coût de l'estimation.

5.3 Coût optimal de la méthode MC

Etant donné un échantillon $(\mathbf{x}_l^k(sc, 2^l) : k = 1, \ldots, N_l)$ de taille N_l de la loi de $\mathbf{x}_l(sc, 2^l)$, nous définissons

$$\hat{\mathbf{x}}_l^{N_l}(sc, 2^l) = \frac{1}{N_l} \sum_{k=1}^{N_l} \mathbf{x}_l^k(sc, 2^l), \quad l, \text{ and, } N_l \text{ are fixed,} \tag{5.3.1}$$

$$\hat{\mathbf{x}}_l(sc, 2^l) := \mathbb{E}[\mathbf{x}_l(sc, 2^l)] = \lim_{N_l \to +\infty} \hat{\mathbf{x}}_l^{N_l}(sc, 2^l), \tag{5.3.2}$$

$$\hat{\mathbf{x}}(T) := \mathbb{E}[\mathbf{x}(T)] = \lim_{l \to +\infty} \hat{\mathbf{x}}_l(sc, 2^l). \tag{5.3.3}$$

Nous rappelons que la méthode MC propose d'estimer $\hat{\mathbf{x}}_l(sc, 2^l)$ (5.3.2) par $\hat{\mathbf{x}}_l^{N_l}(sc, 2^l)$ (5.3.1).

Si nous estimons $\hat{\mathbf{x}}(T)$ (5.3.3) par $\hat{\mathbf{x}}_l^{N_l}(sc, 2^l)$, alors l'erreur a deux sources. L'approximation de $\mathbb{E}[\mathbf{x}(T)]$ par $\mathbb{E}[\mathbf{x}_l(sc, 2^l)]$, et l'erreur de l'approximation de $\hat{\mathbf{x}}_l(sc, 2^l) = \mathbb{E}[\mathbf{x}_l(sc, 2^l)]$ par $\hat{\mathbf{x}}_l^{N_l}(sc, 2^l)$ qui dépend de la taille N_l de l'échantillon.

La précision de $\hat{\mathbf{x}}_l^{N_l}(sc, 2^l)$ comme estimateur de $\hat{\mathbf{x}}(T)$ est égale à

$$\begin{aligned} MSE &= \mathbb{E}[\|\hat{\mathbf{x}}(T) - \hat{\mathbf{x}}_l^{N_l}(sc, 2^l)\|_2^2] \\ &= \|\hat{\mathbf{x}}(T) - \hat{\mathbf{x}}_l(sc, 2^l)\|_2^2 + \mathbb{E}[\|\hat{\mathbf{x}}_l(sc, 2^l) - \hat{\mathbf{x}}_l^{N_l}(sc, 2^l)\|_2^2]. \end{aligned}$$

Nous avons

$$\mathbb{E}[\|\hat{\mathbf{x}}_l(sc, 2^l) - \hat{\mathbf{x}}_l^{N_l}(sc, 2^l)\|_2^2] := \frac{Var_l(sc)}{N_l},$$

où

$$Var_l(sc) = \sum_{i=1}^{p} Var(x_{l,i}(sc, 2^l)).$$

Ici $x_{l,i}(sc, 2^l)$ est la i-th composante de $\mathbf{x}_l(sc, 2^l)$ et $Var(x_{l,i}(sc, 2^l))$ désigne sa variance.

La quantité

$$\|\hat{\mathbf{x}}(T) - \hat{\mathbf{x}}_l(sc, 2^l)\|^2 = e(sc, \Delta t_l) \tag{5.3.4}$$

est une fonction du pas de discrétisation Δt_l. Elle joue le rôle principale dans le calcul de l'erreur MSE.

L'estimateur $\hat{\mathbf{x}}_l^{N_l}(sc, 2^l)$ de $\hat{\mathbf{x}}(T)$ a une précision η^2 si

$$\begin{aligned} MSE &= \mathbb{E}[\|\hat{\mathbf{x}}(T) - \hat{\mathbf{x}}_l^{N_l}(sc, 2^l)\|_2^2] = \eta^2 \\ &= e(sc, \Delta t_l) + \frac{Var_l(sc)}{N_l}. \end{aligned} \tag{5.3.5}$$

Le coût de calcul K pour obtenir $(\mathbf{x}_l^k(sc, 2^l) : \quad k = 1, \ldots, N_l)$ est

$$K(N_l, \Delta t_l) = N_l \frac{T}{\Delta t_l} = N_l 2^l.$$

Nous cherchons donc à minimiser le coût K sous la contrainte (5.3.5). En utilisant la méthode des multiplicateurs de Lagrange

$$L(N_l, \Delta t_l, \lambda) = N_l \frac{T}{\Delta t_l} + \lambda(e(sc, \Delta t_l) + \frac{Var_l(sc)}{N_l} - \eta^2),$$

nous obtenons le choix optimal

$$\begin{aligned} \frac{T}{\Delta t_l} - \lambda \frac{Var_l(sc)}{N_l^2} &= 0, \\ -N_l \frac{T}{(\Delta t_l)^2} + \lambda \frac{\partial e}{\Delta t_l}(sc, \Delta t_l) &= 0, \\ e(sc, \Delta t_l) + \frac{Var_l(sc)}{N_l} &= \eta^2. \end{aligned} \tag{5.3.6}$$

Par conséquent

$$\frac{\partial\, e(sc, \Delta t_l)}{\Delta t_l} = \frac{\eta^2 - e(sc, \Delta t_l)}{\Delta t_l}, \tag{5.3.7}$$

$$e(sc, \Delta t_l) < \eta^2. \tag{5.3.8}$$

Nous proposons de résoudre le dernier système numériquement comme suit. Nous estimons $\hat{\mathbf{x}}(T)$ par $\hat{\mathbf{x}}_L(sc, 2^L)$ avec $L = 16$. D'une part

$$\|\hat{\mathbf{x}}_L(sc, 2^L) - \hat{\mathbf{x}}_l(sc, 2^l)\|_2^2 \approx e(sc, \Delta t_l). \tag{5.3.9}$$

D'autre part

$$\frac{\partial\, e(sc, \Delta t_l)}{\Delta t_l} \approx \frac{e(sc, \Delta t_l) - e(sc, \Delta t_{l+1})}{T 2^{-l-1}}.$$

L'équation (5.3.7) devient

$$3e(sc, \Delta t_l) - 2e(sc, \Delta t_{l+1}) \approx \eta^2.$$

Maintenant nous calculons pour $l \geq l_s$ la quantité

$$3e(sc, \Delta t_l) - 2e(sc, \Delta t_{l+1}) \tag{5.3.10}$$

avec le critère d'arrêt

$$e(sc, \Delta t_l) < \eta^2. \tag{5.3.11}$$

Ayant obtenu le l optimal, nous calculons $Var_l(sc)$ par

$$\sum_{i=1}^{p} \frac{1}{N} \sum_{k=1}^{N} |x_{l,i}^k(sc, 2^l) - \frac{1}{N} \sum_{k=1}^{N} x_{l,i}^k(sc, 2^l)|^2. \tag{5.3.12}$$

Ayant obtenu l et $Var_l(sc)$, nous calculons la taille optimale N_l en utilisant l'équation (5.3.6) et finalement nous obtenons le coût optimal K_l.

5.4 Coût optimal de la méthode MLMC

La méthode MLMC a été initialement développée pour les mathématiques financières [8], [9]. Aujourd'hui elle est utilisée dans beaucoup de domaines.

La méthode MLMC utilise plusieurs niveaux et elle s'utilise de la manière suivante. Dans notre cas, nous considérons les niveaux $l = l_s, l_s + 1, \ldots, l_m < L = 16$. Le plus petit niveau l_s est choisi de sorte que $\Delta t_{l_s} = \frac{1}{2}$. Nous générons un échantillon $(\mathbf{x}_{l_s}^k(sc, 2^{l_s}) : k = 1, \ldots, N_{l_s})$ de taille N_{l_s} de la loi de $\mathbf{x}_{l_s}(sc, 2^{l_s})$, et pour chaque $l = l_s + 1, \ldots, l_m$, nous générons à partir de la même trajectoire et de la même condition initiale des échantillons $(\mathbf{x}_l^k(sc, 2^l) : k = 1, \ldots, N_l)$ et $(\mathbf{x}_{l-1}^k(sc, 2^{l-1}) : k = 1, \ldots, N_l)$ de

la loi de $\mathbf{x}_l(sc, 2^l)$ et $\mathbf{x}_{l-1}(sc, 2^{l-1})$ respectivement. De plus les échantillons $(\mathbf{x}_{l_s}^k(sc, 2^{l_s}) : k = 1, \ldots, N_{l_s})$, $(\mathbf{x}_l^k(sc, 2^l), \mathbf{x}_{l-1}^k(sc, 2^{l-1}) : k = 1, \ldots, N_l)$ pour $l = l_s + 1, \ldots, l_m$ doivent être indépendants. En utilisant la somme télescopique

$$\hat{\mathbf{x}}_{l_m}(sc, 2^{l_m}) = \hat{\mathbf{x}}_{l_s}(sc, 2^{l_s}) + \sum_{l=l_s+1}^{l_m} (\hat{\mathbf{x}}_l(sc, 2^l) - \hat{\mathbf{x}}_{l-1}(sc, 2^{l-1})),$$

la méthode MLMC propose l'estimateur

$$\hat{\mathbf{x}}_{l_m}^{N_{l_m}}(2^{l_m}) = \hat{\mathbf{x}}_{l_s}^{N_{l_s}}(sc, 2^{l_s}) + \sum_{l=l_s+1}^{l_m} \left(\hat{\mathbf{x}}_l^{N_l}(sc, 2^l) - \hat{\mathbf{x}}_{l-1}^{N_l}(sc, 2^{l-1})\right)$$

de $\hat{\mathbf{x}}_{l_m}(sc, 2^{l_m}) := \mathbb{E}[\mathbf{x}_{l_m}(sc, 2^{l_m})]$.

Nous introduisons pour chaque niveau l et pour chaque taille N_l d'échantillon les notations suivantes :

$$\hat{\mathbf{x}}_l^{N_l}(sc, 2^l) - \hat{\mathbf{x}}_{l-1}^{N_l}(sc, 2^{l-1}) := \delta\hat{\mathbf{x}}_l^{N_l}(sc, 2^l).$$

Nous en déduisons

$$\begin{aligned}\hat{\mathbf{x}}_{l_m}^{N_{l_m}}(sc, 2^{l_m}) &= \hat{\mathbf{x}}_{l_s}^{N_{l_s}}(sc, 2^{l_s}) + \sum_{l=l_s+1}^{l_m} \left(\hat{\mathbf{x}}_l^{N_l}(sc, 2^l) - \hat{\mathbf{x}}_{l-1}^{N_l}(sc, 2^{l-1})\right) \\ &:= \hat{\mathbf{x}}_{l_s}^{N_{l_s}}(sc, 2^{l_s}) + \sum_{l=l_s+1}^{l_m} \delta\hat{\mathbf{x}}_l^{N_l}(sc, 2^l), \qquad (5.4.1)\end{aligned}$$

où $\delta\hat{\mathbf{x}}_l^{N_l}(sc, 2^l) := \hat{\mathbf{x}}_l^{N_l}(sc, 2^l) - \hat{\mathbf{x}}_{l-1}^{N_l}(sc, 2^{l-1})$.

La précision de l'estimation de $\hat{\mathbf{x}}(T)$ par l'estimateur $\hat{\mathbf{x}}_{l_m}^{N_{l_m}}(sc, 2^{l_m})$ est calculée par

$$MSE := \mathbb{E}[\|\hat{\mathbf{x}}(T) - \hat{\mathbf{x}}_{l_m}^{N_{l_m}}(sc, 2^{l_m})\|_2^2] = \|\hat{\mathbf{x}}(T) - \hat{\mathbf{x}}_{l_m}(sc, 2^{l_m})\|_2^2 + Var(\hat{\mathbf{x}}_{l_m}^{N_{l_m}}(sc, 2^{l_m})).$$

Si nous posons $V_{l_s} = Var(\mathbf{x}_{l_s}(sc, 2^{l_s}))$, et pour $l = l_s + 1, \ldots, l_m$,

$$V_l = Var(\delta\mathbf{x}_l(sc, 2^l)), \qquad (5.4.2)$$

alors

$$Var(\hat{\mathbf{x}}_{l_m}^{N_{l_m}}(sc, 2^{l_m})) = \sum_{l=l_s}^{l_m} \frac{V_l}{N_l},$$

et

$$MSE = \|\hat{\mathbf{x}}(T) - \hat{\mathbf{x}}_{l_m}(sc, 2^{l_m})\|_2^2 + \sum_{l=l_s}^{l_m} \frac{V_l}{N_l}.$$

Si

$$\|\hat{\mathbf{x}}(T) - \hat{\mathbf{x}}_{l_m}(sc, 2^{l_m})\|_2^2 := e(sc, \Delta t_{l_m}) + \sum_{l=l_s}^{l_m} \frac{V_l}{N_l} = \eta^2, \tag{5.4.3}$$

alors le coût de la méthode MLMC est égal à

$$K = \sum_{l=l_s}^{l_m} K_l = \sum_{l=l_s}^{l_m} N_l \frac{T}{\Delta t_l}. \tag{5.4.4}$$

Nous cherchons à minimiser ce coût sous la contrainte (5.4.3).

Nous estimons pour $l_s \leq l < L = 16$,

$$\|\hat{\mathbf{x}}(T) - \hat{\mathbf{x}}_l(sc, 2^l)\|_2^2$$

par

$$\|\hat{\mathbf{x}}_L(sc, 2^L) - \hat{\mathbf{x}}_l(sc, 2^l)\|_2^2.$$

Nous considérons maintenant l'ensemble $l(\eta)$ des niveaux l tels que

$$e(sc, \Delta t_l) = \|\hat{\mathbf{x}}_L(sc, 2^L) - \hat{\mathbf{x}}_l(sc, 2^l)\|_2^2 \approx \frac{\eta^2}{2}.$$

Pour chaque $lopt \in l(\eta)$, l'équation (5.4.3) devient

$$\sum_{l=l_s}^{lopt} \frac{V_l}{N_l} = \eta^2 - e(sc, \Delta t_{lopt}). \tag{5.4.5}$$

Ayant trouvé $lopt$, nous minimisons K (5.4.4) sous la contrainte (5.4.5) en utilisant les multiplicateurs de Lagrange :

$$\partial_{N_l}(K + \lambda(\frac{V_{l_s}}{N_{l_s}} + \sum_{l=l_s+1}^{lopt} \frac{V_l}{N_l} - (\eta^2 - e(sc, \Delta t_{lopt}))) = 0, \quad l = l_s, \ldots, lopt.$$

Ainsi

$$2^l = \lambda \frac{V_l}{N_l^2}, \quad l = l_s, \ldots, lopt,$$

$$\sum_{l=l_s+1}^{lopt} \frac{V_l}{N_l} = \eta^2 - e(sc, \Delta t_{lopt}).$$

Il en découle $l = l_s, \dots, lopt$ que $N_l = \sqrt{\lambda V_l 2^{-l}}$, et alors

$$\sum_{l=l_s}^{lopt} \frac{\sqrt{V_l 2^l}}{\sqrt{\lambda}} = \eta^2 - e(sc, \Delta t_{lopt}).$$

Ayant $lopt$, nous estimons V_{l_s}, and $(V_l : l = l_s + 1, \dots, lopt)$ par

$$\hat{V}_{l_s} = \sum_{i=1}^{p} \frac{1}{N} \sum_{k=1}^{N} |x^k_{l_s,i} - \hat{x}^k_{l_s,i}|^2, \tag{5.4.6}$$

$$\hat{V}_l := \sum_{i=1}^{p} \frac{1}{N} \sum_{k=1}^{N} |\delta x^k_{l,i} - \delta \hat{x}_{l,i}|^2, \quad l = l_s + 1, \dots, lopt. \tag{5.4.7}$$

Par conséquent la taille N_l optimale pour $l = l_s, \dots, lopt$ est

$$N_l = \frac{1}{\eta^2 - e(sc, \Delta t_{lopt})} \sqrt{V_l 2^{-l}} \sum_{k=l_s}^{l} \sqrt{V_k 2^k}. \tag{5.4.8}$$

Dans le reste de ce chapitre nous allons calculer numériquement ces coûts optimaux pour chacun des deux schémas semi-implicite et explicite d'Euler.

5.5 Schéma semi-implicite d'Euler

Les approximations numériques des EDSM ont été étudiées dans [1, 3, 15, 17]. Voir aussi [10, 11, 22, 23, 24], dans le cas où la dérive est régulière.

Le schéma semi-implicite d'Euler (SIES) de l'équation (5.1.1) est défini par

$$\mathbf{x}_l(k+1) - \mathbf{x}_l(k) = -\nabla\varphi(\mathbf{x}_l(k+1))\Delta t_l - b(\mathbf{x}_l(k))\Delta t_l + \sqrt{\Delta t_l}\mathbf{n}(k+1),$$

où $(\mathbf{n}(k+1) : \quad k = 0, 1, \dots, 2^l - 1)$ est une suite i.i.d. de vecteurs gaussiens standards. Sachant $\mathbf{x}_l(k)$ et $\mathbf{n}(k+1)$, nous avons

$$\mathbf{x}_l(k+1) = prox_{\Delta t_l \varphi}(\mathbf{x}_l(k) - b(\mathbf{x}_l(k))\Delta t_l + \sqrt{\Delta t_l}\mathbf{n}(k+1)). \tag{5.5.1}$$

Les erreurs faible et forte de ce schéma sont définies par

$$e_w(\Delta t_l) = \|\mathbb{E}[\mathbf{x}(T) - \mathbf{x}_l(2^l)]\|_2, \tag{5.5.2}$$

$$e_s(\Delta t_l) = \mathbb{E}[\|\mathbf{x}(T) - \mathbf{x}_l(2^l)\|_2^2]^{\frac{1}{2}}. \tag{5.5.3}$$

D'après ([3]) l'erreur forte de l'équation (5.5.3) est estimée par

$$O((\Delta t_l \ln(\frac{1}{\Delta t_l})^{\frac{1}{4}}). \tag{5.5.4}$$

En prenant $\varphi(\mathbf{x}) = \|\mathbf{x}\|_1$, $\mathbf{b}(\mathbf{x}) = \mathbf{A}^*(\mathbf{A}\mathbf{x} - \mathbf{y})$, le schéma (5.5.1) connu sous le nom de l'algorithme STMALA ([7]).

5.6 Schéma explicite d'Euler

5.6.1 Le schéma EES1

En prenant $\varphi(\mathbf{x}) = \|\mathbf{x}\|_1 + \frac{\|\mathbf{Ax}-\mathbf{y}\|_2^2}{2}$, and $\mathbf{b} = 0$, le schéma EES1 de (5.0.1) est défini par

$$\mathbf{x}_l(k+1) = prox_{\Delta t_l \varphi}(\mathbf{x}_l(k)) + \sqrt{\Delta t_l}\mathbf{n}_{k+1}.$$

Le vecteur $prox_{\Delta t_l \varphi}(\mathbf{x}^{(k)})$ n'est pas explicite, mais pour l grand, nous avons

$$prox_{\Delta t_l \varphi}(\mathbf{x}) \approx prox_{\Delta t_l \|\cdot\|_1}(\mathbf{x} + \mathbf{A}^*(\mathbf{y} - \mathbf{Ax})\Delta t_l).$$

Finalement, nous obtenons le schéma

$$\mathbf{x}_l(k+1) = prox_{\Delta t_l \|\cdot\|_1}\Big(\mathbf{x}_l(k) + \mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{x}_l(k))\Delta t_l\Big) + \sqrt{\Delta t_l}\mathbf{n}_{k+1}, \tag{5.6.1}$$

connu sous le nom PULA algorithm [16].

5.6.2 Le schéma EES2

En prenant $\varphi(\mathbf{x}) = \|\mathbf{x}\|_1$, et $\mathbf{b}(\mathbf{x}) = \mathbf{A}^*(\mathbf{Ax} - \mathbf{y})$, nous obtenons le schéma

$$\mathbf{x}_l(k+1) = prox_{\Delta t_l \|\cdot\|_1}(\mathbf{x}_l(k)) + \mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{x}_l(k)\Delta t_l + \sqrt{\Delta t_l}\mathbf{n}_{k+1}). \tag{5.6.2}$$

5.7 Implémentation numérique

Comme illustration, nous considérons le cas $p = 10$, $n = 7$ et les entrées de la matrice $\mathbf{A}$ sont i.i.d. de loi de Bernoulli à valeurs $\pm\frac{1}{\sqrt{n}}$, $\mathbf{w} \sim \mathcal{N}(0, \frac{1}{2}\mathbf{I})$. Nous simulons un vecteur $\mathbf{x}(true)$ selon la densité proportionnelle à $\exp(-2\|\mathbf{x}\|_1)$. Nous obtenons le vecteur $\mathbf{y} := \mathbf{Ax}(true) + \mathbf{w}$ à partir d'une réalisation de $\mathbf{A}$ et $\mathbf{w}$. L'horizon du temps $T = 10$, le niveau maximal est $L = 16$, et le plus petit niveau est $l_s = 5$.

5.7.1 Simulation des trajectoires pour chaque schéma

Pour chaque schéma sc et pour chaque niveau $l = l_s, l_s + 1, l_s + 2$, nous avons représenté dans les figures 5.1, 5.2 et 5.3 la trajectoire $k \in [0, 2^l] \to \mathbf{x}_l(sc, k)$. Pour $L = 16$, nous avons représenté dans la figure 5.4 uniquement la première composante.

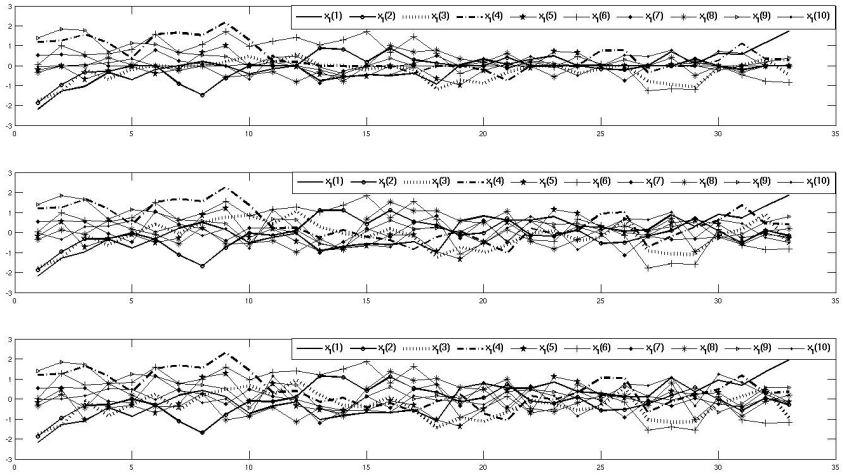

FIGURE 5.1 – Les chaînes de SIES, EES1 et EES2 pour $l = l_s$.

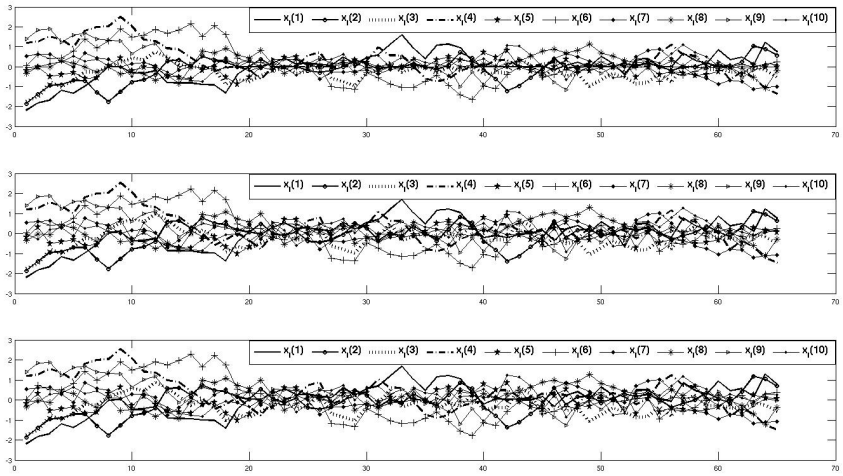

FIGURE 5.2 – Les chaînes de SIES, EES1 et EES2 pour $l = l_s + 1$.

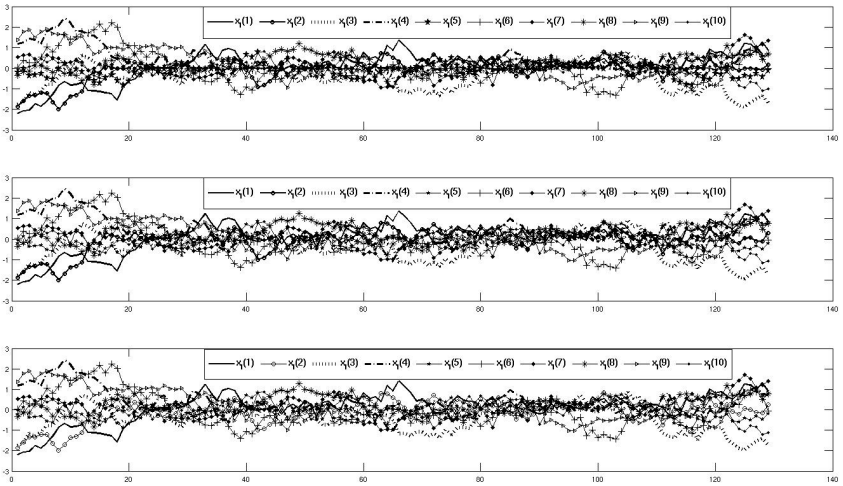

FIGURE 5.3 – les chaînes de SIES, EES1 et EES2 pour $l = l_s + 2$.

5.7.2 Coût optimal de la méthode Monte Carlo

Nous approximons pour chaque schéma $\mathbf{x}(T)$ par $\mathbf{x}_L(sc, 2^L)$ avec $L = 16$, et nous cherchons le niveau optimal $lopt$ et la taille optimal $Nopt$ de l'échantillon tels que

$$MSE := \mathbb{E}[\|\mathbb{E}[\mathbf{x}_L(sc, 2^L)] - \frac{1}{Nopt}\sum_{k=1}^{Nopt} \mathbf{x}_l^k(sc, 2^l)\|_2^2] = \eta^2.$$

Nous avons besoin de calculer

$$e(sc, \Delta t_l) := \|\mathbb{E}[\mathbf{x}_L(sc, 2^L)] - \mathbb{E}[\mathbf{x}_l(sc, 2^l)]\|_2^2$$

pour $l = l_s, \ldots, L-2$. En utilisant Monte-Carlo avec $N = 1000$, nous obtenons

$$e(sc, \Delta t_l) \approx \|\frac{1}{N}\sum_{k=1}^{N} \mathbf{x}_L^k(sc, 2^L) - \frac{1}{N}\sum_{k=1}^{N} \mathbf{x}_l^k(sc, 2^l)\|_2^2.$$

Le tableau 5.1 montre le calcul numérique de $e(sc, \Delta t_l)$ pour chaque schéma et pour $l = 5, \ldots, 13$.

En fixant $\eta^2 \geq \max\Big(e(sc, \Delta t_l), sc = SIES, EES1, EES2, l = 5, \ldots, 13\Big)$, la contrainte $e(sc, l) \leq \eta^2$ a lieu pour chaque niveau $l = 5, \ldots, 13$. Le niveau

l	5	6	7	8	9	10	11	12	13
$e(SIES, \Delta_{t_l})$	0.0050	0.0080	0.0071	0.0022	0.0054	0.0066	0.0056	0.0043	0.0022
$e(EES1, \Delta_{t_l})$	0.0380	0.0025	0.0069	0.0043	0.0016	0.0039	0.0027	0.0032	0.0022
$e(EES2, \Delta_{t_l})$	0.0107	0.0041	0.0044	0.0042	0.0054	0.0111	0.0041	0.0048	0.0065

TABLE 5.1 – Valeurs numériques de $e(sc, \Delta t_l)$ pour chaque schéma et pour $l = 5, \ldots, 13$.

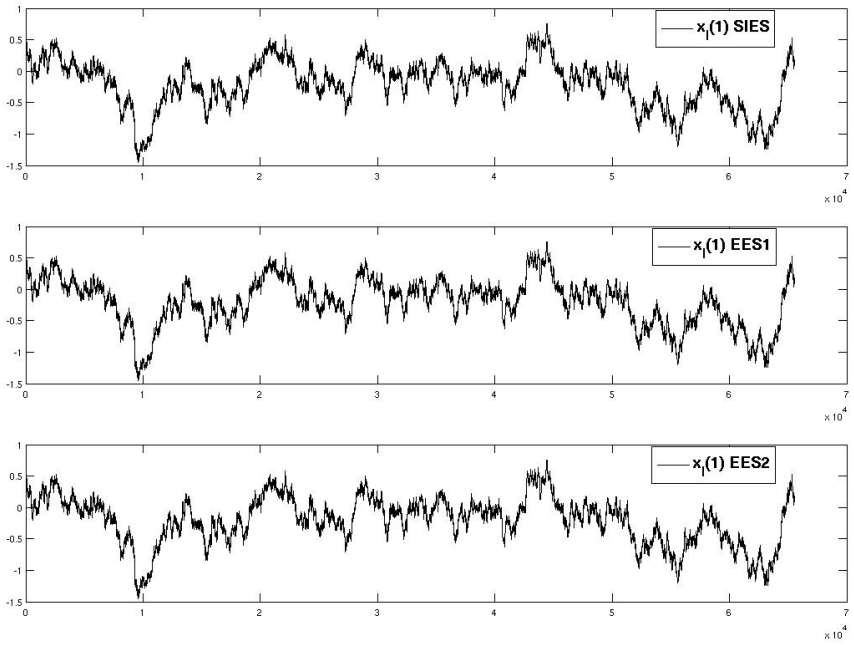

FIGURE 5.4 – La première composante des chaînes SIES, EES1 and EES2 pour $l = 16$.

optimal *lopt* est solution de l'équation

$$3e(sc, l) - 2e(sc, l + 1) \approx \eta^e.$$

Ayant $lopt$, nous calculons

$$Var_{lopt}(sc) := \sum_{i=1}^{p} Var(x_{lopt,i}(sc, 2^{lopt})),$$
$$\approx \sum_{i=1}^{p} \frac{1}{N} \sum_{k=1}^{N} |x^k_{lopt,i}(sc, 2^{lopt}) - \frac{1}{N} \sum_{k=1}^{N} x^k_{lopt,i}(sc, 2^{lopt})|^2,$$

et nous en déduisons la taille optimale $Nopt(sc) = \dfrac{Var_{lopt}(sc)}{\eta^2 - e(sc, \Delta t_{lopt})}$.

La figure 5.5 montre comment trouver graphiquement le niveau optimal $lopt$.

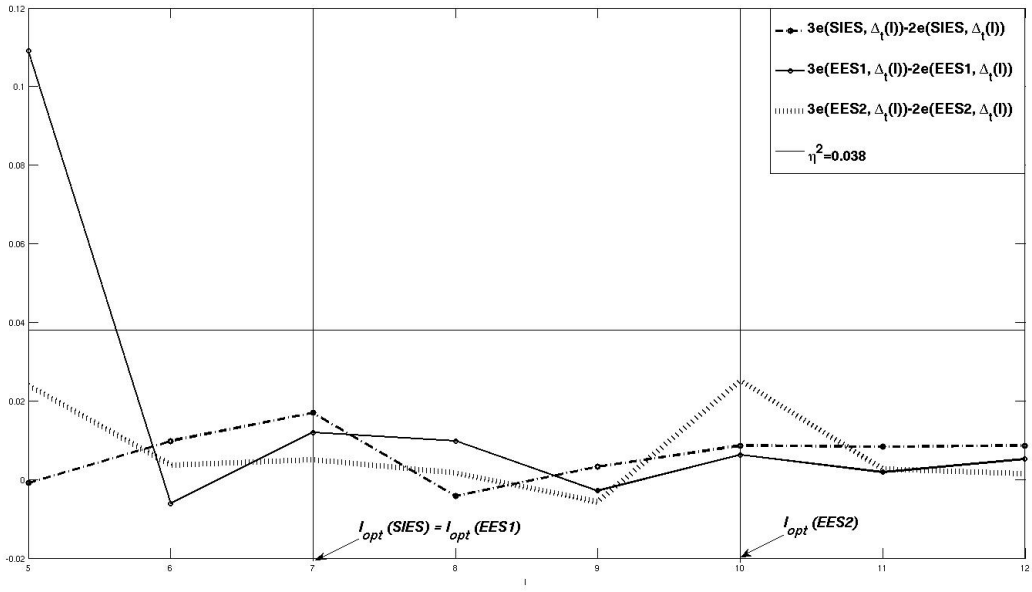

FIGURE 5.5 – Détermination graphique de l_{opt} pour les trois schémas SIES, EES1 et EES2.

Nous avons résumé dans le tableau 5.2 pour les trois schémas les valeurs $lopt$ et $Nopt$ et leur coût. Il ressort de ce tableau que le schéma SIES est le meilleur.

	$lopt$	$Nopt$	Coût
SIES	7	70	8938
EES1	7	81	10427
EES2	10	83	85035

TABLE 5.2 – Niveau optimal et coût de l'algorithme MC avec les trois schémas.

5.7.3 Coût optimal de la méthode MLMC

Dans la figure 5.6 pour chaque schéma nous avons tracé $l \to e(sc, \Delta t_l)$ (5.3.4). Nous en déduisons graphiquement le niveau optimal l_{opt}.

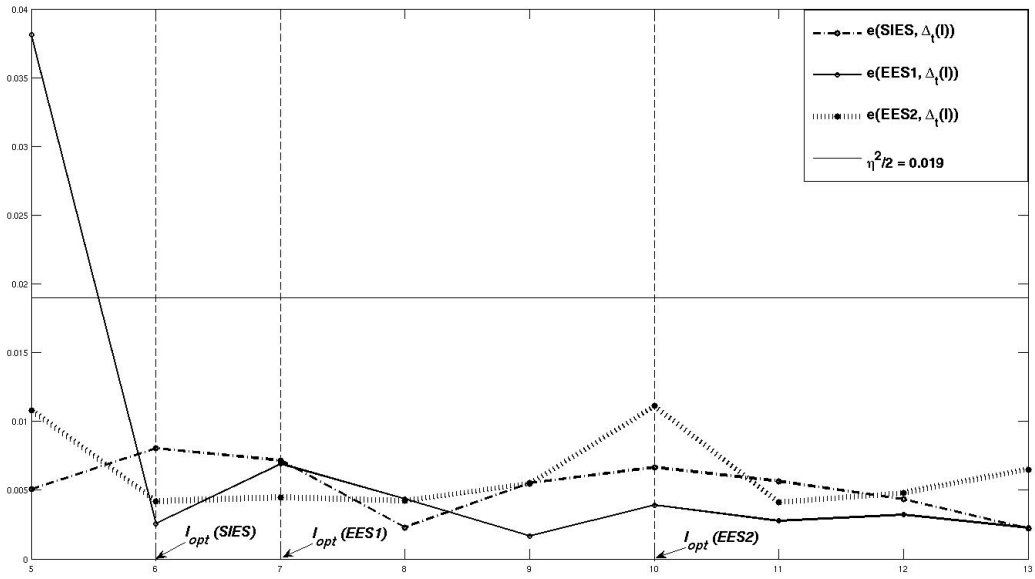

FIGURE 5.6 – Identification graphique de l_{opt} pour chaque schéma.

Nous avons résumé dans le tableau 5.3 pour les trois schémas les valeurs $lopt$ et l_{opt}, $N_{l_s}(opt), \ldots, N_{lopt}(opt)$ ainsi que leur coût. Comme pour la méthode MC, le schéma SIES est gagnant.

	$lopt$	$N_{5-lopt}(opt)$	Coût
SIES	6	74 20	3639.18
EES1	7	132 40 16	8962.85
EES2	10	167 59 23 9 4 2	18029.47

TABLE 5.3 – Niveau optimal et coût de MLMC pour chaque schéma.

N.B. Pour chaque valeur lopt, les tailles optimales des échantillons sont $N_{5-lopt} := N_5(opt), \ldots, N_{lopt}(opt)$, e.g. pour le schéma SIES $lopt = 6$ et $N_{5-6} = 74, 20$.

5.8 MCMC en utilisant les schémas numériques

Nous considérons les chaînes de Markov $MC(k) := \mathbf{x}_{lopt}(sc, k)$ avec $sc = EES1, EES2$ et $MC(k) = RW(k, \sigma^2)$ (marche aléatoire gaussienne de variance σ^2). Nous construisons l'algorithme de Hasting $MCMC(k)$ ayant $(MC(k))$ comme loi de proposition et la loi cible proportionnelle à

$$\exp\Big(- 2(\|\mathbf{x}\|_1 + \frac{\|\mathbf{Ax} - \mathbf{y}\|_2^2}{2})\Big).$$

Nous définissons pour η^2 fixé, le coût de l'algorithme de Hasting est égal au nombre d'itérations N telle que

$$\mathbb{E}\left[\|\mathbb{E}[\mathbf{x}(T)] - \frac{1}{N}\sum_{k=1}^{N} MCMC(k)\|_2^2\right] \approx \eta^2.$$

Nous obtenons trois chaîne $MCMC_{prox}(EES1)$, $MCMC_{prox}(EES2)$, $MCMC_{RW}$. Le tableau 5.4 donne le coût de chaque schéma.

5.8.1 Détails du calcul numérique

Nous répétons le même algorithme M fois et nous notons $(MCMC^{(i)} : i = 1, \ldots, M)$ les échantillons obtenus. Nous approximons

$$\mathbb{E}\left[\|\mathbb{E}[\mathbf{x}(T)] - \frac{1}{N}\sum_{k=1}^{N} MCMC(k)\|_2^2\right]$$

par

$$\frac{1}{M}\sum_{i=1}^{M} \|\mathbb{E}[\mathbf{x}_L(sc, 2^L)] - \frac{1}{N}\sum_{k=1}^{N} MCMC^i(k)\|_2^2.$$

L'espérance mathématique $\mathbb{E}[\mathbf{x}_L(sc, 2^L)]$ est calculée en utilisant la méthode MLMC.

	Coût (MC)	Coût ($MCMC_{prox}$)	Coût ($MCMC_{RW}$)	Coût ($MCMC_{RW}$)
EES1	10427	5340	3990 ($\sigma^2 = 0.3$)	17230 ($\sigma^2 = 0.8$)
EES2	85035	6200	3890 ($\sigma^2 = 0.3$)	16230 ($\sigma^2 = 0.8$)

TABLE 5.4 – Le coût de MC, $MCMC_{prox}$ et $MCMC_{RW}$ en utilisant les schémas EES1 et EES2 pour $\eta^2 = 10^{-2}$.

Le tableau 5.4 montre que $MCMC_{RW}$ avec la loi de proposition $\mathcal{N}(0, 0.3)$ est l'algorithme gagnant. Mais $MCMC_{RW}$ perd contre la méthode MLMC avec le schéma SIES (voir le tableau 5.2).

5.9 Appendice

5.9.1 L'approximation de Yosida

Soit $\varphi : \mathbb{R}^p \to (-\infty, +\infty]$ une fonction convexe et semi continue inférieurement. L'ensemble $\mathcal{P}(\mathbb{R}^p)$ représente les parties de $\mathbb{R}^p$. La sous différentielle $\partial\varphi$

$$\partial\varphi(\mathbf{x}) = \{\mathbf{v} \in \mathbb{R}^p : \quad \varphi(\mathbf{x} + \boldsymbol{h}) \geq \varphi(\mathbf{x}) + \langle h, \mathbf{v}\rangle, \, \forall \, \boldsymbol{h} \in \mathbb{R}^p\}.$$

Son domaine

$$Dom(\partial\varphi) = \{\mathbf{x} : \quad \partial\varphi(\mathbf{x}) \neq \emptyset\}.$$

La régularisation de la sous différentielle $\partial\varphi(\mathbf{x})$ est basée sur l'approximation de Yosida. Pour chaque $\varepsilon > 0$ et $\mathbf{x} \in \mathbb{R}^p$, l'équation

$$\mathbf{x} = \mathbf{z} + \varepsilon\partial\varphi(\mathbf{z})$$

a une unique solution

$$\begin{aligned} \mathbf{z} &= (\mathbf{I} + \varepsilon\partial\varphi)^{-1}(\mathbf{x}) \\ &:= prox_{\varepsilon\varphi}(\mathbf{x}). \end{aligned}$$

Exemple. $\varphi(\mathbf{x}) = \|\mathbf{x}\|_1$. D'abord en dimension $p = 1$, $\varphi(x) = |x|$, nous avons

$$prox_{\varepsilon|\cdot|}(x) = (x + \varepsilon)\mathbf{1}_{[x \leq -\varepsilon]} + (x - \varepsilon)\mathbf{1}_{[x \geq 1]},$$

et alors pour $p \geq 2$

$$prox_{\varepsilon\|\cdot\|_1}(\mathbf{x}) = (prox_{\varepsilon|\cdot|}(x_1), \ldots, prox_{\varepsilon|\cdot|}(x_p)).$$

L'application $prox_{\varepsilon\varphi} : \mathbb{R}^p \to Dom(\partial\varphi)$ est appelée la fonction proximale. L'approximation de Yosida de $\partial\varphi$ est l'application

$$\beta_\varepsilon(\mathbf{x}) := \frac{\mathbf{x} - prox_{\varepsilon\varphi}(\mathbf{x})}{\varepsilon}.$$

Le résultat suivant est bien connu (voir e.g. [15]).

Proposition 5.9.1. *Nous avons*

1. *$prox_{\varepsilon\varphi}$ est une contraction de $\mathbb{R}^p$ à valeurs dans $Dom(\partial\varphi)$.*
2. *β_ε est monotone sur $\mathbb{R}^p$, i.e.*

$$\langle \beta_\varepsilon(\mathbf{x}_1) - \beta_\varepsilon(\mathbf{x}_2), \mathbf{x}_1 - \mathbf{x}_2 \rangle \geq 0,$$

   pour $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^p$, et elle est lipschitzienne avec la constante de Lipschitz $\frac{1}{\varepsilon}$.
3. *Pour $\mathbf{x} \in \mathbb{R}^p$, $\beta_\varepsilon(\mathbf{x}) \in \partial\varphi(prox_{\varepsilon\varphi}(\mathbf{x}))$.*

Nous avons aussi le résultat suivant.

Proposition 5.9.2. *Pour $\varepsilon > 0$, l'application*

$$\mathbf{x} \in \mathbb{R}^p \to \varphi_\varepsilon(\mathbf{x}) = \min\{\varphi(\mathbf{z}) + \frac{\|\mathbf{x} - \mathbf{z}\|_2^2}{2\varepsilon}\}$$

est l'approximation de Yosida de la fonction φ. Nous avons les propriétés suivantes.

1. *φ_ε est convexe dont le domaine $\mathbb{R}^p$.*
2. *φ_ε est de classe C^1 avec $\nabla\varphi_\varepsilon = \beta_\varepsilon$.*
3. *L'argmin de $\varphi_\varepsilon(\mathbf{x})$ est atteint au point $prox_{\varepsilon\varphi}(\mathbf{x})$, et*

$$\varphi_\varepsilon(\mathbf{x}) = \frac{\varepsilon}{2}\|\beta_\varepsilon(\mathbf{x})\|_2^2 + \varphi_\varepsilon(prox_{\varepsilon\varphi}(\mathbf{x})).$$

4. *Nous avons $\varphi_\varepsilon \uparrow \varphi(\mathbf{x})$ pour tout $\mathbf{x} \in \mathbb{R}^p$ lorsque $\varepsilon \downarrow 0$.*

5.9.2 Bang-bang Brownian motion

Dans le cas unidimensionnel

$$dx(t) = -sgn(x(t))dt + dw(t), \quad x(0) = x_0,$$

est connue sous le nom "bang-bang Brownien motion" [13], où la diffusion avec potentiel V [19]. Dans ce cas les probabilités de transition du processus $x\ p(x,t\,|\,x_0,0)$ sont connues [4]. Nous pouvons calculer en utilisant la formule de Girsanov et la loi du triplet mouvement brownien, son temps local et son temps d'occupation ([13]) pour obtenir

$$p(x,t\,|\,x_0) = q(x,t\,|\,x_0)\exp(-2|x|)$$

où

$$q(x,t\,|\,x_0) = \exp\Big(|x_0| + |x|) - \frac{t}{2}\Big)\gamma_t(x - x_0) + F\Big(\frac{t - (|x| + |x_0|)}{\sqrt{t}}\Big),$$

$$F(x) = \int_{-\infty}^{x} \frac{exp(-\frac{u^2}{2})}{\sqrt{2\pi}}du,$$

$$\gamma_t(u) = \frac{exp(-\frac{u^2}{2t})}{\sqrt{2t\pi}}.$$

Nous observons que $p(x,t\,|\,x_0,0) \to \exp(-2|x|)$ lorsque $t \to +\infty$ pour tout x_0. Par conséquent l'EDSM

$$dx(t) = -sgn(x(t))dt + dw(t)$$

est ergodique et ayant la densité invariante $\exp(-2|x|)$.

Conclusion

Dans ce chapitre nous avons approximé la loi du LASSO bayésien en utilisant les méthodes de Monte Carlo, de Monte Carlo à plusieurs niveaux et la méthode MCMC en utilisant les schémas numériques semi-implicite et explicite d'Euler. Nous avons proposé un critère pour calculer le coût optimal de calcul pour chaque méthode. Il ressort de cette étude que la méthode de Monte Carlo à plusieurs niveaux associée au schéma semi-implicite d'Euler est la méthode optimale.

Bibliographie

[1] I. Asiminoaei, A. Rascanu, Approximation and simulation of stochastic variational inequalities-splitting up, Method. Numer. Funct. Anal. and Optimiz. 18 231–282 (1997).

[2] A. Bensoussan, A. Rascanu. Stochastic variational inequalities in infinite dimensional, space. Numer. Funct. Anal. and Optimiz. vol. 18 19–54 (1997).

[3] F. Bernardin, Multivaled stochastic differential equations : convergence of a numerical scheme, Set-Valued Analysis vol. 11 393–415 (2003).

[4] A.N. Borodin, P. Salminen, Handbook of brownian motion-facts and formulas, second edition Birkhäuser (2002).

[5] E. Cépa, Equations différentielles stochastiques multivoques, Thèse Université d'Orléans (1994).

[6] E. Cépa, Problème de Skorohod multivoque, The Annals of Probability vol. 26 no. 2 500–532 (1998).

[7] G. Fort and S. Le Corff, E. Moulines and A. Schreck, A shrinkage-thresholding Metropolis adjusted Langevin algorithm for bayesian variable selection, arXiv :1312.5658v3 [math.ST] (2015).

[8] M.B. Giles, Multilevel Monte Carlo path simulation, Oper. Res. vol. 56 607–617 (2008).

[9] M.B. Giles, Multilevel Monte Carlo methods, Cambridge University Press (2015).

[10] E. Gobet, P.Turkedjiev, Linear regression MDP scheme for discrete backward stochastic differential equations under general conditions, hal-00855760v2 (2014).

[11] C. Graham and D. Talay, Stochastic simulation and Monte Carlo Methods, Mathematical Foundations of Stochastic Simulation, Series : Stochastic Modelling and Applied Probability vol. 68. Springer (2013).

[12] A.G. Jaeschke and M. Yor, A survey and some generalizations of Bessel Processes, Bernoulli 9(2) p. 31–349 (2002).

[13] I. Karatzas and S. Shreve, Trivariate density of Brownian motion, its local and occupation timesvwith application to stochastic control, The Annals of Probability vol. 12 819–828 (1984).

[14] P. Kree, Diffusion equation for multivalued stochastic differential equations, J. Funct Anal. vol. 49 73–90 (1982).

[15] D. Lepingle and T.T. Nguyen, Approximation and simulating multivalued stochastic differential equations, https ://hal.archives-ouvertes.fr/hal-00003500 (2004).

[16] M. Pereyra, Proximal Markov chain Monte Carlo algorithms, arXiv :1306.0187v3, [stat.ME] 3 Jul 2014.

[17] R. Pettersson, Projection scheme for stochastic differential equations with convex constraints, Stochastic Process. Appl. vol. 88 125–134 (2000).

[18] J. Rena, J. Wua and X. Zhang, Exponential ergodicity of non-Lipschitz multivalued stochastic differential equations, Bulletin des Sciences Mathématiques, Vol. 134 Issue 4 P. 391–404 (2010).

[19] H. Risken, The Fokker-Planck equation, Springer (1984).

[20] M.S. Rosin, L.F. Ricketon, A.M. Dimits, R.E. Caflish and B.I. Cohen, Multivel Monte Carlo simulation of Coulomb collisions, Journal of Computational Physics vol. 274 140–157 (2014).

[21] A. Storm, Stochastic differential equations with convex constraint, Stochastics and Stochastics Reports 53 241–247 (1995).

[22] D.Talay, Discretization of stochastic differential equations : Application to simulation, Stochastic Numerical Methods for Partial Differential Equations, In ENUMATH 99-Proceedings of the 3rd European Conference on Numerical Mathematics and Advanced Applications, Jyväskylä, Finland, P. Neittaanmäki, T. Tiihonen and P. Tarvainen (Eds.), World Scientific Singapore (2000).

[23] D. Talay, Simulation of Stochastic Processes and Applications, In Proceedings of FOCM 99, R. DeVore and A. Isarles (Eds.), Cambridge University Press (2000).

[24] D. Talay, Simulation and Numerical Analysis of Stochastic Differential Systems : a Review. In Probabilistic Methods in Applied Physics, P. Kree and W. Wedig (Eds.), vol. 451 of Lecture Notes in Physics, chapter 3, 63–106 Springer-Verlag (1995).

Chapitre 6

Oscillation of Metropolis-Hastings and simulated annealing algorithms around LASSO estimator

In this chapter we study, as the temperature goes to zero, the oscillation of a family of Gibbs measures around LASSO estimator. We derive new criteria for estimating LASSO, choosing the proposal distribution and the temperature in Metropolis-Hastings algorithm. Finally we apply these results to analyse the convergence of Metropolis-Hastings and simulated annealing algorithms.

6.1 Least absolute shrinkage and selection operator (LASSO)

Let $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w}$ be the classical linear regression problem, where $\mathbf{y} \in \boldsymbol{R}^n$ are the observations, $\mathbf{x} \in \boldsymbol{R}^p$ is the unknown signal to recover, $\mathbf{w} \in \boldsymbol{R}^n$ is the noise, and $\mathbf{A}$ be a known matrix which maps the signal domain $\boldsymbol{R}^p$ into the observation domain $\boldsymbol{R}^n$ with $n \leq p$. The matrix $\mathbf{A}$ is in general ill-conditioned, which makes difficult to use the least squares estimate. Penalization is a popular way to compute an approximation of $\mathbf{x}$ from the observations $\mathbf{y}$. The general framework proposes to recover the vector $\mathbf{x}$

using the minimization

$$\mathbf{x}(\mathbf{y},t) \in \arg\min\{P(\mathbf{x}) + \frac{\|\mathbf{A}\mathbf{x}-\mathbf{y}\|^2}{2t} : \quad \mathbf{x} \in \boldsymbol{R}^p\}. \tag{6.1.1}$$

Here $\|\cdot\|$ denotes the Euclidean norm. This requires to define a penalization P to enforce some prior information on the signal $\mathbf{x}$. The term $\frac{\|\mathbf{A}\mathbf{x}-\mathbf{y}\|^2}{2}$ reflects Gaussian prior on the noise $\mathbf{w}$. The parameter $t > 0$ reflects the noise level and the degree of sparsity of the signal $\mathbf{x}$.

The l^1 penalization i.e. the sum of the absolute values $P(\mathbf{x}) = \|\mathbf{x}\|_1$ of the components of $\mathbf{x}$, which is the focus of this paper, was first introduced in [16] and called LASSO. It is also called Basis Pursuit De-Noising method [5]. It was introduced to induce sparsity in the minimizer $\mathbf{x}(\mathbf{y},t)$ (6.1.1). A large number of theoretical results has been provided for the l^1 penalization. In our work we select the term LASSO and we keep it for the rest of the article. The most popular algorithms to find LASSO estimator are LARS algorithm [6], ISTA and FISTA algorithms see e.g. [2] and the review article [14].

In our work we consider the family of probabilities (called also Gibbs measures)

$$P_T^{\mathbf{y},t} := \frac{\exp\left(-\frac{1}{T}F(\mathbf{x},\mathbf{y},t)\right)d\mathbf{x}}{\int_{\boldsymbol{R}^p}\exp\left(-\frac{1}{T}F(\mathbf{x},\mathbf{y},t)\right)d\mathbf{x}}, \tag{6.1.2}$$

where $T > 0$ is called the temperature and

$$F(\mathbf{x},\mathbf{y},t) = \|\mathbf{x}\|_1 + \frac{\|\mathbf{A}\mathbf{x}-\mathbf{y}\|^2}{2t}$$

is called the objective function. The family of the probabilities (6.1.2) oscillates around the set of LASSO estimators as $T \to 0$. More precisely, any sequence $(\mathbf{P}_{T_k}^{\mathbf{y},t} : T_k \to 0)$ is tight [11], [1] i.e. we can extract a convergent subsequence from $(P_{T_k}^{\mathbf{y},t})$. If

$$P_{T_k}^{\mathbf{y},t} \to P^{\mathbf{y},t},$$

then $P^{\mathbf{y},t}$ concentrates on $\arg\min\{F(\mathbf{x},\mathbf{y},t) : \quad \mathbf{x} \in \boldsymbol{R}^p\}$. The contents of the present work are organized as follows. In Section 2 we recall some well known results of LASSO estimator and we establish preliminary results. The main body of this paper consists of Section 3, i.e. we give a precise scaling of the asymptotic of the measures (6.1.2) as $T \to 0$. Using this scaling and

Metropolis-Hastings algorithm with the target (6.1.2) we show how to estimate LASSO. In Sections 4-7 we investigate Metropolis-Hastings algorithm with small temperature. We establish new criteria of the choice of the proposal distribution and the temperature. We also analyse the convergence of Metropolis-Hastings algorithm using our new criteria. In Section 8 we apply these criteria to analyse the simulated annealing algorithms.

6.2 LASSO estimator properties

First, we need some notations. The column vector $sgn(\mathbf{x}) \in \boldsymbol{R}^p$ has the components $sgn(x_i) = 1$ if $x_i > 0$, $sgn(x_i) = -1$ if $x_i < 0$ and $sgn(0)$ is any element of $[-1, 1]$. We will denote, for each subset $I \subset \{1, \dots, p\}$ and for each vector $\mathbf{v} \in \boldsymbol{R}^p$, $\mathbf{v}(I) = (v(i) : i \in I) \in \boldsymbol{R}^{|I|}$. Here $|I|$ denotes the cardinality of I. The notation $\mathbf{v} \leq \mathbf{w}$ means $v(i) \leq w(i)$ for all $i = 1, 2, \dots, p$. The scalar product is denoted by $\langle \cdot, \cdot \rangle$, and $(\mathbf{e}_i : \quad i = 1, 2, \dots, p)$ denotes the canonical basis of $\boldsymbol{R}^p$.

Now we recall a well known properties of LASSO estimator see e.g. [17].

Lemme 6.2.1. *The vector $\mathbf{x}(\mathbf{y}, t)$ is a minimizer of the map $\mathbf{x} \to \|\mathbf{x}\|_1 + \frac{\|\mathbf{A}\mathbf{x}-\mathbf{y}\|^2}{2t}$ if and only if the vector $\xi(\mathbf{y}, t) := \frac{\mathbf{A}^*(\mathbf{y}-A\mathbf{x}(\mathbf{y},t))}{t}$ is equal to $sgn(\mathbf{x}(\mathbf{y}, t))$. The vectors $\xi(\mathbf{y}, t)$, $\mathbf{A}\mathbf{x}(\mathbf{y}, t)$ and the l^1-norm $\|\mathbf{x}(\mathbf{y}, t)\|_1$ are constant on the set of LASSO estimators. Here $\mathbf{A}^*$ denotes the transpose of the matrix $\mathbf{A}$.*

The sets $I_0 = \{i \in \{1, \dots, p\} : \quad \mathbf{x}_i(\mathbf{y}, t) = 0\}$, $\partial I_0 = \{i \in I_0 : |\xi_i(\mathbf{y}, t)| = 1\}$ will play an important role in the Gibbs measures scaling. The set $S = \{1, \dots, p\} \setminus I_0$ is the support of the LASSO estimator $\mathbf{x}(\mathbf{y}, t)$ i.e. $S = \{i \in \{1, \dots, p\} : \mathbf{x}_i(\mathbf{y}, t) \neq 0\}$. If the LASSO estimator $\mathbf{x}(\mathbf{y}, t)$ is not unique, then the sets $I_0, \partial I_0$ and S depend on $\mathbf{x}(\mathbf{y}, t)$.

Lemme 6.2.2. *If the matrix $[\langle \mathbf{A}\mathbf{e}_i, \mathbf{A}\mathbf{e}_j \rangle, \quad i, j \in (S \cup \partial I_0)]$ is invertible, then the set of LASSO estimators is a singleton.*

Preuve 6.2.3. *This result is known see e.g. [17], but for the sake of completeness we give the proof. Observe that the invertibility of the matrix $[\langle \mathbf{A}\mathbf{e}_i, \mathbf{A}\mathbf{e}_j \rangle, \quad i, j \in (S \cup \partial I_0)]$ is equivalent to say that the linear operator $\mathbf{A}_{S \cup \partial I_0} : \boldsymbol{R}^{S \cup \partial I_0} \to \boldsymbol{R}^n$ is injective. Here $\mathbf{A}_{S \cup \partial I_0}$ denotes the sub-matrix of $\mathbf{A}$ having the columns indexed by $S \cup \partial I_0$. The inverse of $\mathbf{A}_{S \cup \partial I_0}$ defined from $\boldsymbol{R}^{S \cup \partial I_0}$ into its range $R(\mathbf{A}_{S \cup \partial I_0})$ is denoted by $\mathbf{A}^{-1}_{S \cup \partial I_0}$. Now, we recall a result of Grasmair et al. [8] Lemma 3.10. Let $M(\mathbf{x}(\mathbf{y}, t)) := \max\{|\xi_i(\mathbf{y}, t))| : i \in I_0 \setminus \partial I_0\}$, and for any couple $\mathbf{x}^{(1)}, \mathbf{x}^{(2)} \in \boldsymbol{R}^p$, let us define, for some fixed*

$\xi \in sgn(\mathbf{x}^{(2)})$, $D(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}) := \|\mathbf{x}^{(1)}\|_1 - \|\mathbf{x}^{(2)}\|_1 - \langle \xi, \mathbf{x}^{(1)} - \mathbf{x}^{(2)} \rangle$. *The result of Grasmair et al. tells us that for all* x,

$$\|\mathbf{x} - \mathbf{x}(\mathbf{y},t)\| \leq \|\mathbf{A}^{-1}_{S\cup\partial I_0}\| \|\mathbf{A}(\mathbf{x} - \mathbf{x}(\mathbf{y},t))\| + \frac{1 + \|\mathbf{A}^{-1}_{S\cup\partial I_0}\| \|\mathbf{A}\|}{1 - M(\mathbf{x}(\mathbf{y},t))} D(\mathbf{x}, \mathbf{x}(\mathbf{y},t)),$$

where $\|\mathbf{B}\|$ *denotes the operator norm of the matrix* $\mathbf{B}$. *If* $\mathbf{x}$ *is another LASSO, then from Lemma (6.2.1), we have* $\mathbf{Ax} = \mathbf{Ax}(\mathbf{y},t)$ *and* $D(\mathbf{x}, \mathbf{x}(y,t)) = 0$, *which achieves the proof.*

In the sequel $X_T(\mathbf{y},t)$ will denote a random vector having the probability distribution (6.1.2). If the set of LASSO estimators is a singleton $\mathbf{x}(\mathbf{y},t)$, then we can show that $X_T(\mathbf{y},t) \to \mathbf{x}(\mathbf{y},t)$ in probability as $T \to 0$ see e.g. [1].

We present two preliminary propositions prior to our principal result.

Proposition 6.2.4. *Let* $\mathbf{x}(\mathbf{y},t)$ *be any LASSO estimator and* $m(\mathbf{y},t) = F(\mathbf{x}(\mathbf{y},t),\mathbf{y},t)$ *be the minimum of the objective function* $F(\mathbf{x},\mathbf{y},t)$. *The function* $F(\mathbf{x},\mathbf{y},t) - m(\mathbf{y},t)$ *is equal to*

$$\sum_{i=1}^{p} |x_i| \left(1 - sgn(x_i)\xi_i(\mathbf{y},t)\right) + \frac{\|\mathbf{A}(\mathbf{x} - \mathbf{x}(\mathbf{y},t))\|^2}{2t}. \tag{6.2.1}$$

If, for $i \in S$, $\mathbf{x}_i$ *is near to* $\mathbf{x}_i(\mathbf{y},t)$, *then* $F(\mathbf{x},\mathbf{y},t) - m(\mathbf{y},t)$ *becomes*

$$\sum_{i\in I_0} |x_i| \left(1 - sgn(x_i)\xi_i(\mathbf{y},t)\right) + \frac{\|\mathbf{A}(\mathbf{x} - \mathbf{x}(\mathbf{y},t))\|^2}{2t}. \tag{6.2.2}$$

Preuve 6.2.5. *From the equality* $\|\mathbf{Ax} - \mathbf{y}\|^2 = \|\mathbf{A}(\mathbf{x} - \mathbf{x}(\mathbf{y},t))\|^2 + 2\langle \mathbf{A}(\mathbf{x} - \mathbf{x}(\mathbf{y},t)), \mathbf{Ax}(\mathbf{y},t) - \mathbf{y}\rangle + \|\mathbf{Ax}(\mathbf{y},t) - \mathbf{y}\|^2$, *we have*

$$\begin{aligned} &F(\mathbf{x},\mathbf{y},t) = \\ &\|\mathbf{x}\|_1 + \frac{\|\mathbf{A}(\mathbf{x} - \mathbf{x}(\mathbf{y},t))\|^2}{2t} + \frac{\langle \mathbf{A}(\mathbf{x} - \mathbf{x}(\mathbf{y},t)), \mathbf{Ax}(\mathbf{y},t) - \mathbf{y}\rangle}{t} + \frac{\|\mathbf{Ax}(\mathbf{y},t) - \mathbf{y}\|^2}{2t} \\ &= \|\mathbf{x}\|_1 + \frac{\|\mathbf{A}(\mathbf{x} - \mathbf{x}(\mathbf{y},t))\|^2}{2t} + \frac{\langle \mathbf{x} - \mathbf{x}(\mathbf{y},t), \mathbf{A}^*(\mathbf{Ax}(\mathbf{y},t) - \mathbf{y})\rangle}{t} + \frac{\|\mathbf{Ax}(\mathbf{y},t) - \mathbf{y}\|^2}{2t}. \end{aligned}$$

From the equality $\xi(\mathbf{y},t) = \frac{\mathbf{A}^*(y - \mathbf{Ax}(\mathbf{y},t))}{t}$ *(see Lemma 6.2.1), we have*

$$\begin{aligned} \frac{\langle \mathbf{x} - \mathbf{x}(\mathbf{y},t), \mathbf{A}^*(\mathbf{Ax}(\mathbf{y},t) - \mathbf{y})\rangle}{t} &= -\langle \mathbf{x} - \mathbf{x}(\mathbf{y},t), \xi(\mathbf{y},t)\rangle \\ &= -\langle \mathbf{x}, \xi(\mathbf{y},t)\rangle + \|\mathbf{x}(\mathbf{y},t)\|_1. \end{aligned} \tag{6.2.3}$$

Now formulas (6.2.1) and (6.2.2) are an easy consequence of the formula (6.2.3).

Now we can announce our last proposition.

Proposition 6.2.6. *Let $\partial I_0 = K_1 \cup K_2$ be a partition such that $K_1, K_2 \neq \emptyset$ and $E(K_1, K_2) := E_{-1}(K_1) \cap E_1(K_2)$ with $E_{-1}(K_1) = \{\mathbf{x} \in \boldsymbol{R}^p : sgn(x_i)\xi_i(\mathbf{y}, t) = -1, \forall\, i \in K_1\}$, and $E_{+1}(K_2) = \{\mathbf{x} \in \boldsymbol{R}^p : sgn(x_i)\xi_i(\mathbf{y}, t) = 1, \forall\, i \in K_2\}$. If $[\langle \mathbf{A}\mathbf{e}_i, \mathbf{A}\mathbf{e}_j\rangle, \quad i, j \in (S \cup \partial I_0)]$ is invertible, then the set of LASSO estimators is a singleton and the probability of the event $E_T(K_1, K_2) := [X_T(\mathbf{y}, t) \in E(K_1, K_2)]$ tends to 0 as $T \to 0$. As a consequence, we have $\mathbf{P}(E_T(\emptyset, \partial I_0)) \to 1$ as $T \to 0$.*

Preuve 6.2.7. *The uniqueness of LASSO is shown in the Lemma 6.2.2. Now, we prove the rest of our Proposition. We have $\mathbf{P}\left(X_T(\mathbf{y}, t) \in E(K_1, K_2)\right) = \frac{A_T(K_1, K_2)}{B_T}$, where*

$$A_T(K_1, K_2) = \int_{E(K_1, K_2)} \exp(-\frac{1}{T}F(\mathbf{x}, \mathbf{y}, t))d\mathbf{x}, \ \int_{\boldsymbol{R}^p} \exp(-\frac{1}{T}F(\mathbf{x}, \mathbf{y}, t))d\mathbf{x} = B_T.$$

We know that $X_T(\mathbf{y}, t) \to \mathbf{x}(\mathbf{y}, t)$ in probability as $T \to 0$. It follows that the PDF (6.1.2) becomes more and more concentrated around $\mathbf{x}(\mathbf{y}, t)$. Hence, it is sufficient to consider, for small δ,

$$A_T(K_1, K_2, \delta) = \int_{E(K_1, K_2, \delta)} \exp\left(-\frac{1}{T}F(\mathbf{x}, \mathbf{y}, t)\right) d\mathbf{x},$$
$$B_T(\delta) = \int_{\|\mathbf{x} - \mathbf{x}(\mathbf{y}, t)\|_\infty \leq \delta} \exp\left(-\frac{1}{T}F(\mathbf{x}, \mathbf{y}, t)\right) d\mathbf{x},$$

where $E(K_1, K_2, \delta) = E(K_1, K_2) \cap \{\mathbf{x} : \|\mathbf{x} - \mathbf{x}(\mathbf{y}, t)\|_\infty \leq \delta\}$ and $\|\mathbf{x}\|_\infty = \max(|x_i| : i = 1, \ldots, p)$. From Proposition 6.2.4 formula (6.2.2), we have

$$A_T(K_1, K_2, \delta) = \exp\left(-\frac{m(\mathbf{y}, t)}{T}\right)$$
$$\int_{E(K_1, K_2, \delta)} \exp\left(-\frac{1}{T}(\sum_{l \in I_0} |x_l|(1 - sgn(x_l)\xi_l(\mathbf{y}, t)) + \frac{\|\mathbf{A}(\mathbf{x} - \mathbf{x}(\mathbf{y}, t))\|^2}{2t})\right) d\mathbf{x}$$
$$= \exp\left(-\frac{m(\mathbf{y}, t)}{T}\right) \int_{E(K_1, K_2, \delta)}$$
$$\exp\left(-\frac{1}{T}(\sum_{l \in I_0 \setminus K_2} |x_l|(1 - sgn(x_l)\xi_l(\mathbf{y}, t)) + \frac{\|\mathbf{A}(\mathbf{x} - \mathbf{x}(\mathbf{y}, t))\|^2}{2t})\right) d\mathbf{x}.$$

Using the change of variables

$$\mathbf{u} = \frac{\mathbf{x}(I_0 \setminus K_2)}{T}, \quad \frac{\mathbf{x}(S \cup K_2) - \mathbf{x}(\mathbf{y}, t, S \cup K_2)}{\sqrt{T}} = \mathbf{v}, \qquad (6.2.4)$$

we get $A_T(K_1,K_2,\delta)=\exp\left(-\frac{m(\mathbf{y},t)}{T}\right)T^{\frac{|I_0\setminus K_2|+p}{2}}C_T(K_1,K_2)$, *where* $C_T(K_1,K_2)=$
$\int_{\tilde{E}_T(K_1,K_2,\delta)}$
$\exp\left(-(\sum_{i\in I_0\setminus K_2}|u_i|(1-sgn(u_i)\xi_i(\mathbf{y},t))+\frac{\|\sqrt{T}\sum_{i\in I_0\setminus K_2}u_i\mathbf{Ae}_i+\sum_{i\in(S\cup K_2)}v_i\mathbf{Ae}_i\|^2}{2t})\right)d\mathbf{u}d\mathbf{v}$,
and

$$\tilde{E}_T(K_1,K_2,\delta)=\{(u,v)\in\boldsymbol{R}^p:\mathbf{u}\in[-\frac{\delta}{T},\frac{\delta}{T}]^{|I_0\setminus K_2|},\mathbf{v}\in[-\frac{\delta}{\sqrt{T}},\frac{\delta}{\sqrt{T}}]^{|S\cup K_2|},$$
$$sgn(\mathbf{u}_{K_1})=-\xi_{K_1}(\mathbf{y},t),\ sgn(\mathbf{v}_{K_2})=\xi_{K_2}(\mathbf{y},t),$$
$$sgn(\sqrt{T}\mathbf{v}_S+\mathbf{x}(\mathbf{y},t,S))=sgn(\mathbf{x}(\mathbf{y},t,S))\}.$$

From the same calculation we can show that $B_T(\delta)=\sum_{K_1',K_2':\partial I_0=K_1'\cup K_2'}A_T(K_1',K_2',\delta)$. *We emphasize that the couple* $K_1'=\emptyset,K_2'=\partial I_0$ *is an element of the latter sum. Moreover, the quantity* $\frac{p+|I_0\setminus K_2|}{2}$ *is minimal at* $K_2=\partial I_0$. *From this we derive that*

$$\frac{A_T(K_1,K_2,\delta)}{B_T(\delta)}=\frac{T^{\frac{|I_0\setminus K_2|-|I_0\setminus\partial I_0|}{2}}C_T(K_1,K_2)}{C_T(\emptyset,\partial I_0)+\sum_{K_1',K_2'\neq\emptyset:\partial I_0=K_1'\cup K_2'}T^{\frac{|I_0\setminus K_2'|-|I_0\setminus\partial I_0|}{2}}C_T(K_1',K_2')}$$

converges to 0 as $T\to 0$, *because* $C_T(K_1',K_2')\to C_0(K_1',K_2')\neq 0$ *as* $T\to 0$ *for any partition* K_1',K_2' *of* ∂I_0.

6.3 Main results

Now we are ready to announce our main results.

Theorem 6.3.1. *Suppose that the matrix* $[\langle\mathbf{Ae}_i,\mathbf{Ae}_j\rangle,\quad i,j\in(S\cup\partial I_0)]$ *is invertible. Then the random vector* $\left(\frac{X_T(\mathbf{y},t,i)}{T}:\quad i\in(I_0\setminus\partial I_0)\right)$,
$\left(\frac{X_T(\mathbf{y},t,i)-x(\mathbf{y},t,i)}{\sqrt{T}}:\quad i\in(S\cup\partial I_0)\right)$ *converges in distribution for* $T\to 0$ *to the random vector*
$(X_i(\mathbf{y},t):\quad i\in(I_0\setminus\partial I_0)),(X_i(\mathbf{y},t):\quad i\in(S\cup\partial I_0))$ *having the PDF proportional to*

$$\prod_{i\in(I_0\setminus\partial I_0)}\exp\left(-|x_i|(1-sgn(x_i)\xi_i(\mathbf{y},t))\right)$$
$$\exp\left(-\frac{\|\sum_{i\in(S\cup\partial I_0)}x_i\mathbf{Ae}_i\|^2}{2t}\right)\prod_{i\in\partial I_0}\mathbf{1}_{[sgn(x_i)\xi_i(\mathbf{y},t))=1]}.$$

Preuve 6.3.2. *Let $I = I_0 \setminus \partial I_0$ and $J = S \cup \partial I_0$ and $\mathbf{a}, \mathbf{b} \in \boldsymbol{R}^p$. We want to prove that :*

$$\mathbf{P}\left(\mathbf{a}(I) \leq \frac{X_T(\mathbf{y},t,I)}{T} \leq \mathbf{b}(I), \mathbf{a}(J) \leq \frac{X_T(\mathbf{y},t,J) - \mathbf{x}(\mathbf{y},t,J)}{\sqrt{T}} \leq \mathbf{b}(J)\right)$$

converges to

$$\mathbf{P}\left(\mathbf{a}(I) \leq X(\mathbf{y},t,I) \leq \mathbf{b}(I), \mathbf{a}(J) \leq X(\mathbf{y},t,J) - \mathbf{x}(\mathbf{y},t,J) \leq \mathbf{b}(J)\right)$$

as $T \to 0$.

As we showed in Proposition 6.2.6, it is sufficient to consider, for small δ,

$$\begin{aligned}
&\mathbf{P}(\mathbf{a}(I) \leq \frac{X_T(\mathbf{y},t,I)}{T} \leq \mathbf{b}(I), \mathbf{a}(J) \leq \frac{X_T(\mathbf{y},t,J) - \mathbf{x}(\mathbf{y},t,J)}{\sqrt{T}} \leq \mathbf{b}(J),\\
&\|X_T(\mathbf{y},t) - \mathbf{x}(\mathbf{y},t)\|_\infty \leq \delta)\\
&= \sum_{K_1,K_2:\partial I_0 = K_1 \cup K_2} \mathbf{P}\left(\dots \mid E_T(K_1,K_2,\delta)\right) \mathbf{P}\left(E_T(K_1,K_2,\delta)\right),
\end{aligned}$$

where the events $E_T(K_1,K_2,\delta)$ are defined in Proposition 6.2.6. As we are interested in the limit as $T \to 0$ and thanks to Proposition 6.2.6 only the term $\mathbf{P}\left(\cdots \mid E_T(\emptyset,\partial I_0,\delta)\right)\mathbf{P}\left(E_T(\emptyset,\partial I_0,\delta)\right)$ is needed. More precisely we have only to study the term

$$\begin{aligned}
&\mathbf{P}\Big(\mathbf{a}(I) \leq \frac{X_T(\mathbf{y},t,I)}{T} \leq \mathbf{b}(I),\\
&\mathbf{a}(J) \leq \frac{X_T(\mathbf{y},t,J) - \mathbf{x}(\mathbf{y},t,J)}{\sqrt{T}} \leq \mathbf{b}(J) \mid E_T(\emptyset,\partial I_0,\delta)\Big) = \frac{A_T(\delta)}{B_T(\delta)}
\end{aligned}$$

where

$$\begin{aligned}
A_T(\delta) &= \int_{T\mathbf{a}(I)}^{T\mathbf{b}(I)} \int_{\sqrt{T}\mathbf{a}(J)+\mathbf{x}(\mathbf{y},t,J)}^{\sqrt{T}\mathbf{b}(J)+\mathbf{x}(\mathbf{y},t,J)} \exp(-\frac{1}{T}F(\mathbf{x},\mathbf{y},t))\mathbf{1}_{E(\emptyset,\partial I_0,\delta)}(\mathbf{x})d\mathbf{x}\\
B_T(\delta) &= \int \exp(-\frac{1}{T}F(\mathbf{x},\mathbf{y},t))\mathbf{1}_{E(\emptyset,\partial I_0,\delta)}d\mathbf{x}.
\end{aligned}$$

From Proposition 6.2.4 we have

$$\begin{aligned}
&A_T(\delta) = \exp\left(-\frac{m(\mathbf{y},t)}{T}\right) \int_{T\mathbf{a}(I)}^{T\mathbf{b}(I)} \int_{\sqrt{T}\mathbf{a}(J)+\mathbf{x}(\mathbf{y},t,J)}^{\sqrt{T}\mathbf{b}(J)+\mathbf{x}(\mathbf{y},t,J)}\\
&\exp\left(-\frac{1}{T}(\sum_{i\in I}|x_i|(1 - sgn(x_i)\xi_i(\mathbf{y},t)) + \frac{\|\mathbf{A}(\mathbf{x} - \mathbf{x}(\mathbf{y},t))\|^2}{2t})\right)\mathbf{1}_{E(\emptyset,\partial I_0,\delta)}(\mathbf{x})d\mathbf{x}.
\end{aligned}$$

Using the change of variables

$$\mathbf{u} = \frac{\mathbf{x}(I)}{T}, \quad \frac{\mathbf{x}(J) - \mathbf{x}(\mathbf{y}, t, J)}{\sqrt{T}} = \mathbf{v}, \tag{6.3.1}$$

we get

$$A_T(\delta) = \exp\left(-\frac{m(\mathbf{y},t)}{T}\right) T^{\frac{|I|+p}{2}} \int_{\mathbf{a}(I)}^{\mathbf{b}(I)} \int_{\mathbf{a}(J)}^{\mathbf{b}(J)} d\mathbf{u}d\mathbf{v}$$
$$\exp\left(-(\sum_{i\in I} |u_i|(1 - sgn(u_i)\xi_i(\mathbf{y},t)) + \frac{\|\sqrt{T}\sum_{i\in I} u_i\mathbf{A}\mathbf{e}_i + \sum_{i\in J} v_i\mathbf{A}\mathbf{e}_i\|^2}{2t})\right)$$
$$\mathbf{1}_{\tilde{E}_T(\emptyset,\partial I_0,\delta)}(\mathbf{u},\mathbf{v}).$$

Now we are going to study $T^{-\frac{|I|+p}{2}} B_T \exp\left(\frac{m(\mathbf{y},t)}{T}\right)$*. From the change of variables formula (6.3.1), we have*

$$\lim_{T\to 0} T^{-\frac{|I|+p}{2}} B_T(\delta) \exp\left(\frac{m(\mathbf{y},t)}{T}\right) = \lim_{T\to 0} T^{-\frac{|I|+p}{2}} \int_{-\delta\le\mathbf{x}\le\delta}$$
$$\exp\left(-\frac{1}{T}(\sum_{i\in I} |x_i|(1 - sgn(x_i)\xi_i(\mathbf{y},t)) + \frac{\|\mathbf{A}(\mathbf{x} - \mathbf{x}(\mathbf{y},t))\|^2}{2t})\right) \mathbf{1}_{E(\emptyset,\partial I_0,\delta)}(\mathbf{x})d\mathbf{x}$$
$$= \int \exp\left(-(\sum_{i\in I} |u_i|(1 - sgn(u_i)\xi_i(\mathbf{y},t)) + \frac{\|\sum_{i\in J} v_i\mathbf{A}\mathbf{e}_i\|^2}{2t})\right)$$
$$\prod_{i\in\partial I_0} \mathbf{1}_{[sgn(v_i)\xi_i(\mathbf{y},t)=1]} d\mathbf{u}d\mathbf{v},$$

which achieves the proof.

Theorem 6.3.1 tells us that the random vectors $(X_i(\mathbf{y},t), i \in I_0 \setminus \partial I_0)$, $(X_i(\mathbf{y},t), i \in S \cup \partial I_0)$ are independent. Moreover, the components $(X_i(\mathbf{y},t), i \in I_0 \setminus \partial I_0)$ are mutually independent. We can show, for $i \in I_0 \setminus \partial I_0$, that the first moment of $X_i(\mathbf{y},t)$ is equal to

$$m_1(\xi_i(\mathbf{y},t)) := \frac{2\xi_i(\mathbf{y},t))}{1 - |\xi_i(\mathbf{y},t)|^2}.$$

If ∂I_0 is empty, then $(X_i(\mathbf{y},t), i \in S)$ is centred and Gaussian with the covariance matrix $t[\langle \mathbf{A}\mathbf{e}_i, \mathbf{A}\mathbf{e}_j\rangle : i,j \in S]^{-1}$. If ∂I_0 is not empty, then the distribution of $(X_i(\mathbf{y},t), i \in S \cup \partial I_0)$ is equal to the distribution of a centred Gaussian vector $(G_i, i \in S \cup \partial I_0)$, with the covariance matrix $t[\langle \mathbf{A}\mathbf{e}_i, \mathbf{A}\mathbf{e}_j\rangle : i,j \in S \cup \partial I_0]^{-1}$, conditional to $(\xi_i(\mathbf{y},t)sgn(G_i) = 1, i \in \partial I_0)$. Now we are going to show, for small T, that the component $X_T(\mathbf{y},t,i)$ with $i \in I_0 \setminus \partial I_0$

behaves like the one dimensional random variable $X_T(t\xi_i(\mathbf{y},t),t)$ drawn from the PDF proportional to $\exp\left(-\frac{1}{T}(|x|+\frac{(x-t\xi_i(\mathbf{y},t))^2}{2t})\right)$. We recall that in the one dimensional case $n=p=1$ with the data y and $A=1$, the objective function $F(x,y,t)=|x|+\frac{(x-y)^2}{2t}$. In this case LASSO $x(y,t)=sgn(y)(|y|-t)\mathbf{1}_{[|y|>t]}$. Hence for $|y|<t$, $I_0=\{1\}$, $\partial I_0=S=\emptyset$, for $|y|=t$, $I_0=S=\emptyset$, $\partial I_0=\{1\}$, and finally for $|y|>t$, $I_0=\partial I_0=\emptyset$, $S=\{1\}$. In the one dimensional case Proposition 6.2.6 and Theorem 6.3.1 can be summarized as following.

Theorem 6.3.3. *1) If $|y|<t$, then $\frac{X_T(y,t)}{T}\to X(y,t)$ in distribution as $T\to 0$, where $X(y,t)$ is the random variable having the PDF*

$$x\to\frac{1-\frac{y^2}{t^2}}{2}\exp\left(-|x|(1-sgn(x)\frac{y}{t})\right).$$

2) Knowing that $[X_T(t,t)<0]$, $\frac{X_T(t,t)}{T}\to-\mathcal{E}(2)$ in distribution as $T\to 0$, where $\mathcal{E}(2)$ is the random variable having the exponential distribution with the parameter 2, i.e. the PDF of $\mathcal{E}(2)$ is equal to $2\exp(-2x)1_{[x>0]}$. More precisely we have, for any $a<b<0$, $\mathbf{P}(a<\frac{X_T(t,t)}{T}<b\,|\,X_T(t,t)<0)\to\mathbf{P}(a<-\mathcal{E}(2)<b)$ as $T\to 0$. Moreover $\mathbf{P}(X_T(t,t)<0)\to 0$ as $T\to 0$.
3) Knowing that $[X_T(t,t)>0]$, the random variable $\frac{X_T(t,t)}{\sqrt{T}}\to|N(0,t)|$, where $N(0,t)$ is the standard Gaussian with the variance t.
4) We have for $|y|>t$ that $\frac{X_T(y,t)-x(y,t)}{\sqrt{T}}\to N(0,t)$ as $T\to 0$.

Let us show how to sample from the limit $X(y,t)$ given in Theorem 6.3.3. If $|y|<t$, then the density of $X(y,t)$ is a mixture of exponential probability distributions i.e. is equal to $\frac{1-\frac{y}{t}}{2}f_{X_-(y,t)}(x)+\frac{1+\frac{y}{t}}{2}f_{X_+(y,t)}(x)$, where $X_-(y,t),X_+(y,t)$ are independent variables having respectively the exponential distribution $-\mathcal{E}(1+\frac{y}{t})$, $\mathcal{E}(1-\frac{y}{t})$. Hence, $X(y,t)$ has the same PDF as $X_{b(\frac{1+\frac{y}{t}}{2})}(y,t)$, where $X_-(y,t)$, $X_+(y,t)$, $b(\frac{1+\frac{y}{t}}{2}))$ are independent with the respective PDF

$$-\mathcal{E}(1+\frac{y}{t}),\mathcal{E}(1-\frac{y}{t}),$$
$$\mathbf{P}(b(\frac{1+\frac{y}{t}}{2})=-)=\frac{1-\frac{y}{t}}{2},\mathbf{P}(b(\frac{1+\frac{y}{t}}{2})=+)=\frac{1+\frac{y}{t}}{2}.$$

In Figure 1, we plot the probability density function of $X(y,t)$.

We finish this section by showing how to estimate LASSO using Theorem 6.3.1. We sample, for small T, $X_T(\mathbf{y},t)$ using one algorithm among MCMC

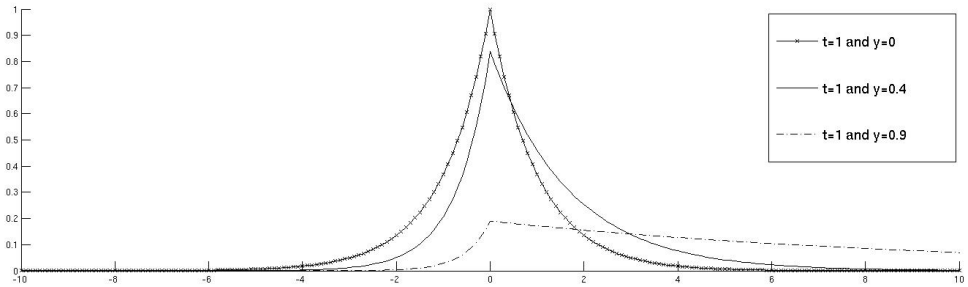

FIGURE 6.1 – The density of $X(y,t)$.

algorithms and we get a sequence $(\theta^{(n)} : \quad n = 1, \ldots, N)$. There is a great variety of different MCMC methods see e.g. [15] and the references herein. In our work, due to lack of space and time, we only use the random-walk Metropolis-Hastings algorithm [10], [13]. We approximate $\mathbb{E}(X_T(\mathbf{y}, t))$ by $\bar{\theta} := \frac{1}{N}\sum_{n=1}^{N}\theta^{(n)}$ and we propose an estimator LASSOMH (say) of LASSO based on Theorem 3.1. The construction of LASSOMH works as following. We estimate the vector $\xi(\mathbf{y}, t)$ by $\bar{\xi} := \frac{A^*(\mathbf{y}-\mathbf{A}\bar{\theta})}{t}$. We fix $u \in (0,1)$, $c > 0$ and estimate the sets $I_0 \setminus \partial I_0$, ∂I_0 respectively by $(I_0 \setminus \partial I_0)(u) := \{i : |\bar{\theta}_i| \leq \frac{2Tu}{1-u^2}\}$ and $\partial I_0(u,c) = \{i : \frac{2Tu}{1-u^2} < |\bar{\theta}_i| \leq c\sqrt{T}\}$. Finally LASSOMH has the component $\bar{\theta}_i$ for $i \notin I_0(u,c) := (I_0 \setminus \partial I_0)(u) \cup \partial I_0(u,c)$ and $\bar{\theta}_i = 0$ if not. As an illustration we consider the case $p = 10$, $n = 7$, $t = 1$, and the entries of the matrix $\mathbf{A}$ are independent Bernoulli random variables with values $\pm\frac{1}{\sqrt{n}}$, and $\mathbf{w} \sim \mathcal{N}(0,t)$. We simulate the vector $\mathbf{x}$ from the PDF $2^{-p}\exp(-\|\mathbf{x}\|_1)$. We get the data $\mathbf{y} := \mathbf{A}\mathbf{x} + \mathbf{w}$ from a realization of $\mathbf{A}$ and $\mathbf{w}$. Using FISTA algorithm we get the estimators $LASSO(FISTA)$, $\xi(FISTA)$, $I_0(FISTA)$ and $\partial I_0(FISTA)$ respectively of LASSO, $\xi(\mathbf{y}, t)$, I_0 and ∂I_0. We display our estimators and FISTA estimators in Table 1.

We showed how to use the random-walk Metropolis-Hastings algorithm with a small temperature in order to calculate LASSO. In the sequel we suppose that LASSO and the sets I_0, ∂I_0 and S are known, and we are going to define new criteria for choosing the proposal distribution and the temperature.

i	1	2	3	4	5	6	7	8	9	10
$\bar{\theta}$	-1.8840	0.0162	-0.3298	0.3997	-0.0033	0.6388	-0.1617	0.8318	-0.1734	0.0384
$\bar{\xi}$	-0.8203	0.3802	-0.4169	0.7824	-0.3398	1.0086	-0.0834	1.0086	0.3398	-0.6341
LASSOMH	-1.8840	0	0	0	0	0.6388	0	0.8318	0	0
LASSOFISTA	-1.9409	0	0	0	0	1.0592	0	0.9136	0	0
$\xi(FISTA)$	-1	0.4750	-0.9690	0.8438	-0.5335	1	-0.5804	1	0.5335	-0.6427

TABLE 6.1 – $t = 1$, $T = 0.08$, $u = 0.7863$, $c = 1.4275$, $(I_0 \setminus \partial I_0)(u) = \{2, 3, 5, 7, 9, 10\}$, $\partial I_0(u, c) = \{4\}$. Observe that $I_0(u, c) = I_0(FISTA)$ and as $|\xi_3(FISTA)| \approx 1$, $|\xi_4(FISTA)| \approx 1$ we propose $\{3, 4\} := \partial I_0(FISTA)$. It follows that $\partial I_0(u, c) \subset \partial I_0(FISTA)$.

6.4 Choosing the proposal distribution and the temperature in the random-walk Metropolis-Hastings algorithm

6.4.1 Choosing the proposal distribution in the random-walk Metropolis-Hastings algorithm

We want, for small T, to sample from $X_T(\mathbf{y}, t)$ using the random-walk Metropolis-Hastings algorithm with a family of proposal distributions. If the variance of the proposal is too small or too large, the random-walk Metropolis-Hastings algorithm will converge slowly. A number of works have suggested informal guidelines for choosing the proposal by monitoring variances and accept/reject ratios (also called Probability of acceptance) see e.g. [3], [4]. Gelman et al. [7], see also [12] example 5.3 chapter 5, propose to adjust the proposal such that the acceptance rate is around $\frac{1}{2}$ for one or two dimensional target distributions, and around $\frac{1}{4}$ for larger dimensions. Haario et al. [9] suggested a method called Adaptive Proposal (AP). AP algorithm may be viewed as the random-walk Metropolis-Hastings where the proposal distribution depends on time. In the sequel we propose new criterion. The idea is to fix, for some integer $d \geq 1$, a test function $\varphi : \boldsymbol{R}^p \to \boldsymbol{R}^d$, and apply Theorem 6.3.1 for small T as following : $\mathbb{E}[\varphi(X_T(\mathbf{y}, t) - \mathbf{x}(\mathbf{y}, t))] \approx \mathbb{E}[\varphi(TX_{I_0 \setminus \partial I_0}(\mathbf{y}, t), \sqrt{T}X_{S \cup \partial I_0}(\mathbf{y}, t))]$. Here $X_I(\mathbf{y}, t) := (X_i(\mathbf{y}, t), i \in I)$ for any subset $I \subset \{1, \ldots, p\}$. Knowing the test function φ we calculate the expectation $\mathbb{E}[\varphi(TX_{I_0 \setminus \partial I_0}(\mathbf{y}, t), \sqrt{T}X_{S \cup \partial I_0}(\mathbf{y}, t))]$. The best proposal based on the latter approximation for sampling $X_T(\mathbf{y}, t)$ will produce a sequence $(\theta^{(n)} : \quad n = 1, \ldots, N)$ such that $\frac{1}{N}\sum_{n=1}^{N} \varphi(\theta^{(n)} - \mathbf{x}(\mathbf{y}, t))$ is the nearest to

$\mathbb{E}[\varphi(TX_{I_0\setminus\partial I_0}(\mathbf{y},t),\sqrt{T}X_{S\cup\partial I_0}(\mathbf{y},t))]$. More precisely, for each proposal q, we consider M chains with the size N, i.e. $(\theta_i^{(n)}: \quad i=1,\dots,M, n=1,\dots,N)$, and calculate the objective function

$$f(q,\varphi)=\frac{1}{M}\sum_{i=1}^{M}\|\frac{1}{N}\sum_{n=1}^{N}\varphi(\theta_i^{(n)}-\mathbf{x}(\mathbf{y},t))-\mathbb{E}[\varphi(TX_{I_0\setminus\partial I_0}(\mathbf{y},t),\sqrt{T}X_{S\cup\partial I_0}(\mathbf{y},t))]\|.$$

We say that the proposal q^* is the best among a finite family F of proposal distributions if q^* is the minimizer of $q\in F\to f(q,\varphi)$. As a first illustration we consider the case $p=10$, $n=7$, already used in Table 1. Using FISTA algorithm we get LASSO $\mathbf{x}(\mathbf{y},t)$ and $\xi(\mathbf{y},t)$ with $S=\{1,6,8\}$, $I_0\setminus\partial I_0=\{2,5,7,9,10\}$, $\partial I_0=\{3,4\}$, and $det([\langle\mathbf{Ae}_i,\mathbf{Ae}_j\rangle: i,j\in S\cup\partial I_0]^{-1})=0.7463$. A comparison between the acceptance rate and our criterion $f_1(q):=f(q,\varphi)$ with $\varphi(\mathbf{x})=\mathbf{x}_{I_0\setminus\partial I_0}$ and the proposal distribution $q=\mathcal{N}(0,\sigma^2\mathbf{I}_{p\times p})$ is displayed in Table 2. In our second illustration we will consider in Section

Proposal	$f_1(q)$	The acceptance rate
$\sigma^2=0.09$	0.2445	0.2234
$\sigma^2=1$	0.9846	0.0024
$\sigma^2=6$	732.84	0.9620

TABLE 6.2 – $t=1$, $T=0.1$, $M=500$, $N=5000$. The variance $\sigma^2=0.09$ wins according to the two criteria.

6 the one dimensional case, $n=p=1$ and $A=1$. In this case LASSO $x(y,t)$ is explicit for all y and t. We will discuss in details this simple case by considering the following test functions : $\varphi_1(x)=x$, $\varphi_2(x)=x^2$ and $\varphi_3(x)=(x,x^2)$.

6.4.2 Choosing the temperature in the random-walk Metropolis-Hastings algorithm

We propose, for $i=1,\dots,p$, $X_T(\mathbf{y},t,i)$ as an estimator of LASSO component $\mathbf{x}(\mathbf{y},t,i)$. We approximate respectively the bias $\mathbb{E}\left(X_T(\mathbf{y},t,i)-\mathbf{x}(\mathbf{y},t,i)\right)$ and the mean square error $\mathbb{E}\left(|X_T(\mathbf{y},t,i)-\mathbf{x}(\mathbf{y},t,i)|^2\right)$ respectively by $T\mathbb{E}\left(X(\mathbf{y},t,i)\right)$ and $T^2\mathbb{E}\left(X^2(\mathbf{y},t,i)\right)$ for $i\in I_0\setminus\partial I_0$, $\sqrt{T}\mathbb{E}\left(X(\mathbf{y},t,i)\right)$ and $T\mathbb{E}\left(X^2(\mathbf{y},t,i)\right)$ for $i\in S\cup\partial I_0$. The idea for choosing the temperature is to fix an upper bound b_{ap} of the approximate bias or an upper bound MSE_{ap} of the approximate mean square errors and then derive the temperature as a function of b_{ap} or MSE_{ap}. In Section 7 we will illustrate our criteria in the one dimensional case, $n=p=1$ and $A=1$.

6.5 Choice of the proposal distribution in the one dimensional case

We distinguish three cases.
1) We have for $y \in [0,t)$ that $\mathbb{E}[X(y,t)] = \frac{2\frac{y}{t}}{(1-\frac{y^2}{t^2})} := m_1(\frac{y}{t})$ and

$$\mathbb{E}[X^2(y,t)] = \{\frac{1-\frac{y}{t}}{(1+\frac{y}{t})^2} + \frac{1+\frac{y}{t}}{(1-\frac{y}{t})^2}\} := m_2(\frac{y}{t}).$$

a) If the test function $\varphi = \varphi_1$, then for each proposal q we calculate the objective function

$$f(q,\varphi_1) := \frac{1}{M}\sum_{i=1}^{M} |\frac{1}{N}\sum_{n=1}^{N} \theta_i^{(n)} - Tm_1(\frac{y}{t})| := f_1(q)$$

We say that the proposal q^* is the best among a finite family F of proposal distributions if q^* is the minimizer of $q \in F \to f_1(q)$.
b) If the test function $\varphi = \varphi_2$, then for each proposal q we calculate the objective function

$$f(q,\varphi_2) := \frac{1}{M}\sum_{i=1}^{M} |\frac{1}{N}\sum_{n=1}^{N} (\theta_i^{(n)})^2 - T^2 m_2(\frac{y}{t})| := f_2(q)$$

We say that the proposal q^* is the best among a finite family F of proposal distributions if q^* is the minimizer of $q \in F \to f_2(q)$.
c) If the test function $\varphi = \varphi_3$, then the best proposal is the minimizer of $q \in F \to \sqrt{f_1(q) + f_2(q)}$.

In order to illustrate these results we consider $M = 600$ chains with the size $N = 5000$, with the proposal distribution $\mathcal{N}(0,\sigma^2)$ and different values of σ^2.

Proposal	$f_1(q)$	$f_2(q)$	$\sqrt{f_1(q)+f_2(q)}$	The acceptance rate
$\sigma^2 = 0.09$	0.0538	0.0550	0.329	0.7556
$\sigma^2 = 1$	0.0360	0.0898	0.354	0.1672
$\sigma^2 = 6$	0.0368	0.0918	0.358	0.0284

TABLE 6.3 – $t = 1, y = 0.5$, $T = 0.1$, $M = 600$, $N = 5000$; the variance $\sigma^2 = 1$ wins according to the criterion f_1 and $\sigma^2 = 0.09$ wins according to the criteria f_2, $\sqrt{f_1 + f_2}$ and the acceptance rate.

2) We have for $y = t$ that $\mathbb{E}[X(t,t)] = \sqrt{\frac{2t}{\pi}}$, and $\mathbb{E}[X^2(t,t)] = t$.

a) If the test function is $\varphi = \varphi_1$, then for each proposal q we calculate the objective function

$$f(q,\varphi_1) := \frac{1}{M}\sum_{i=1}^{M} |\frac{1}{N}\sum_{n=1}^{N} \theta_i^{(n)} - \sqrt{\frac{2Tt}{\pi}}| := f_1(q).$$

We say that the proposal q^* is the best among a finite family F of proposal distributions if q^* is the minimizer of $q \in F \to f_1(q)$.

b) If the test function is $\varphi = \varphi_2$, then for each proposal q we calculate the objective function

$$f(q,\varphi_2) := \frac{1}{M}\sum_{i=1}^{M} |\frac{1}{N}\sum_{n=1}^{N} (\theta_i^{(n)})^2 - Tt| := f_2(q).$$

We say that the proposal q^* is the best among a finite family F of proposal distributions if q^* is the minimizer of $q \in F \to f_2(q)$.

c) If the test function is $\varphi = \varphi_3$, then the best proposal is the minimizer of $q \in F \to \sqrt{f_1(q) + f_2(q)}$.

Proposal	$f_1(q)$	$f_2(q)$	$\sqrt{f_1(q)+f_2(q)}$	The acceptance rate
$\sigma^2 = 0.09$	0.17569	0.01957	0.44189	0.8484
$\sigma^2 = 1$	0.17391	0.01994	0.44029	0.2310
$\sigma^2 = 6$	0.17471	0.02069	0.44205	0.0362

TABLE 6.4 – $t = y = 1$, $T = 0.1$, $M = 600$, $N = 5000$; $\sigma^2 = 1$ wins according to the criteria f_1, $\sqrt{f_1 + f_2}$, and The acceptance rate, but $\sigma^2 = 0.09$ wins according to the criterion f_2.

3) We have for $y > t$ that $\mathbb{E}[X(y,t)] = 0$ and $\mathbb{E}[X^2(y,t)] = t$.

a) If the test function φ_1, then for each proposal q we calculate the objective function

$$f(q,\varphi_1) := \frac{1}{M}\sum_{i=1}^{M} |\frac{1}{N}\sum_{n=1}^{N} \theta_i^{(n)} - (y-t)| := f_1(q).$$

We say that the proposal q^* is the best among a finite family F of proposal distributions if q^* is the minimizer of $q \in F \to f_1(q)$.

b)If the test function φ_2, then for each proposal q we calculate the objective function

$$f(q,\varphi_2) := \frac{1}{M}\sum_{i=1}^{M} |\frac{1}{N}\sum_{n=1}^{N} (\theta_i^{(n)})^2 - (y-t)^2 - Tt| := f_2(q).$$

We say that the proposal q^* is the best among a finite family F of proposal distributions if q^* is the minimizer of $q \in F \to f_2(q)$.
c) If the test function is $\varphi = \varphi_3$, then the best proposal is the minimizer of $q \in F \to \sqrt{f_1(q) + f_2(q)}$.

Proposal	$f_1(q)$	$f_2(q)$	$\sqrt{f_1(q) + f_2(q)}$	The acceptance rate
$\sigma^2 = 0.09$	0.0270	0.0135	0.2011	0.910
$\sigma^2 = 1$	0.0078	0.0040	0.1087	0.385
$\sigma^2 = 6$	0.0174	0.0088	0.1614	0.0716

TABLE 6.5 – $t = 1, y = 3,$, $T = 0.1$, $M = 600$, $N = 5000$; $\sigma^2 = 1$ wins according to the four criteria.

6.6 Choice of the temperature in Metropolis-Hastings algorithm

We distinguish three cases.
1) The case $y \in [0, t)$. The idea is to approximate the bias $\mathbb{E}[X_T(y,t) - x(y,t)] = \mathbb{E}[X_T(y,t)]$ by $T\mathbb{E}[X(y,t)]$ and the mean square $\mathbb{E}[(X_T(y,t) - x(y,t))^2] = \mathbb{E}[(X_T(y,t))^2]$ by $T^2\mathbb{E}[X^2(y,t)]$.
a) Controlling the approximate bias : Fixing the approximate bias

$$T\mathbf{E}[X(y,t)] = T\frac{2\frac{y}{t}}{(1-\frac{y^2}{t^2})} = Tm_1(\frac{y}{t}) := b_{ap},$$

we get, for $y \neq 0$, the temperature $T(b_{ap}, \frac{y}{t}) := \frac{b_{ap}}{m_1(\frac{y}{t})}$. Observe that our criterion is not defined for $y = 0$. The Figure 2 (a) shows, for fixed b_{ap}, that $u \in (0,1) \to T(b_{ap}, u)$ is decreasing.
b) Controlling the approximate mean square error : Fixing the approximate mean square error

$$T^2\mathbb{E}[X^2(y,t)] = T^2\{\frac{1-\frac{y}{t}}{(1+\frac{y}{t})^2} + \frac{1+\frac{y}{t}}{(1-\frac{y}{t})^2}\} = T^2 m_2(\frac{y}{t}) := MSE_{ap},$$

we get the temperature $T(MSE_{ap}, \frac{y}{t}) = \sqrt{\frac{MSE_{ap}}{m_2(\frac{y}{t})}}$. Observe, contrary to criterion a), that the temperature $T(MSE_{ap}, 0)$ is defined for $y = 0$. Moreover, if the couple (b_{ap}, MSE_{ap}) satisfies the constraint $\frac{b_{ap}^2}{m_1^2(\frac{y}{t})} = \frac{MSE_{ap}}{m_2(\frac{y}{t})}$,

then $T(b_{ap}, \frac{y}{t}) = T(MSE_{ap}, \frac{y}{t})$. The Figure 2 (b) shows, for fixed MSE_{ap}, that $u \in (0,1) \to T(MSE_{ap}, u)$ is also decreasing.

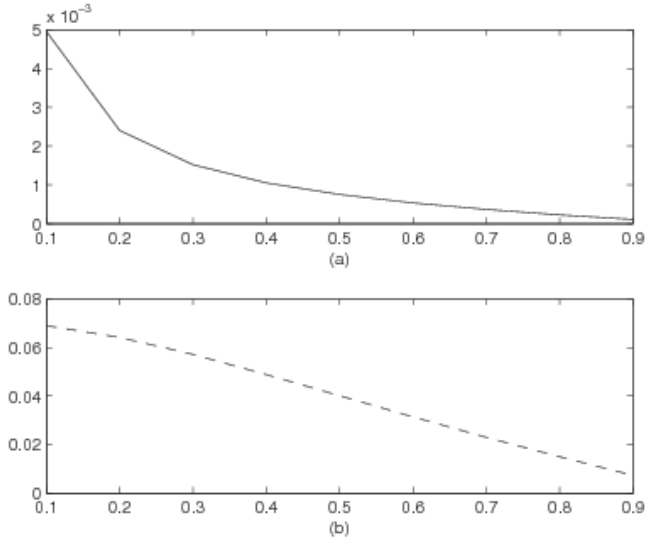

FIGURE 6.2 – (a) $b_{ap} = 0.001$, $u \in (0,1) \to T(b_{ap}, u)$ and (b) $MSE_{ap} = 0.01$, $u \in (0,1) \to T(MSE_{ap}, u)$.

2) The case $y = t$.
a) Controlling the approximate bias : Fixing the approximate the bias

$$\sqrt{T}\mathbb{E}[|\mathcal{N}(0,t)|] = \sqrt{\frac{2Tt}{\pi}} := b_{ap},$$

we get the temperature $T(b_{ap}, t) = \dfrac{\pi b_{ap}^2}{2t}$.
b) Controlling the approximate mean square error : Fixing the approximate mean square error

$$T\mathbb{E}[N^2(0,t)] = Tt := MSE_{ap},$$

we get the temperature $T(MSE_{ap}, t) = \dfrac{MSE_{ap}}{t}$.
3) The case $y > t$. Here the approximate bias $\sqrt{T}\mathbb{E}[N(0,t)] = 0$, and we have criterion based on the approximate mean square error, i.e.

$$T\mathbb{E}[\mathcal{N}^2(0,t)] = Tt := MSE_{ap}.$$

We get the temperature $T(MSE_{ap}, y) = \dfrac{MSE_{ap}}{t}$.

6.7 Criterion for convergence of Metropolis Hastings algorithm

Metropolis-Hastings algorithm produces a Markov chain (θ^n_{MH}) such that for any suitable measurable function h

$$\mathbb{E}[h(X_T(y,t))] = \lim_{N\to+\infty} \frac{\sum_{n=0}^{N} h(\theta^n)}{N}.$$

We address the problem of the convergence of the series $\frac{\sum_{n=0}^{N} h(\theta^n)}{N}$ in the cases $h(x) = x$, $h(x) = x^2$ and $y \in (0,t)$. We fix the approximate bias b_{ap}. We run Metropolis-Hastings algorithm with the temperature $T(b_{ap}, \frac{y}{t})$ (see Section 6 paragraph 1) a)) and we calculate the sums $b_{MH} := \frac{1}{N}\sum_{n=0}^{N} \theta^n$, and $MSE_{MH} := \frac{1}{N}\sum_{n=0}^{N} (\theta^n)^2$ for different sizes N. We decide that the chain is convergent at the iteration N if $b_{MH} \approx b_{ap}$ and $MSE_{MH} \approx MSE(b_{ap})$ where $MSE(b_{ap}) := \frac{b_{ap}^2 m_2(\frac{y}{t})}{m_1^2(\frac{y}{t})}$. Observe that the latter constraint guarantees the equality $T(b_{ap}, \frac{y}{t}) = T(MSE(b_{ap}), \frac{y}{t})$. In The table 6 we fix $b_{ap} = 0.01$. We vary N and calculate the values of b_{MH} and MSE_{MH}. We show for $N = 80000$, that $b_{MH} \approx b_{ap}$ and $MSE_{MH} \approx MSE(b_{ap}) = 0.00035$.

N_{MH}	4000	5000	8000	10000	30000	40000	80000
b_{MH}	0.0067	0.0074	0.0084	0.0085	0.0088	0.0091	0.0090
MSE_{MH}	0.0017	0.0014	0.0010	8.7017e-04	4.8341e-04	4.4582e-04	3.6576e-04

TABLE 6.6 – b_{MH} and MSE_{MH} values : $y = 0.5$, $t = 1$, $b_{ap} = 0.01$, $MSE(b_{ap}) = 3.5e-04$, $T(b_{ap}, \frac{y}{t}) = 0.0075$, $Proposal = \mathcal{N}(0, 0.009)$.

6.8 Choice of geometric tempering in simulated annealing algorithm

We want, for $y \in [0,t)$, to estimate LASSO $x(y,t)$ using simulated annealing's algorithm. A popular choice of the sequence of temperature $(T_n : n = 1,2,\ldots)$ in simulated annealing's algorithm is geometric tempering i.e. $T_n := \frac{T_0}{q^n}$ for some $q > 1$ and for arbitrary initial temperature T_0 see e.g.[12]. Observe, for $y \in [0,t)$, that the temperature T_n coincides with our temperature $T(b_{ap}(n), \frac{y}{t})$ with the approximate bias $b_{ap}(n) = \frac{1}{q^n}$

and the initial temperature $T_0 = \dfrac{1}{m_1(\frac{y}{t})}$. The temperature T_n also coincides with our temperature $T(MSE_{ap}(n), \frac{y}{t})$ with $MSE_{ap}(n) = \frac{1}{q^{2n}}$ and the initial temperature $T_0 = \frac{1}{\sqrt{m_2(\frac{y}{t})}}$. We propose to compare simulated annealing's algorithms using three initial temperatures $T_0 = 1$, $T_0 = m_1(\frac{y}{t})$ and $T_0 = \frac{1}{\sqrt{m_2(\frac{y}{t})}}$ and the same $q > 1$.

We consider, for each initial temperature T_0, $M = 500$ simulated annealing's algorithm with $q = 1.001$. We get a sequence $(\theta^n_{SA}(m) : n, m)$. We propose $b_{SA}(N) := \frac{1}{M}\sum_{m=1}^{M} \theta^N_{SA}(m)$ as an estimator of LASSO $x(y,t) = 0$. We select the iteration N_{SA} and the corresponding temperature T_{SA} giving the best estimate of LASSO. In figure 3 we plot $N \to b_{SA}(N) := \frac{1}{M}\sum_{m=1}^{M} \theta^N_{SA}(m)$ and we remark that for the three temperatures the best iteration N_{SA} is around 7000.

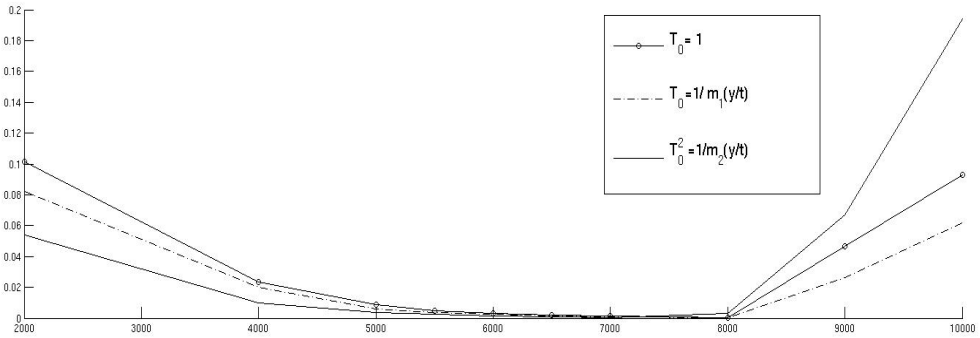

FIGURE 6.3 – $N \to b_{SA}(N)$ with different values of T_0 .

6.9 Conclusion

In this paper we treated LASSO using Gibbs measures. We showed that the scaling of the Gibbs measures as the temperature goes to zero depends on the support and the null components of LASSO. We obtained as a by-product a link between these scaling and the geometric tempering in simulated annealing algorithm. Our results can be easily extended to the analysis sparsity problem i.e. the minimization of the objective function $\|\mathbf{Dx}\|_1 + \frac{\|\mathbf{Ax}-\mathbf{y}\|^2}{2t}$ with $\mathbf{D} \neq \mathbf{I}$.

Bibliographie

[1] K.B. Athreya, C.-R. Hwang, Gibbs measures asymptotics, Sankhya A. Vol.72 Part. 1 191–207 (2010).

[2] A. Beck, M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problem. SIAM J. Imaging Sci. 183–202 (2009).

[3] J. Besag, P.J. Green, Spatial statistics and Bayesian computation, J. R. Stat. Soc. Ser. B Stat. Methodol. 55 25–38 (1993).

[4] J. Besag, P. J. Green, D. Higton, K. Mengersen, Bayesian computation and stochastic systems, Statist. Sci. 10 3–66 (1995).

[5] S. Chen, D.L. Donoho, M. Saunders, Atomic decomposition by basis pursuit, SIAM J. Sci. Comput. Vol. 20, no. 1 33–61 (1998).

[6] B. Efron, T. Hastie, I. Johnstone, R. Tibshirani, Least angle regression, Ann. Statist. vol. 37 407–499 (2004).

[7] A. Gelman, W.R. Gilks, G.O. Roberts : Weak convergence and optimal scaling of random walk Metropolis algorithms, Ann. Appl. Probab. vol. 7 110–120 (1997).

[8] M. Grasmair, M. Haltmeier, O. Scherzer, Necessary and sufficient conditions for linear convergence of l^1-regularization, Comm. Pure Appl. Math. vol. 64 no. 2 161–182 (2011).

[9] H. Haario, E. Saksman, J. Tamminen, Adaptive proposal distribution for random walk Metropolis algorithm, Comput. Statist. vol. 14 375–395 (1999).

[10] W.K. Hastings, Monte Carlo sampling methods using Markov chains and their application, Biometrika vol. 57 97–109 (1970).

[11] C.-R. Hwang, Laplace's method revised, weak convergence of probability measures, Ann. Probab. vol. 8 1177–1182 (1980).

[12] A. M. Johansen, L. Evers : Monte-Carlo Methods, Lecture Notes, University of Bristol, 1–128 (2007).

[13] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller, Equations of state calculation by fast computing machines, J. Chem. Phys. vol. 21 1087–1092 (1953).

[14] N. Parikh, S. Boyed, Proximal algorithms, Foundation and Trends in Optimization vol. 1 no. 3 123–231 (2003).

[15] M. Pereyra, Proximal Markov chain Monte Carlo algorithms, arXiv :1306.0187v3, [stat.ME] (2014).

[16] R. Tibshirani, Regression shrinkage and selection via LASSO, J. R. Stat. Soc. Ser. B Stat. Methodol. vol. 58 no. 1 267–288 (1996).

[17] R. Tibshirani, The LASSO problem and uniqueness, Electron. J. Stat. vol. 7 1456–1490 (2013).