

DOSSIER DE CANDIDATURE AU POSTE DE MAÎTRE DE CONFÉRENCES DE L'UNIVERSITÉ PARIS 7 (DENIS DIDEROT)

Nom : Yohann de Castro

Thèse : Constructions déterministes pour la régression parcimonieuse

Soutenue le : 3/12/2011

Directeurs : J.-M. Azaïs et F. Barthe

Rapporteurs : G. Lugosi et A. Tsybakov

Jury : J.-M. Azaïs, F. Barthe, F. Gamboa, G. Lécué, A. Tsybakov

Mots-clefs : Problèmes en grandes dimensions, Compressed Sensing, Graphes
expanseurs déséquilibrés, Largeurs de Gelfand, Complétion de
matrices, ACP robuste.

Section : 25, 26 et 27

Numéro dans le SI local : 672

Référence GALAXIE : 4073

Table des matières

Curriculum Vitæ	3
1 Publications	5
2 Activités administratives	6
3 Activités d'enseignement	7
Lettre de recommandation pour l'enseignement	9
4 Résumé des résultats obtenus	11
4.1 Présentation des travaux effectués	11
4.2 Constructions à partir de sections "presque"-Euclidiennes	12
4.3 Constructions à partir de graphes expanseurs déséquilibrés	12
4.4 Reconstruction de mesures signées sur la droite réelle	12
4.5 Autres travaux	13
Lettre de recommandation pour la recherche	15
5 Programme de recherche	19
5.1 Largeur de Gelfand et Compressed Sensing	19
5.2 Graphes expanseurs et complétion de matrices	20
5.3 Dé-bruitage par relaxation convexe sur l'espace des mesures signées	20
Bibliographie	22
Rapport de pré-soutenance du Pr Tsybakov	23
Rapport de pré-soutenance du Pr Lugosi	25
Rapport de soutenance	29
Attestation de réussite au doctorat	31
Déclaration de candidature	33
Copie pièce d'identité	35

Yohann de Castro

État Civil Né le 20 décembre 1983 à Muret (Hte-Garonne)

Français, Pacsé

Thèmes de Recherche

Mots-Clefs Problèmes inverses en grandes dimensions, Compressed Sensing, Complétion de matrices, ACP robuste.

Formation Académique

- 2008 – 2012 **Doctorat en Mathématiques**, *Institut de Mathématiques de Toulouse*, mention très honorable.
 ◦ Thèse : *Constructions déterministes pour la régression parcimonieuse*, soutenue le 3 décembre 2011.
 ◦ Directeurs de thèse : J.-M. Azaïs et F. Barthe.
 ◦ Rapporteurs : G. Lugosi (absent) et A. Tsybakov.
 ◦ Examineurs : F. Gamboa (président du jury) et G. Lécué.
- 2009 – 2010 **Agrégation de Mathématiques**, *admis 2^e décile*.
- 2006 – 2007 **Master 2 Recherche "Mathématiques Vision Apprentissage"**, *ÉNS Cachan*, mention très bien.
 ◦ Mémoire : *Théorèmes de reconstruction et géométrie asymptotique*.
 ◦ Directeurs : E. Candès et Y. Meyer.
- 2005 – 2006 **Master 2 Recherche "Analyse Arithmétique et Géométrie"**, *Université Paris-Sud, Orsay*, mention bien.
- 2004 – 2009 **Magistère "Mathématiques Fondamentales et Appliquées et Informatique"**, *ÉNS Ulm*, mention bien.

Situation Administrative

- 2009 – 2012 **Allocataire Moniteur Normalien**, *Institut de Mathématiques de Toulouse*, Toulouse.
- 2004 – 2009 **Élève fonctionnaire**, *ÉNS Ulm*, Paris.
 inclut un congé sans traitement en 2007-2008

Activités d'Enseignement

- 2009 – 2012 **Monitorat**, *Université Paul Sabatier*, Toulouse.
- 2008 – 2009 **Interrogations orales de Mathématiques en \mathcal{MP}^*** , *Lycée Pierre de Fermat*, Toulouse.
- 2005 – 2007 **Interrogations orales de Mathématiques en \mathcal{MP}^*** , *Lycée Charlemagne*, Paris 4^e.

Publications

- Y. de Castro, *Quantitative Isoperimetric Inequalities on the Real Line*, *Annales Mathématiques Blaise Pascal*, volume 18, n° 2 (2011), p. 311-331.
- Y. de Castro and F. Gamboa, *Exact Reconstruction using Beurling Minimal Extrapolation*, accepté à *Journal of Mathematical Analysis and Applications*, 2011.
- Y. de Castro, *A Remark on the Lasso and the Dantzig Selector*, soumis à *Electronic Journal of Statistics*, arXiv :1108.5533, 2011.
- Y. de Castro, *Error Prediction and Model Selection via Unbalanced Expander Graphs*, soumis à *IEEE IT*, arXiv :1010.2457, 2011.

Communications

- 2011
- Groupe de travail *Complétion de matrices et ACP Robuste*, IMT, 4 exposés (Automne).
 - Séminaire de Statistique de l'ENSAE, Malakov, Novembre.
 - Rencontres des doctorants en Statistique, IMT, Septembre.
 - 41^e école d'été de Probabilités et Statistique de St-Flour, St-Flour, Juillet.
 - École d'été *Sparsity and Model selection* organisée par l'IFUM, Montevideo, Février.
- 2010
- Rencontres des doctorants en Statistique, IMT, Septembre.
 - 9^e rencontres *Jeunes Probabilistes et Statisticiens*, Le Mont-Dore, Mai.
 - Groupe de travail sur les matrices aléatoires, IMT, Janvier.

Vie Scientifique

Séjours/Collaborations à l'étranger

- 2011 *Isaac Newton Institute for Mathematical Sciences*, Cambridge, 2 mois.
- 2007 Avec E. Candès, *California Institute of Technology*, Pasadena, 4 mois.

Responsabilités

- Co-organisateur du groupe de travail "*Complétion de matrices et ACP robuste*" à l'IMT en 2011/2012.
- Organisateur des "*Rencontres des doctorants en Statistique 2011*" à l'IMT.
- Co-organisateur du séminaire étudiant inter-équipes de l'IMT pour l'année 2011/2012.

Expertise

1 Rapport pour *Bernoulli Journal*.

Divers

- 2007 – 2008 Voyage autour du monde pendant 9 mois, <http://9mois-de-voyage.blogspot.com/>.
- 2004 – 2005 Création d'un atelier d'écriture à la Maison d'arrêt de Fresnes, association GENEPI.

Compétences

Informatique

- Maîtrise de \LaTeX , Matlab, Octave, Scilab, R, et Maple.
- Maîtrise du langage Html.

Langues

Français (langue maternelle), Anglais (courant), Espagnol (conversant).

Centres d'intérêt

La randonnée (Aconcagua 6962m, Annapurnas, Andes, Alpes, Pyrénées) et le cyclisme.

1 Publications

Paru

1. Y. de Castro, *Quantitative Isoperimetric Inequalities on the Real Line*, Annales Mathématiques Blaise Pascal, volume 18, n° 2 (2011), p. 311-331.

À paraître

2. Y. de Castro and F. Gamboa, *Exact Reconstruction using Beurling Minimal Extrapolation*, accepté à *Journal of Mathematical Analysis and Applications*, 2011.

Soumis

3. Y. de Castro, *A Remark on the Lasso and the Dantzig Selector*, soumis à *Electronic Journal of Statistics*, arXiv :1108.5533, 2011.
4. Y. de Castro, *Error Prediction and Model Selection via Unbalanced Expander Graphs*, soumis à *IEEE IT*, arXiv :1010.2457, 2011.

Autres Publications

5. Y. de Castro, *Constructions déterministes pour la régression parcimonieuse*, disponible sur TEL, tel.archives-ouvertes.fr/tel-00656449/, 2012.

Correspondance avec les chapitres de ma thèse

1. Chapitre V
2. Chapitre IV
3. Chapitre II
4. Chapitre III

2 Activités administratives

- Co-organisateur (avec J. Bigot) du groupe de travail "Complétion de Matrices et ACP robuste" à l'Institut de mathématiques de Toulouse en Automne 2011.
- Organisateur des rencontres des doctorants en statistique de l'IMT en Septembre 2011.
- Co-organisateur (avec E. Boissard) du séminaire bi-mensuel des étudiants de l'IMT en 2011/2012.
- Rapporteur pour *Bernoulli Journal*.
- Mise en place la page web de l'école d'été "Sparsity and Model selection" de l'Institut Franco-Uruguayen de Mathématiques et membre de l'équipe pédagogique de cette école d'hiver.

3 Activités d'enseignement

Depuis septembre 2009 jusqu'en août 2012, j'effectue mon service de monitorat (64h par année universitaire) au sein de l'Université Paul Sabatier à Toulouse. Avant cela j'ai enseigné en tant que "khôlleur" en classes préparatoires \mathcal{MP}^* pendant 3 années. Par ailleurs, j'ai obtenu l'Agrégation de mathématiques en 2010 (classé 48ième) lors de ma première inscription.

Détail des enseignements

- 2005/2007 :
J'ai été khôlleur deux ans au Lycée Charlemagne (Paris 4ième) en \mathcal{MP}^* . J'ai dispensé plus de 80 h de khôlles sur les deux années à raison de 2 h par semaine pendant 2*20 semaines. Le programme des classes préparatoires recouvre l'algèbre générale, l'algèbre linéaire, l'analyse de Fourier, l'analyse différentielle, les séries entières, les séries de fonctions... J'ai encadré un binôme sur un sujet de TIPE (projet personnel effectué au cours de l'année et à présenter aux concours des grandes écoles) lié à mon sujet de mémoire de master 2 : "*Reconstruction exacte en Compressed Sensing*".
- 2008/2009 :
J'ai effectué 40 h de khôlles en \mathcal{MP}^* au Lycée Pierre de Fermat à Toulouse à raison de 2 h par semaine pendant 20 semaines.
- 2009/2010 :
 - 36 h de TD "*Méthodes et outils de calcul*" en L3 SVS (biologie) à raison de 2 h par semaine pendant 12 semaines. Ce TD pour non-spécialistes développe des outils statistiques et probabilistes (tests, intervalles de confiance, théorème central limite...) autour d'exemples concrets (gestion d'une maternité, évolution d'une population de bactéries...).
 - 24 h de Cours-TD "*Mathématiques*" en L3 SVS à raison de 2 h par semaine pendant 12 semaines. Ce Cours-TD est une introduction aux Probabilités avec une composante Statistique à travers les tests.
- 2010/2011 :
 - 48 h de TD "*Méthodes et outils mathématiques*" en L2 SVS à raison de 4 h par semaine pendant 12 semaines.
 - 12 h de Cours-TD "*Mathématiques*" en L1 SVS à raison de 2 h par semaine pendant 12 semaines. Ce Cours-TD est une introduction aux équations différentielles et à la théorie de l'intégration.
- 2011/2012 :
 - 16 h de TD "*Théorie du signal*" en M1 IMAT (master professionnel "ingénierie mathématiques") à raison de 2 h par semaine pendant 8 semaines. Dans ce TD, on étudie la transformée de Fourier des distributions tempérées, les systèmes linéaires à temps invariant (LTI), la notion de fonction de transfert et de réponse impulsionnelle.
 - (à venir au second semestre) 20 h de Cours-TD "*Méthodes et outils de calculs*" en L2 Biochimie à raison de 2 h par semaine pendant 10 semaines.
 - 5 h de TP "*Signal et ingénierie médicale*" en M2Pro IMAT sur le "Compressed Sensing". Ce TP s'intéresse à des problèmes de reconstructions d'images médicales à partir d'un faible nombre de mesures aléatoires. On y revisite les résultats récents du Compressed Sensing avec une finalité concrète d'ingénierie médicale. Le TP est réalisé sous Matlab.
 - 15 h de TP Matlab "*Plans d'expériences*" en M2Pro IMAT à raison de 3 h par semaine pendant 5 semaines. Ce TP explore, des critères AIC/BIC au lasso en passant par la D-optimalité et les plans de mélanges, les outils numériques pour le modèle linéaire à l'aide des logiciels Matlab et Scilab.

Cours doctoral

Dans le cadre de l'école d'été "*Parcimonie et sélection de modèle*" organisée par l'Institut Franco-Uruguayen de Mathématiques (IFUM) à Montevideo, j'ai eu l'occasion de donner un cours à des étudiants en Master 2 recherche et des doctorants d'Amérique latine. J'ai dispensé 4 h de cours "*A short geometric tour of the sparse regression*" sur le modèle parcimonieux et le Compressed Sensing.

Synthèse

Mes activités d'enseignement m'ont permis d'appréhender la pluralité du métier d'enseignant tant par la diversité des audiences que par celle des missions à accomplir. J'ai essayé autant que possible de garder un même enseignement deux années de suite afin d'apporter aux étudiants un plus grand recul. Je suis particulièrement resté attaché à mon premier cours en L3 SVS et j'ai d'ailleurs enseigné dans cette filière tout au long de mon monitorat. J'y ai rencontré un public non-spécialiste impliqué et réceptif. Ce fût une expérience très enrichissante qui m'a permis de faire un lien fort entre un cours académique et des problèmes concrets. En troisième année, j'ai pu enseigner à des élèves de master professionnel et parachever ce lien. Les TP Matlab et Scilab furent une expérience gratifiante où j'ai pu partager mes acquis de master et d'Agrégation et mettre en oeuvre des sujets abordés au cours de ma thèse. Durant ces trois années de monitorat, j'ai dispensé un large spectre d'enseignements à des publics variés : des étudiants en Licence SVS et d'autres en Master 2 professionnel prêts à s'engager dans leur future carrière d'ingénieur. D'une manière générale, j'ai pris beaucoup de plaisir au contact des étudiants et des équipes dont j'ai fait partie. Il m'est particulièrement agréable de pouvoir concilier, avec grand profit, l'exercice pédagogique et mon activité de recherche.

Toulouse, le 6 Mars 2012

M. Yohann de Castro est moniteur au Département de Mathématiques de l'Université Paul Sabatier de Toulouse depuis Septembre 2009.

Durant ces trois années, il a effectué la plupart de ses enseignements dans les diverses licences de Biologie (L1 SVS, L2 CBPS, L2 SVS, L2 Biochimie) dans des UE dont je suis, ou ai été, le responsable.

Personnellement, je suis entièrement satisfait du travail qu'a fourni M. de Castro. Il a activement participé à l'élaboration et l'amélioration des supports pédagogiques et s'est naturellement investi dans les tâches telles que les surveillances d'examen, les soutiens et les expériences pédagogiques.

Par ailleurs, le sérieux ainsi que la personnalité de M. de Castro font de lui un collègue avec qui il est agréable de travailler et sur qui on peut compter.

Je le recommande donc très chaleureusement !



Philippe Monnier
Maître de Conférence
05 61 55 60 28
monnier@math.univ-toulouse.fr

4 Résumé des résultats obtenus

Mes travaux portent sur le *Compressed Sensing*, théorie dont l'objet est la reconstruction d'un signal de grande dimension à partir de peu d'observations aléatoires. Ce domaine a connu un essor important lors de la dernière décennie tant par les nombreuses applications qu'il a bouleversé (imagerie médicale, génétique, analyse micro-array...) que par le nouvel éclairage qu'il apporte sur la Statistique. C'est un domaine porteur et compétitif qui occupe une grande place dans la communauté Statistique aujourd'hui.

J'ai commencé à travailler sur le *Compressed Sensing* en 2007 sous la direction de E. Candès lors d'un séjour de 4 mois à Caltech, et avec Y. Meyer à l'ENS de Cachan lors de mon mémoire de Master 2. J'ai naturellement continué sur cette problématique lors de mon doctorat sous la direction de Jean-Marc Azaïs et Franck Barthe à l'Institut de Mathématiques de Toulouse de septembre 2008 à décembre 2011.

4.1 Présentation des travaux effectués

Le cadre classique du *Compressed Sensing* consiste à reconstruire une cible à partir d'un petit nombre d'observations aléatoires et en utilisant un programme linéaire facilement et efficacement implémentable. On considère le modèle linéaire en grandes dimensions :

$$y = X\beta^* + z, \quad (\text{ML})$$

où $y \in \mathbb{R}^n$ est l'observation, $X \in \mathbb{R}^{n \times p}$ est la matrice de design (dont les entrées sont tirées au hasard selon une loi connue), $\beta^* \in \mathbb{R}^p$ est la cible, et $z \in \mathbb{R}^n$ est le bruit. On se place dans le cas où $n \ll p$, i.e. le système d'équations représenté par X est fortement sous-déterminé. On considère alors les estimateurs lasso :

$$\beta^\ell \in \arg \min_{\beta \in \mathbb{R}^p} \left\{ \frac{1}{2} \|y - X\beta\|_{\ell_2}^2 + \lambda_\ell \|\beta\|_{\ell_1} \right\}, \quad (\text{LASSO})$$

où $\lambda_\ell > 0$ est un paramètre à régler, et sélecteur Dantzig :

$$\beta^d \in \arg \min_{\beta \in \mathbb{R}^p} \|\beta\|_{\ell_1} \text{ sachant que } \|X^\top (y - X\beta)\|_{\ell_\infty} \leq \lambda_d, \quad (\text{DS})$$

où $\lambda_d > 0$ est un paramètre à régler. Ces deux estimateurs peuvent être implémentés efficacement.

L'objet de mes travaux doctoraux était de proposer des constructions déterministes s'abstrayant de l'aspect aléatoire de la matrice de design X . Mon approche fût de regarder du côté de la théorie des sous-espaces "presque"-Euclidiens, des graphes expanseurs déséquilibrés, et de la reconstruction des mesures signées sur la droite réelle :

1. Dans un premier temps, j'ai étendu un résultat de Kashin et Temlyakov [KTo7] au cas d'observations bruitées. Ce résultat montre que les erreurs de prédiction (i.e. estimer $\|X\beta^l - X\beta^*\|_{\ell_1}$) et d'estimation (i.e. estimer $\|\beta^l - \beta^*\|_{\ell_1}$) des estimateurs classiques (lasso, sélecteur Dantzig) sont liés à la distorsion (rapport entre la norme ℓ_1 et la norme ℓ_2) du noyau de la matrice de design.
2. Dans un second temps, je me suis intéressé aux designs construits à partir de matrices d'adjacence de graphes expanseurs déséquilibrés. J'ai montré qu'il y avait un lien entre les erreurs de prédiction et d'estimation, et la constante d'expansion de ces graphes.
3. En collaboration avec F. Gamboa, je me suis intéressé à la reconstruction de mesures signées discrètes sur la droite réelle. On a montré que, parmi toutes les mesures signées dont quelques moments "généralisés" sont fixés, la mesure de variation totale la plus faible est la mesure dont le cardinal du support est le plus petit. Ce résultat nous permet de construire une large classe de matrices de designs à partir de matrices de Vandermonde.

4.2 Constructions à partir de sections "presque"-Euclidiennes

En 2007, Kashin et Temlyakov [KT07] ont montré que les erreurs des estimateurs classiques dépendaient de la distorsion du noyau de la matrice de design. Cette dernière mesure l'écart entre une boule euclidienne et l'intersection du noyau avec la boule unité ℓ_1 . Un autre point de vue est de dire que la distorsion contrôle le rapport entre la norme ℓ_1 et la norme ℓ_2 sur le noyau du design. Cependant leur résultat concernait seulement le cas d'observations non-bruitées (i.e. $z = 0$ dans (ML)).

Dans l'article [dC11b], j'ai étendu ce résultat au cas d'observations bruitées. En particulier, j'ai montré que le lasso (LASSO) et le sélecteur Dantzig (DS) satisfont des *inégalités oracles* en termes de la distorsion du noyau du design. Pour cela, j'ai introduit une nouvelle condition sur la matrice de design X , la *condition UDP*. Celle-ci est la plus faible condition à satisfaire pour garantir que les erreurs (de prédiction et d'estimation) sont comparables aux erreurs que l'on obtient dans un cadre "idéal" (avec un "oracle") où l'on connaît à l'avance le support des plus grands coefficients de la cible β^* . De plus, cette condition peut être exprimée en terme seulement de la distorsion du noyau de la matrice de design X . Ainsi, j'ai montré qu'un estimateur appréhende d'autant mieux l'information importante de la cible (i.e. ses plus grands coefficients) que la distorsion du noyau de la matrice de design est petite.

4.3 Constructions à partir de graphes expenseurs déséquilibrés

Le Compressed Sensing a de fortes interactions avec la théorie des codes correcteurs. En effet, une matrice de design X peut être vue comme la matrice "Parity-Check" d'un code. Il est connu que de "bons" codes (en termes de distance relative et de "taux R ") peuvent être construits à partir de matrices d'adjacence de graphes bi-parti particuliers : les *expenseurs déséquilibrés*. Une question naturelle est d'étudier les performances de matrices de design X issues de matrices d'adjacence de graphes expenseurs déséquilibrés. Dans l'article [dC10], on montre que de telles matrices donnent des erreurs de prédiction et d'estimation comparables aux erreurs obtenues avec un oracle. On exhibe le lien entre la constante d'expansion et l'écart avec l'oracle des erreurs de prédiction et d'estimation.

Dans mon mémoire de thèse, je montre que ces matrices de design satisfont la propriété UDP. De plus, cette condition peut être exprimée en terme seulement de la constante d'expansion (et non plus en terme de la distorsion du noyau du design).

4.4 Reconstruction de mesures signées sur la droite réelle

En collaboration avec Fabrice Gamboa, nous avons étudié la reconstruction fidèle de mesures signées σ de support fini sur la droite réelle. Celle-ci peut être décrite comme suit. Soit I un borélien de \mathbb{R} que l'on suppose borné. Soit σ une mesure de support fini inclus dans I :

$$\sigma = \sum_{i=1}^s \sigma_i \delta_{x_i},$$

où σ_i sont les poids et δ_x est la masse de Dirac en x . On souligne que ni les poids, ni les points x_i ne sont connus. On observe $n + 1$ moments généralisés :

$$c_k(\sigma) = \int_I u_k d\sigma, \quad (1)$$

où $k = 0, \dots, n$ et $\mathcal{F} := \{u_0, \dots, u_n\}$ est une famille de fonctions continues sur la fermeture de I . On suppose que \mathcal{F} est un M -système homogène. La définition précise de ce dernier se trouve dans [dCG11]. On rappelle qu'ils englobent les moments standards, la transformation de

Stieltjes, la transformée de Laplace, la fonction caractéristique... On introduit l'estimateur GME (pour "Generalized Minimal Extrapolation") :

$$\sigma^* \in \arg \min_{\mu \in \mathcal{M}} \|\mu\|_{TV} \quad \text{tel que } \mathcal{K}_n(\mu) = \mathcal{K}_n(\sigma), \quad (\text{GME})$$

où $\|\cdot\|_{TV}$ est la norme de variation totale, $\mathcal{K}_n(\mu) := \{c_0(\mu), \dots, c_n(\mu)\}$, et \mathcal{M} représente l'ensemble des mesures signées sur I .

Dans l'article [dCG11], on montre que l'on peut reconstruire fidèlement toute mesure signée de support fini à partir de l'observation de peu de moments généralisés. Dans le cas où la mesure cible σ est positive et de support de taille s , on montre que $n = 2s + 1$ moments suffisent. Ce résultat peut être transposé à notre problème originel de régression parcimonieuse. On montre ainsi que tout vecteur $\beta^* \in \mathbb{R}^p$ de support de taille s et ayant ses coefficients positifs, est l'unique solution du programme suivant :

$$\beta^{bp} \in \arg \min_{X\beta=y} \|\beta\|_{\ell_1}. \quad (\text{BP})$$

où $X \in \mathbb{R}^{n \times p}$ est une matrice de Vandermonde "généralisée" de taille telle que $n = 2s + 1$, et $y = X\beta^*$ est l'observation (ici le bruit est nul, $z = 0$). Ce résultat était connu auparavant mais nous l'avons étendu à une classe plus large de matrices. L'estimateur β^{bp} est appelé basis pursuit.

4.5 Autres travaux

Dans [dC11a], on s'intéresse à des inégalités isopérimétriques quantitatives optimales sur la droite réelle et l'on prouve que parmi les ensembles de mesure donnée et d'asymétrie donnée (distance à la demi-droite, i.e. distance aux ensembles de périmètre minimal), les intervalles ou les complémentaires d'intervalles ont le plus petit périmètre. Les travaux exposés utilisent uniquement des outils géométriques et complètent le remarquable résultat [CFMP11] en précisant la stabilité des inégalités isopérimétriques dans le cas des mesures log-concaves symétriques sur la droite réelle.

LETTRE DE RECOMMANDATION DE YOHANN DE CASTRO
POUR UN POSTE DE MAÎTRE DE CONFÉRENCE

Guillaume Lecué
Bureau 4B093
Laboratoire d'analyse et mathématiques appliquées
Université Paris Est – Marne-la-vallée
Cité Descartes – Champs-sur-Marne
5 Boulevard Descartes
77454 Marne-la-vallée Cedex 2
France

Tel : +33 (0) 1 60 95 77 32
Tel : +33 (0) 6 60 80 91 39
Fax : +33 (0) 1 60 95 75 45
guillaume.lecue@univ-mlv.fr

Au président du comité de sélection,

Yohann de Castro a effectué sa thèse sous la direction de Jean-Marc Azaïs et Franck Barthe à L'université Toulouse III - Paul Sabatier. C'est au cours de cette période que j'ai pu interagir avec Yohann notamment autour des problèmes du Compressed Sensing et des méthodes par pénalisation ℓ_1 lors de différents séminaires ou au cours de différentes écoles thématiques. J'ai aussi fait partie du jury de thèse de Yohann en Décembre dernier.

Les connaissances acquises et les résultats obtenus par Yohann durant sa thèse ont suscité mon intérêt à plusieurs reprises. Premièrement, la construction de matrices déterministes pour le problème du Compressed Sensing est un sujet d'une réelle importance pour de nombreuses communautés mathématiques. Yohann fait partie du petit nombre de chercheurs à avoir étudié ce problème très difficile. Les matrices déterministes qu'il a proposées avec Fabrice Gamboa, et pour lesquelles ils ont obtenues des résultats de reconstruction,

ont rendu les recherches de Yohann très attractives pour de nombreux chercheurs. Les connaissances que Yohann a acquises sur les graphes expandeurs au cours de ses recherches en thèse sont d'un grand intérêt à mes yeux car elles me semblent être la voie la plus prometteuse pour la construction de matrices déterministes en Compressed Sensing. Un autre aspect du problème de Compressed Sensing que Yohann a particulièrement développé concerne le “certificat dual”. Cet aspect dual du problème d'optimisation de l'algorithme du Basis Pursuit a conduit Yohann à relier le problème du Compressed Sensing (et plus généralement le problème de reconstruction exacte de mesures signées positives à support fini) à des problèmes d'interpolation de polynôme dans la même veine que les premiers résultats de [Donoho et Tanner] et [Candes et Tao] en Compressed Sensing. C'est grâce à ce parallèle entre propriété d'interpolation de certains polynômes et reconstruction exacte que Yohann a proposé la construction de matrices déterministes ne nécessitant que $2s + 1$ mesures pour la reconstruction de tout vecteur s -sparse à coordonnées positives.

Deuxièmement, Yohann a étudié les propriétés de reconstruction/estimation du LASSO sous une nouvelle “hypothèse de distortion” de la matrice de design qui a été introduite par lui-même. Cette propriété est en un certain sens aussi fondamentale que la propriété de section euclidienne du noyau de la matrice de mesures en Compressed Sensing. Ce genre d'hypothèse permet donc d'étudier un problème d'estimation grâce à des outils de la géométrie des espaces de Banach qui ont été développés depuis plus de soixante ans. Une profonde étude de cette condition de distortion reste encore à faire et peut mener Yohann à une nouvelle compréhension de l'algorithme LASSO. Par ailleurs, Yohann a montré que la matrice d'adjacence d'un graphe expander vérifie cette condition de distortion.

Finalement, l'intérêt développé par Yohann pour les applications en traitement du signal et en analyses de données en grandes dimensions lui permettent d'interagir avec un grand nombre de statisticiens du plus théorique au plus appliqué. Ce champs d'interaction est un atout considérable pour tout mathématicien.

Yohann a acquis une grande compréhension des problèmes de Compressed Sensing et de traitement des données en grandes dimensions au cours de sa thèse. Malgré la très forte compétitivité dans ces deux milieux, Yohann a su obtenir des résultats originaux et d'un grand intérêt. Il n'est pas de doute à mes yeux que Yohann est amené à obtenir d'autres très beaux résultats en Compressed Sensing dans les années à venir. Je recommande donc avec la plus grande conviction la candidature de Yohann pour un poste de Maître de conférence.

Je me tiens à votre disposition pour plus d'informations et je vous pris d'accepter mes salutations distinguées.

Paris, le 19 Mars 2012.
Guillaume Lécué.

5 Programme de recherche

Ce programme de recherche comporte trois composantes dont deux sont un prolongement naturel de ma thèse de doctorat. Tout d'abord, je souhaite étendre les interactions entre la théorie de l'approximation dans les espaces de Banach (sous-espaces presque-Euclidiens, largeur de Gelfand...) avec le Compressed Sensing à de nouveaux estimateurs encore mal connus sur le plan théorique : group lasso, elastic net, procédure à plusieurs étapes... Ces estimateurs sont largement employés en génétique (analyse micro-array), en économie, et en imagerie médicale. Dans un second temps, il me semble pertinent d'aborder la question de la "dérandomisation" de l'approche classique de la complétion de matrices. De nombreuses questions de premier plan restent encore ouvertes comme par exemple les interactions entre graphes (propriété d'expansion, vitesse de convergence des marches aléatoires...) et l'estimation de matrices de faible rang. Ces questions théoriques trouvent un écho tout particulier dans les nombreuses applications incidentes à ce domaine comme la physique quantique avec la tomographie des états quantiques [Gro11]. Enfin, j'aimerais questionner la stabilité de l'estimateur "Generalized Minimal Extrapolation" qui a permis de construire toute une nouvelle classe de matrices de design pour le Compressed Sensing. La question de l'estimation avec bruit semble s'imposer d'elle même. Cela permettrait d'offrir de nouvelle procédure de compression de données en Informatique.

Mon programme s'articule autour du Compressed Sensing, la théorie des graphes, la théorie de l'approximation dans les espaces de Banach, l'étude des processus empiriques et la relaxation convexe dans les espaces de Banach.

5.1 Largeur de Gelfand et Compressed Sensing

Motivations

La théorie de l'approximation dans les espaces de Banach permet d'exprimer l'erreur minimax de la meilleure procédure d'estimation d'une classe de vecteurs \mathcal{C} en fonction de la largeur de Gelfand de \mathcal{C} , voir par exemple [CDD09]. Dans la même veine, l'article [dC11b] montre que la largeur de Gelfand de la boule ℓ_1 (donnant la plus petite largeur des sections presque-Euclidiennes) permet d'exprimer l'erreur d'estimation et de prédiction des estimateurs lasso et sélecteur Dantzig. À ma connaissance, aucun résultat de ce type n'existe pour des estimateurs comme le *group lasso*, l'*elastic net*, ou des procédures à plusieurs étapes comme le lasso repondéré. Ces estimateurs sont intensivement utilisés dans la pratique et encore mal connus d'un point de vue théorique. Ils sont très présents en traitement d'image et en génétique où certains groupes de loci ont tendance à être s'exprimer simultanément (e.g. les loci influençant le taux de cholestérol).

Objectif

Nous souhaitons étendre le résultat de [dC11b] à des estimateurs prenant en compte des structures de groupes de coordonnées du type *group lasso*, ou des estimateurs à plusieurs étapes comme le lasso repondéré. Appliquer ces résultats à la matrice du Génome.

Mise en oeuvre

Un première étape a été franchie dans [dC11b] où la condition UDP fait le lien entre largeur de Gelfand et inégalités oracles. Nous avons bon espoir qu'une telle démarche donnerait des résultats pertinents dans le cas du group lasso et de l'elastic net. Pour cela il est nécessaire d'exhiber la plus faible condition pour avoir des inégalités oracles. Après avoir exhibé cette condition, il faudra calculer la largeur de Gelfand correspondante (dépendant des normes mises en

jeu dans la condition exhibée). Les techniques classiques pour cela sont inspirées des techniques de la concentration de la mesure et des processus empiriques.

5.2 Graphes expenseurs et complétion de matrices

Motivations

Le problème de complétion de matrices a émergé récemment avec, en particulier, les travaux de D. Gross [Gro11]. Il a une place de choix en physique quantique puisqu'il est une généralisation du problème de tomographie des états quantiques. D'une certaine manière, il peut être vu comme la généralisation non-commutative du Compressed Sensing. En particulier, l'inégalité de Bernstein non-commutative joue un rôle important dans cette approche. De ce fait, nous sommes tenus de "randomiser" les observations de notre matrice dans notre étude théorique. D'une manière plus générale, les lieux des coefficients observés peuvent être vus comme la matrice d'adjacence d'un graphe. Une étude simple dans le cas d'une matrice de rang 1 a montré que les propriétés du graphe (connectivité) jouaient un rôle prépondérant.

Objectif

"Dérandomizer" l'approche classique qui utilise des observations aléatoires de la matrice cible. Exhiber les liens entre graphe (les observations étant vus comme une matrice d'adjacence) et l'estimation de la matrice cible. Appliquer ces résultats à la tomographie des états quantique. Étendre ces résultats à l'ACP robuste et l'appliquer au traitement d'images (détection de mouvement sur une caméra).

Mise en oeuvre

Un graphe aléatoire a de très bonne propriété d'expansion. Il semblerait que ce soit l'une des raisons pour lesquelles l'approche aléatoire donne des résultats quasi-optimaux. Nous souhaitons reprendre l'argument de D. Gross [Gro11] du *certificat dual inexact* et le vérifier lorsque les observations sont faites selon la matrice d'adjacence d'un graphe expenseur. Il faudra en particulier calculer la norme opérateur d'une projection matricielle qui dépend du graphe.

5.3 Dé-bruitage par relaxation convexe sur l'espace des mesures signées

Motivations

L'article [dCG11] étudie la reconstruction d'une mesure signées à partir de l'observation d'un petit nombre de ses moments. Cette démarche a permis de donner une nouvelle classe de matrices de design pour la régression parcimonieuse des vecteurs positifs. Dans la pratique, nous observons une version bruitée des moments. De même les questions de stabilité de la procédure décrite dans [dCG11] pourraient être mise en perspective avec un problème de reconstruction avec observations bruitées.

Objectif

Étendre l'estimateur *Generalized Minimal Extrapolation* (GME) au cas d'observation bruitée. En déduire des résultats pour la régression parcimonieuse dans \mathbb{R}^p . Appliquer ces résultats au Compressed Sensing.

Mise en oeuvre

Afin de palier à la perturbation due au bruit, il sera nécessaire d'étudier le spectre de l'opérateur des moments. Les techniques d'optimisation convexe (conditions KKT) permettront de restreindre cette étude à un cône bien choisi. Une autre voie consiste à décrire les conditions d'optimalité comme un problème d'interpolation. La théorie de l'approximation polynomiale pourra être utile.

Références

- [CDD09] A. Cohen, W. Dahmen, and R. DeVore. Compressed sensing and best k-term approximation. *J. Amer. Math. Soc.*, 22(1) :211–231, 2009.
- [CFMP11] A. Cianchi, N. Fusco, F. Maggi, and A. Pratelli. On the isoperimetric deficit in gauss space. *American Journal of Mathematics*, 133(1) :131–186, 2011.
- [CGLP11] D. Chafaï, O. Guédon, G. Lecué, and A. Pajor. Interactions between compressed sensing, random matrices, and high dimensional geometry. *forthcoming book in preparation available on the Internet*, 2011.
- [dC10] Y. de Castro. Error prediction and variable selection via unbalanced expander graphs. arXiv :1010.2457, 2010.
- [dC11a] Y. de Castro. Quantitative isoperimetric inequalities on the real line. *Annales Mathématiques Blaise Pascal*, 18(2) :311–331., 2011.
- [dC11b] Y. de Castro. A remark on the lasso and the dantzig selector. arXiv, 2011.
- [dCG11] Y. de Castro and F. Gamboa. Exact reconstruction using beurling minimal extrapolation. en revision à *Journal of Mathematical Analysis and Applications*, 2011.
- [Gro11] D. Gross. Recovering low-rank matrices from few coefficients in any basis. *Information Theory, IEEE Transactions on*, 57(3) :1548–1566, 2011.
- [KT07] B.S. Kashin and V.N. Temlyakov. A remark on compressed sensing. *Mathematical notes*, 82(5) :748–755, 2007.

RAPPORT SUR LA THÈSE DE DOCTORAT DE YOANN DE CASTRO INTITULÉE

“Constructions déterministes pour la régression parcimonieuse”

Le problème principal étudié dans le travail de thèse de Yoann de Castro est celui de l'estimation statistique dans le modèle de régression linéaire de grande dimension $y = X\beta^* + \xi$. Le vecteur d'observations $y \in \mathbf{R}^n$ et la matrice déterministe $X \in \mathbf{R}^{n \times p}$ sont donnés, ξ est un bruit gaussien et l'objectif est d'estimer le vecteur inconnu $\beta^* \in \mathbf{R}^p$ ou de faire la prévision, *i.e.*, d'estimer $X\beta^*$. La difficulté du problème vient du fait que la dimension p peut être (beaucoup) plus grande que le nombre d'observations n . Dans ce cas, les approches classiques telles que la méthode des moindres carrés ne donnent pas de bon résultat. Ces dernières années, plusieurs méthodes alternatives ont été proposées et étudiées dont les plus célèbres sont le Lasso et le Dantzig selector. Pour les modèles de grande dimension ($p \gg n$), ces méthodes s'avèrent le plus souvent efficaces en pratique mais leur étude théorique se heurte aux fortes restrictions. Notamment, il est impossible d'obtenir un bon estimateur du vecteur β^* sans hypothèse supplémentaire spécifique sur sa structure et sur celle de la matrice X . Une hypothèse naturelle sur β^* convenable pour la plupart des applications est celle de parcimonie : seulement s coordonnées du vecteur β^* sont non-nulles et $s \ll p$. La situation avec les hypothèses sur X est plus délicate. Plusieurs travaux récents proposent différentes conditions sur X permettant d'établir de bonnes inégalités d'oracle pour le Lasso et le Dantzig selector (la condition de cohérence, d'isométrie restreinte, de valeur propre restreinte, de compatibilité etc.). La question importante est à savoir quelles sont les hypothèses minimales. Le travail de Yoann de Castro donne des éléments de réponse à cette question. Il développe une approche originale en établissant un parallèle avec le problème géométrique de longue date de la construction des sections presque-Euclidiennes de la boule ℓ_1 dans \mathbf{R}^p ayant la distorsion la plus petite possible. S'inspirant de ce cercle d'idées, Yoann de Castro introduit une nouvelle condition qu'il appelle propriété universelle de distorsion et il montre qu'elle constitue l'hypothèse sur X la plus faible parmi celles connues dans la littérature pour obtenir des inégalités d'oracle pour le Lasso et le Dantzig selector. Sous la propriété universelle de distorsion, il développe explicitement au Chapitre 2 de sa thèse les inégalités d'oracle pour l'estimation en norme ℓ_1 et pour la prévision. La question étudiée au Chapitre 3 est celle de la construction de matrices X vérifiant la propriété universelle de distorsion. Comme cette propriété est la plus faible parmi celles connues dans la littérature, toutes les constructions vérifiant, par exemple, l'isométrie restreinte conviennent. Or, souvent ces constructions sont

randomisées et l'isométrie restreinte n'est vérifiée qu'avec une certaine probabilité, tandis que Yoann de Castro s'intéresse à la propriété exacte et donc il cherche des matrices X déterministes. La nouveauté intéressante consiste dans le fait que, sous certaines hypothèses, on peut utiliser les matrices d'adjacence des graphes expandeurs. Yoann de Castro montre que de telles matrices X vérifient la propriété universelle de distorsion.

Le Chapitre 4 propose une extension du problème de reconstruction exacte de vecteurs parcimonieux à un espace de dimension finie (un travail avec F. Gamboa). Yoann de Castro propose ici une nouvelle méthode pour le problème des moments qu'il appelle *support pursuit*. Il montre que le support pursuit permet de reconstruire une mesure positive de support fini σ à partir de n moments généralisés de σ dès lors que n est strictement plus grand que deux fois le cardinal du support de σ et les moments généralisés sont obtenus à partir d'un M -système de fonctions homogène. Ce résultat généralise celui de Donoho et Tanner (2005) d'une part en l'étendant à un espace de dimension finie et d'autre en proposant toute une série de nouvelles constructions déterministes pour la matrice X de codage dans le problème de type "compressed sensing". Ces matrices sont notamment obtenues comme les matrices de Vandermonde généralisées associées aux divers M -systèmes (polynômes algébriques ou trigonométriques, systèmes associés aux transformations de Stieltjes et Laplace etc.)

Le Chapitre 5 présente un résultat sur les inégalités isopérimétriques. Notamment, Yoann de Castro montre que parmi les ensembles de mesure donnée et d'asymétrie donnée, les intervalles ou les complémentaires d'intervalles ont le plus petit périmètre. Ce résultat est démontré pour les mesures log-concaves sur la droite. Il étend à cette famille de mesures le travail de Cianchi et al. (2011) qui traite les mesures gaussiennes.

En conclusion, Yoann de Castro étudie dans sa thèse des problèmes innovants et difficiles sur le plan mathématique. Il montre clairement une excellente maîtrise des outils mathématiques variés et il obtient des résultats très originaux et intéressants. Je suis donc tout à fait favorable à la soutenance de sa thèse.

Paris, le 10 novembre 2011,



Alexandre Tsybakov

Review of the thesis
Constructions déterministes pour la
régression parcimonieuse
by Yohann de Castro

The principal topic of this thesis is high-dimensional linear regression. In particular, “oracle inequalities” are derived for the lasso and the Danzig selector. This is an important topic of great interest in modern statistical theory, as these methods of linear regression have proved to be very effective and efficiently computable in a large variety of application. A lot of effort has been devoted to the better theoretical understanding of how and under what circumstances these methods work and the Chapters 2 and 3 of this thesis contribute to this line of research. In particular, the author attempts to offer verifiable conditions (in the sense of efficiently computable verification) on the design matrix that guarantee that lasso (and related methods) work well. This is an important problem because, although it has been shown that random design matrices work well with high probability, deterministic constructions with good performance guarantees are called for.

To this end, the author introduces a new condition, the so-called “universal distortion property.” It is shown that this condition implies oracle inequalities for both the lasso and the Danzig selector both for variable selection and for error prediction. In Chapter 3 it is shown that unbalanced bipartite expander graphs may be used to construct design matrices that satisfy the universal distortion property. Many fascinating connections with convex geometry and coding theory are revealed.

Chapter 4 is dedicated to an interesting problem related to ℓ_1 -penalized regression: the problem of reconstruction of signed measures from linear measurements. The performance of the “support pursuit” algorithm is studied in a quite general framework. The main result extends a result of Donoho and Tanner for a deterministic design for exact recovery of nonnegative sparse vectors to general homogeneous M-systems.

The topic of Chapter 5 is somewhat independent of the rest of the thesis. Here the author establishes sharp quantitative isoperimetric inequalities on the real line for log-concave measures, extending an earlier result for the gaussian measure. The main result (Theorem V.9) is very nice and potentially useful.

In summary, I think this thesis contains independent work of high quality. Numerous novel results are given in very interesting and important topics.

The thesis is written with a mathematical rigor, and reflects the work of an original and talented researcher. In my opinion this is a very solid thesis and I am happy to recommend that the Ph.D. degree is given to Yohann de Castro.

Below I describe a few remarks and questions.

Remarks.

- Definition II.2 is of central importance and its main importance lies in its verifiability. I would have liked to see a detailed discussion about why and how one can verify this condition. In particular, can this condition be verified in polynomial time for any matrix?
- On p.25, just before (II.12), the case (II.5) is referred to as “optimal”. In what sense is this optimal?
- It would be nice to have a discussion about Table 1 on p. 29. In particular, are the reverse implications true for some value of the parameters of UDP?
- I did not understand the last column in Table 2.
- It would be interesting to understand the advantages of general M-systems as opposed to moment and trigonometrical families. In what practical problems can this extension be important?
- I do not understand the first remark on page 68: why is the symmetry assumption not restrictive? Not all log-concave densities can be shifted so that this assumption is satisfied. (The numbering is missing so I am not sure if (iv) refers to this assumption.)
- The bibliographic survey is nice but seems to miss some important references like the papers of Bunea, Tsybakov and Wegkamp; Koltchinskii; Bartlett, Mendelson and Neeman. Of course the area is huge and it is difficult to be exhaustive but these results seem to be closely related to the oracle inequalities given here.

Minor remarks.

- The thesis is generally well written but English needs to be checked carefully. I do not give an exhaustive list of grammatical and spelling problems but there are quite a few.
- p.22, l.-8: singular value

- p.27, middle of page: Restricted Isometry (and not Isoperimetric) Property
- ϕ and Φ should be the same in Section 4.2, p.30.
- p.39, l. 3: Unless otherwise specified...
- (III.16) and (III.17) only hold with high probability so this should be mentioned.
- Definition IV.1: “Extrema Jordan type” sounds strange. Maybe “Extremal” would be more adequate.
- p.50, l.9: finite linear combination
- p.55, l.-8: pioneering
- Lemma V.3: the notation d has not been introduced.

Gábor Lugosi
 ICREA Research Professor
 Pompeu Fabra University
 Barcelona.

PROCES VERBAL DE SOUTENANCE DU 03/12/2011 A 10h00

ANNEE UNIVERSITAIRE 2011/2012

Etudiant : M. YOHANN DE CASTRO né le : 20/12/1983

Version de diplôme : DOCT. UNIV. DE MATHEMATIQUES APPLIQUEES

Titre des travaux : Constructions déterministes pour la régression parcimonieuse

Ecole doctorale : Ecole Doctorale Math., Info., Télécomm.

Formation doctorale : Mathématiques Appliquées

Directeur : M. JEAN MARC AZAIS

Codirecteur : M. FRANCK BARTHE

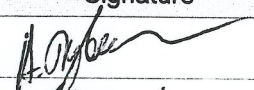



Lieu de soutenance : IMT-AMPHI SCHWARTZ

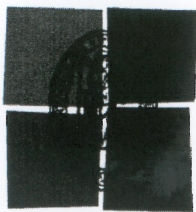
La soutenance est publique.

Résultat : admis

Mention :

Membres du Jury

Nom	Qualité	Etablissement	Rôle	Signature
M. ALEXANDRE TSYBAKOV	PROFESSEUR DES UNIVERSITES	UNIVERSITE PARIS 6	Rapporteur	
M. JEAN MARC AZAIS	PROFESSEUR DES UNIVERSITES	UNIVERSITE DE TOULOUSE III	Membre	
M. FABRICE GAMBOA	PROFESSEUR DES UNIVERSITES	UNIVERSITE DE TOULOUSE III	Membre	
M. GUILLAUME LECUE	CHARGE DE RECHERCHE	UNIVERSITE PARIS 5	Membre	



Université
de Toulouse



DOCTORAT DE L'UNIVERSITE DE TOULOUSE
délivré par l'Université Toulouse III – Paul Sabatier

Spécialité :

RAPPORT DE SOUTENANCE

Nom du (ou de la) candidat(e) :

Date de soutenance :

Président du jury :

Yohann DE CASTRO a présenté avec beaucoup de maturité ses travaux sur les constructions déterministes pour la régression parcimonieuse.

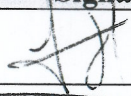

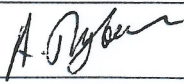

L'exposé a été particulièrement bien construit, motivé par les applications concrètes, illustrées par des exemples originaux en traitement de l'image.

Yohann DE CASTRO a su mettre en avant de façon élégante les principales idées mathématiques de sa thèse qui contient des résultats originaux sur un sujet difficile et très compétitif.

Le jury unanime et convaincu que Yohann DE CASTRO a toutes les qualités requises pour faire une brillante carrière de chercheur ou d'enseignant chercheur, lui décerne le titre de Docteur de l'Université de Toulouse 3 Paul Sabatier avec la mention très honorable.

Le diplôme national de Docteur est délivré par l'Université Toulouse III – Paul Sabatier sans aucune mention.

Les Membres du Jury :

Nom	Signature	Nom	Signature
Gambou			
Lecuo			
Tsybakov			
AZAR			



Université
de Toulouse

Université Paul Sabatier - Toulouse III

ATTESTATION DE RÉUSSITE AU DIPLÔME

Le Président de l'Université de Toulouse et Le Président de Toulouse III attestent que

le Diplôme national de docteur en Mathématiques appliquées de l'Université de Toulouse
a été décerné à

Monsieur DE CASTRO YOHANN

né le 20 décembre 1983 à MURET (031)

au titre de l'année universitaire 2011/2012

Titre des travaux : Constructions déterministes pour la régression parcimonieuse
Date de soutenance : 3 décembre 2011
Etablissement soutenance : UNIVERSITE DE TOULOUSE III
Jury : M. FABRICE GAMBOA, Président du jury, PROFESSEUR DES UNIVERSITES
UNIVERSITE DE TOULOUSE III
M. ALEXANDRE TSYBAKOV, Rapporteur du jury, PROFESSEUR DES UNIVERSITES
UNIVERSITE PARIS 6
M. JEAN MARC AZAIS, Membre du jury, PROFESSEUR DES UNIVERSITES
UNIVERSITE DE TOULOUSE III
M. GUILLAUME LECUE, Membre du jury, CHARGE DE RECHERCHE
UNIVERSITE PARIS 5
Ecole doctorale : Ecole Doctorale Math., Info., Télécomm.

Fait à TOULOUSE, le 6 décembre 2011

Le Président de Toulouse III

Gilbert CASAMATTA

Président de l'Université de Toulouse

N° étudiant : 20107299



Gilles FOURTANIER

*Ce document doit être joint au dossier transmis à
l'intention du comité de sélection*

DÉCLARATION DE CANDIDATURE AU RECRUTEMENT SUR UN EMPLOI DE MAÎTRE DE CONFÉRENCES

(Campagne 2012)

(décret n° 84-431 du 6 juin 1984 modifié)

Authentification : 4958be9c09abd8c77cde46d7ffba4bcc (1331743806498)

adressée au chef d'établissement de : UNIVERSITE PARIS 7 (DENIS DIDEROT)

Poste(s) : n°4073

Publié le : 23 février 2012 à 00:02

Section(s) C.N.U : 26 (Mathématiques appliquées et applications des mathématiques) et 25 (Mathématiques)

Profil : Statistiques du risque et Fouilles de données

Localisation : PARIS

Article 26-I-1

Chaire : Non

Je soussigné(e) M.

Nom patronymique : DE CASTRO

Nom d'usage ou marital :

Prénom : YOHANN

Date et lieu de naissance : 20/12/1983 - MURET

Nationalité : Française

N° de qualification : 12226224405

Adresse postale et électronique à laquelle seront acheminées toutes les correspondances		
25 AVENUE DE LA GARONNETTE APT 14		
Code postal : 31000	Ville : TOULOUSE	Pays : FRANCE
Téléphone : 0680161668	Télécopie :	
Adresse électronique : yohann.decastro@gmail.com		

Fonctions et établissement actuel : Allocataire-moniteur-UNIVERSITE PAUL SABATIER - TOULOUSE III

Date de création :

28/02/2012 à 21:02

Date de dernière modification :

28/02/2012 à 21:02

Titres universitaires français :

Diplôme le plus récent : Thèse

Titre : Constructions déterministes pour la régression parcimonieuse

Date de soutenance : 03/12/2011

Lieu de la soutenance : UNIVERSITE PAUL SABATIER - TOULOUSE III

Mention : Mathématiques

Directeur : JEAN-MARC AZAIS ET FRANCK BARTHE

Composition du jury : GABOR LUGOSI (ABSENT)

ALEXANDRE TSYBAKOV

GUILLAUME LECUE

JEAN-MARC AZAIS

FRANCK BARTHE

Activités en matière d'enseignement :

Pré 2009 : 3 ans Khôlleur en MP* (Lycées Charlemagne, Paris puis P. de Fermat, Toulouse).

2009 - 2011 : Cours-TD en L2 Biologie (Statistique et Probabilité).

Cours sur le modèle parcimonieux à Montevideo.

2011-2012 : TD en M1 IMAT (Master 1 Ingénieur) (Transformée de Fourier, filtre de Wiener, ...), 2 TP en M2 IMAT (Master 2 Ingénieur) "Plans d'expériences" et "Ingénierie et imagerie médicale".

thème de recherche et mots clés :

Thèmes de recherche : Statistiques en grandes dimensions, Régression parcimonieuse, Compressive sensing, Processus empiriques, Matrices aléatoires.

Mots clés: "Compressed Sensing", Lasso, Modèle linéaire, "Derandomization", Largeur de Gelfand, Graphes Expanseurs, Polytopes convexes.

Activités en matière d'administration et autres responsabilités collectives :

Co-organisateur des Rencontres des Doctorants en Statistique de l'IMT. Co-organisateur du séminaire étudiant de l'IMT.

Co-organisateur d'un groupe de travail sur la complétion de matrice.

Autres titres et diplômes :

Travaux, ouvrages, articles, réalisations :

Liste des pièces à fournir par le candidat :

pièces obligatoires

déclare faire acte de candidature sur l'emploi ci-dessus désigné :

Fait à Toulouse le 14/03/2012

Signature

