Présentation du cours "Introduction mathématique au *Compressed Sensing*"

Guillaume Lecué
ENSAE – CREST
Bureau E31
3 avenue Pierre Larousse
92245 Malakoff France

Tel: +33 (0) 1 41 17 64 22
`guillaume.lecue@ensae.fr`
`http://lecueguillaume.github.io/`

L'objectif de ce cours est d'étudier des problèmes de statistiques en grandes dimensions afin de dégager trois idées fondamentales de cette thématique qui pourront par la suite être appliquées dans de nombreux autres problèmes liés aux sciences des données. Nous enonçons briévement ces principes :

1. Un grand nombre de données réelles appartiennent à des espaces de grandes dimensions dans lesquels les méthodes classiques de statistiques sont inefficaces (cf. *fléau de la dimension*). Néanmoins ces données sont pour la plupart d'entre elles structurées. Si bien que la "vraie" dimension du problème n'est plus celle de l'espace ambiant mais plutôt celle de la structure qui contient l'information utile des données. On parle de **données structurées ou parcimonieuses**. La construction de bases ou dictionnaires permettant de révéler les structures de faible dimension de ces données est une composante importante de la statistique en grande dimension.
2. En première approche, la recherche de ces structures de faible dimension semble nécessiter le lancement d'une recherche combinatoire dans un espace de grande dimension. De telles procédures ne peuvent pas être utilisées en pratique. Une composante importante de la statistique en grande dimension est alors de proposer et d'analyser des algorithmes qui peuvent être implémentés même dans des espaces de grande dimension. Pour cela, deux approches ont reçu une attention particulière : la **relaxation convexe** (couplé à la boîte à outils de l'optimisation convexe) et les **algorithmes itératifs** qui permettent de résoudre parfois des problèmes d'optimisation non-convexe.
3. Finalement, la troisième composante est le rôle joué par **l'aléatoire** dans la statistique en grande dimension. Il s'avère que les structures de faibles dimensions sont généralement révélées par des objets aléatoires et que, jusqu'à maintenant, on ne sait pas exhiber ces structures à l'aide de mesures déterministes aussi efficacement que le font, par exemple, les matrices aléatoires.

Un cours de statistiques en grande dimension peut donc couvrir plusieurs pans des mathématiques dont la théorie de l'approximation, l'optimisation convexe et les probabilités. Dans ce cours, nous étudierons principalement l'aspect algorithmique et probabiliste de cette théorie. La théorie de l'approximation ne sera que très brièvement abordée au travers de l'exemple des images.

Ce cours abordera le paradigme de la statistique en grande dimension principalement autour de trois thématiques :

1. **Compressed sensing**: problème de reconstruction exacte et approchée d'un signal de grande dimension à partir d'un petit nombre de mesures linéaires de ce vecteur sachant qu'il a un petit support;
2. **complétion de matrice / système de recommandation**: comment compléter une matrice à partir de l'observation d'un petit nombre de ses entrées sachant que cette matrice est de faible rang;
3. **détection de communauté dans les graphes**: trouver les sous-graphes de forte densité dans des 'grands' graphes.

Nous abordons donc le problème de la statistique en grande dimension au travers de trois objets/ types de données clefs pour la science des données : les vecteurs de grande dimension mais parcimonieux, les matrices de grande taille mais de faible rang et finalement, les graphes de 'grande' taille dont les noeuds sont organisés en communautés.

Le problème de Compressed Sensing sera utilisé comme le principale vecteur pédagogique pour l'apprentissage des trois idées clefs de la statistique en grandes dimensions mentionnés précédemment. On y consacrera donc 8 séances divisées comme suit : 5 (ou 4) séances de cours, 2 séances d'exercices et 1 (ou 2) séances de pratiques informatiques (le temps consacrait en TP dépendra du goût des élèves au fil du cours). Puis nous consacrerons les 4 dernières séances aux problèmes de complétion de matrices et de détection de communautés: 1 séance de cours/exercices et 1 séance d'informatique pour chacune des deux thématiques.

D'un point de vue de la *technique mathématiques* nous mettrons l'accent sur les thématiques suivantes :

1. concentration de variables aléatoires et calcul de complexité;
2. méthodes et analyse d'algorithmes en optimisation convexe.

Les séances de travaux pratiques informatiques s'effectueront avec le langages **Python**. On mettra particulièrement l'accent sur les librairies classiques en science des données: **sklearn**, **cvxopt** et **networkx**.

Cours 1 : Introduction au Compressed Sensing et motivations.

Introduction aux idées de parcimonie et de mesures linéaires aléatoires au travers d'exemples concrêts:

1. Imagerie par résonance magnétique
2. Single pixel Camera
3. Codes correcteurs d'erreurs.
4. Spectroscopie de masse et notion de parcimonie avancées.

Mise en avant de l'idée de développement parcimonieux d'un signal dans une base adaptée. Présentation de l'aspect algorithmique de la statistique en grande dimension:

1. minimisation ℓ_0 et la notion de "NP-Hardness"
2. relaxation convexe et l'algorithme du Basis Pursuit,
3. algorithmes itératifs (*Matching pursuit*, *reweighted basis pursuit*, etc.)

Borne de Shannon-Nyquist.

Cours 2 : Condition RIP pour les matrices Sous-Gaussiennes.

On introduira les notion suivantes:

1. *Restricted isometry property*
2. Sections Euclidiennes de la boule unité B_1^p
3. la *null-space property*
4. propriété de reconstruction exacte des vecteurs sparses par le Basis Pursuit.

L'accent sera mis sur le rôle de l'aléa dans cette théorie au travers des mesures et matrices aléatoires. Des liens avec la géométrie des espaces de Banach en grandes dimensions seront mis en avant lors des séances de TD (Johnson/Lindentsrauss, Gelfand, Kashin et Dvoretski).

Sur l'aspect techniques mathématiques, on présentera les outils suivants:

1. l'inégalité de Bernstein pour les variables sous-exponentielles,
2. l'argument volumique pour le calcul d'entropie,
3. l'argument d'ε-réseau sur la boule B_2^p.

On montrera la condition RIP pour les matrices sous-gaussiennes. On montrera un résultat d'optimalité sur le nombre de mesures concernant la méthode de Basis Pursuit.

Si le temps le permet, on introduira la méthode dite *de petite boule*.

Cours 3 (TD) : séance d'exercices autour de la concentration et de la complexité

L'objectif de cette séance d'exercices est de familiariser les élèves aux matrices aléatoires, aux inégalités de concentration et au calcul de complexité. L'accent sera mis sur le compromis classique en statistiques : concentration/complexité grâce aux exercices suivant :

1. ε-argument pour $B_1^p \cap rB_2^p$.
2. calcul du nombre d'entropie par la méthode de Maurey.
3. inégalité de Dudley (argument de chaînage)
4. inégalité de Sudakov et calcul de nombres d'entropies.
5. Le théorème de Gordon "Escape trough the mesh"
6. Le théorème de Kashin (cf. [2])
7. Théorème de Johnson-Lindenstrauss.
8. Fenêtres de Gelfand de B_r^p, $0 < r \leq 1$.
9. Introduction au théorème de Dvoretsky et à l'inégalité de Pajor-Tomcjak.
10. "smallest singular value of a rectangular sub-gaussian matrix" selon Rudelson.

Cours 4 : Algorithmes classiques en compressed sensing

Réécriture du Basis Pursuit comme un problème de programmation linéaire. Méthode de résolution de programmation linéaire.

Introduction aux méthodes de descente de gradient et aux méthodes proximales. Méthode de Douglas-Rachford.

Cours 5 (TD) : séance d'exercices sur l'optimisation convexe

Cette séance d'exercice sera consacrée à des problèmes d'optimisation convexe.

Cours 6 : Stabilité et robustesse

Présentation des problème de stabilité (à la parcimonie) et robustesse (au bruit). Etude de la stabilité et robustesse des algorithmes vus jusqu'ici (BP, OMP, etc.). Présentation du LASSO, du Dantzig et du BPDN.

Cours 7 : Mesures structurées et ouvertures sur d'autres problèmes

On présentera des résultats de reconstruction exacte pour des mesures de Fourier, temps/fréquence et circulaires.

Pour des raisons de difficultés techniques, plusieurs résultats ne seront que mentionnés en cours mais les élèves seront renvoyés aux démonstrations des notes de cours :

1. On mentionnera que la condition RIP pour les mesures sous-exponentielles est sous-optimale. On montrera en revanche que la *Null-space property* est satisfaite avec le nombre minimal de mesures pour ces matrices aléatoires sous-exponentielles (toujours en renvoyant aux notes de cours).
2. On mentionnera la preuve de RIP pour les matrices de Fourier extraites aléatoires.
3. On introduira le problème $\Lambda(1)$,

.

Ouverture sur d'autres sujets:

1. Introduction aux problèmes suivants: codes correcteurs d'erreurs, "phase recovery", séparation de sources (par ACP robuste ou factorisation de matrice), "one bit Compressed sensing" (lien avec le *deep learning*).
2. Généralisation à d'autres problèmes (cf. [8])
3. Aspect géométriques (cône, sous-gradient, dimension statistique) et parcimonie (cf. [1] or [6])

Note : selon le goût des élèves pour la théorie ou la pratique, on peut envisager de remplacer le cours 7 par une séance de TP en semaine 6.

Cours 8 (TP) : Implémentation et test des algorithmes

Implémentation du basis pursuit, matching pursuit et de la version itérative du weighted Basis Pursuit. Introduction aux libraires python : *sklearn* et *cvxopt*.

On construira le diagramme de transition de phase pour chaque algorithme pour différentes matrices de mesures. On fera aussi des tests de robustesse et de stabilité.

Si le temps le permet on étudiera l'implémentation du LARS pour le lASSO.

Cours 9 : Complétion de matrices et systèmes de recommandation

Introduction de la problématique des systèmes de recommandation grâce au prix de *Netflix*. Rappels sur les procédures classique (item-based et user-based). Présentation d'algorithmes plus évolués:

1. problème de minimisation de la norme nucléaire,
2. factorisation de matrices: NMF
3. introduction au deep learning avec la *restricted Boltzman machine* (cf. [7]).
4. *Alternating least square* pour la complétion de matrices (selon [4]).

Cours 10 (TP) : Complétion de matrices et systèmes de recommandation

Construction de systèmes de recommandation grâce aux librairies *sklearn* (ou *nimfa* ou *graphlab*) sur la base de donnée *movielens* et sur des images. Test efficacitéé.

Cours 11 : Détection de communautés dans les graphes

Introduction aux graphes. Présentation de quelques algorithmes de détection de communautés. Relaxation convexe et théorème de Grothendieck (cf. [3])

Cours 12 (TP) : Détection de communautés dans les graphes

Utilisation de la librairie *networkx*. Utilisation de l'algorithme de Louvain et de la librairie *igraph* (sous R si lŠinstallation sous python est trop difficile). Visualisation avec *gephi* du graphe facebook des élèves avec *graphviz* (si API encore disponible). Construction de modèle aléatoire de graphes (par exemple, le "stochastic block model").

Quelques exemples d'exercices

1. L'algorithme de minimisation ℓ_0 et le problème de partitionnement par des ensembles de cardinal 3.
2. Vérifier RIP est NP-hard.
3. borne de Welsh et *frames* equi-angulaires.
4. exercices sur le *fléau de la grande dimension*.
5. L'équivalence entre Johnson-Lindenstrauss et RIP par [5].
6. théorème de Kashin par [2].
7. Les matrices de 0 et de 1 ne peuvent pas vérifier RIP pour le nombre optimal de mesures.
8. Calcul d'entropie dans l'espace des matrices.

9. moment and tails
10. non-commutative khintchin inequality for fourier

Quelques exemple de projets scientifiques

1. Principes d'incertitude discrets.
2. Construction de matrices déterministes pour le Compressed Sensing. Gershyrin result cannot do that s^2.
3. deterministic mat + random signal.
4. L'algorithme de *de Prony*.
5. Introduction aux *expanders*. RIP-1. Indyk.
6. Méthode de minimisation ℓ_p^n pour $0 < p < 1$.
7. Le *one-bit compressed sensing*.
8. Rudelson paper about the smallest singular value of sign matrix.
9. Kramher and ward RIP - JL
10. Plan and Verhsynin logistic
11. golfing scheme

Bibliography

[1] Venkat Chandrasekaran, Benjamin Recht, Pablo A. Parrilo, and Alan S. Willsky. The convex geometry of linear inverse problems. *Found. Comput. Math.*, 12(6):805–849, 2012.

[2] Simon Foucart. A simple proof of kashinâĂŹs decomposition theorem. Technical report, Drexel University, 2012.

[3] Olivier Guédon and Roman Vershynin. Community detection in sparse networks via grothendieck's inequality. Technical report, 2014.

[4] Trevor Hastie, Rahul Mazumder, Jason D. Lee, and Reza Zadeh. Matrix completion and low-rank svd via fast alternating least squares. Technical report, Statistics Department and ICME Stanford University, 2014.

[5] Felix Krahmer and Rachel Ward. New and improved Johnson-Lindenstrauss embeddings via the restricted isometry property. *SIAM J. Math. Anal.*, 43(3):1269–1281, 2011.

[6] Emile Richard, Guillaume Obozinski, and Jean-Philippe Vert. Tight convex relaxations for sparse matrix factorization. Technical report, 2012.

[7] Ruslan Salakhutdinov, Andriy Mnih, and Geoffrey Hinton. Restricted boltzmann machines for collaborative filtering. Technical report, Toronto University, 2010.

[8] Joel A. Tropp. Convex recovery of a structured signal from independent random linear measurements. Technical report, To appear in Sampling Theory, a Renaissance., 2014.