

Sujet de thèse
École doctorale EEA de Lyon

Merci de compléter l'ensemble des rubriques et de lire les notes de bas de page.

Etablissement d'inscription : Ecole Centrale de Lyon ¹
École doctorale : ED 160 EEA de Lyon dirigée par Mr Delachartre Philippe
Intitulé du doctorat : Automatique ²
Sujet de la thèse : Performance assessment of a class of robust estimators via measure concentration inequalities
Unité de recherche : Ampère ³ , dirigée par Christian Vollaire
Directeur/trice de thèse : Mr BAKO Laurent
Co-directeur/trice de thèse (le cas échéant)⁴ :
Co-directeur/trice de thèse en entreprise (le cas échéant) :

¹ A impérativement choisir dans la liste suivante : Ecole Centrale de Lyon, INSA de Lyon, Université Claude Bernard Lyon 1

² A impérativement choisir dans la liste suivante : Automatique // Electronique, Micro et Nano-électronique, Optique et Laser // Génie Electrique // Ingénierie pour le vivant Traitement du signal et de l'Image)

³ A impérativement choisir dans la liste suivante : Laboratoire Ampère, CITI, CREATIS, INL, LAGEP, LGEF

⁴ Un/une co-encadrant-e n'est pas nécessairement co-directeur/trice de thèse puisque pour remplir ce rôle, il est nécessaire d'être habilité à diriger des recherches (pour plus de précision, voir le règlement intérieur de l'ED EEA, section 3.

Collaboration(s)/partenariat(s) extérieur(s) éventuels⁵ :

Domaine et contexte scientifiques :

In many engineering fields such as control system design, signal processing, or machine learning, one is frequently confronted with the necessity of either estimating a set of signals (e.g., the state of a dynamic system) or inferring a mathematical relationship between a number of signals of interest (e.g., the model of a dynamic system). In both situations, the problem is to design a possibly set-valued map, called estimator, which associates the available dataset to a subset of the space of possible values of the to-be-estimated quantity. Therefore, the performance of the estimator is naturally measured in terms of its capability to return a set of estimates which are as close as possible, in the sense of an appropriate metric, to the true (but unknown) quantity of interest. While asymptotic properties are traditionally studied (e.g. in system identification [7]), we note that in practice, only a finite (and sometimes small) number of data samples are available for constructing an estimate. This observation calls for finite-sample analysis similar to the one proposed in [8,9].

Mots-clefs : estimation theory, optimization, recursive algorithms, performance assessment, measure concentration inequalities

Objectifs de la thèse :

The goal of this research project is to analyze the properties of a class of robust optimization-based estimators operating on a finite number of data samples, for example the type of estimators discussed in [1,2]. More precisely, the focus is to develop some relevant metrics for assessing the performance of such finite-sample estimators in a stochastic setting. Ideally, one would like to derive the full distribution of the estimation error conditioned on the input data. In effect, given such a distribution it becomes possible to construct uncertainty sets of any level of confidence. However, this quest is almost hopeless for many non trivial estimators of practical interest. For this reason, most of existing performance measurement criteria are essentially reduced to first and second order statistics such as the mean and the variance of the error. As global measures however, these statistics taken alone do not meet the need for finer characterization of the estimation error required for use in modern decision-making systems. In this work, we might conveniently view the estimator as a system (an operator, similarly as in control theory) which would be static if the estimator operates in a batch mode, or dynamic if instead, it processes the data sample after sample, in an adaptive fashion. In line with this view, the analysis will consist in characterizing the salient intrinsic properties of the

⁵ Hors contrats doctoraux fléchés UMI par l'établissement, les sujets de thèse en cotutelle ne sont pas acceptés.

estimator in terms of robustness, consistency, generalization capacity of the finite-sample based estimate.

Verrous scientifiques :

There are three main challenges to be addressed in this research program:

- First, finite-sample guarantees are generally harder to assess than asymptotic properties. An intuitive reason for that is that fundamental theorems of statistics such as those of the Central Limit (CLT), Law of Large Numbers (LLN), etc, have a limited significance for finite-sample scenarios.
- Second, optimization-based estimators have a non trivial structure, that is, the estimates outputed by the estimator may not be explicitly expressible as a function of the data. This increases considerably the challenge related to mathematical analysis. It might also be a source of significant conservatism. The analysis method should be able to overtake somehow this limitation.
- Finally, performance assessment itself relies on an estimation (bounding) process and will consist in deriving formally the least conservative error bounds possibles in terms of measure concentration inequalities. Computing tight certificates of performance can also involve numerical optimization.

Contributions originales attendues :

By taking inspiration from some recent contributions to the topic of robust estimation (the so-called distributionally robust approach) [3,4,5], a first interesting result could be to formalize a general estimation framework that would admit most of the classical ones as its special cases. After that the focus may be the fundamental analysis (of the derived general framework) whose goal is to achieve performance characterization in terms of measure concentration inequalities [6]. To sum up, original contributions are expected in the form, for example of :

- A generalized optimization-based robust estimation framework in the sense of insensitivity to a moderate number of outliers and small variation of the estimates as a consequence of dense noise
- Analysis results within this framework characterizing properties of resilience/robustness, stability, consistency, out-of-sample performance in finite time,

- Substantial discussion on the possible exploitation of the analysis results for experiment design and selection of the design parameters for the estimator itself so as to enhance its performances.

Programme de recherche et démarche scientifique proposée :

The research can be conducted along the following plan:

- Step 1 (4 months) : Bibliographic study. A natural starting point is to conduct a bibliographic search on robust statistical estimation, performance assessment metrics for estimators and analysis methods, in particular those techniques dedicated to finite sample estimators. This step will allow the student to get familiar with estimation theory and to the technical tools that will be necessary for carrying out a deep mathematical analysis.
- Step 2 (4 months) : Setting up relevant metrics for the analysis. It is possible to go from some general purpose performance metrics to more specific application-dependent ones. Since an estimate is designed for a specific engineering goal (control, prediction, filtering, decision-making, etc), it might be of interest to provide some performance guarantees (up to some confidence level) with respect to the intended use of the estimate.
- Step 3 (16 months): Proper analysis work. This part is the heart of the PhD thesis and deserves more time and effort. First publishable results might be expected at the end of the first year or the beginning of the second year.
- Step 4 (6 months): Application of the analysis results. Part of the work should be devoted to investigating how the obtained (general) analysis results can be specialized to different specific estimation problems of interest in engineering and to discussing how those results can have an impact in practice, e.g. in experiment design.
- Step 5 (6 months): Writing of thesis of dissertation.

Encadrement scientifique :

- **Description du comité d'encadrement :** [à compléter avec le rôle dans l'encadrement scientifique (en termes de compétences scientifiques, etc.) et le pourcentage d'implication du directeur de thèse ⁶ et des autres membres du comité⁷]

Nom Prénom	Labo / Equipe	Compétences scientifiques	Taux d'encadrement %
Mr BAKO Laurent	Ampère Conception Intégrée de Systèmes Dynamiques/Automatique pour l'Ingénierie des Systèmes (AIS)	Estimation, optimisation, identification de systèmes,	50
Mr HUILLERY Julien	Convergence Energie-Information/Automatique pour l'Ingénierie des Systèmes	Traitement du Signal, Estimation	50

- Le comité d'évaluation de l'HCERES ayant demandé à l'école doctorale de limiter la taille du comité d'encadrement à deux membres (directeur de thèse compris), il est impératif de ne proposer des comités d'encadrement de taille plus importante que si cela est absolument nécessaire⁸ et **de le justifier soigneusement.**
- **Intégration au sein du (ou des) laboratoire(s) (Département/Equipe(s) impliquée(s)) (pourcentage du temps travail au sein de ce ou ces laboratoire(s)) :**

⁶ Le directeur de thèse doit être un HdR rattaché à l'ED EEA ou en passe de le devenir avant juin de l'année en cours ou bénéficier d'une dérogation du Conseil Scientifique lors du dépôt du sujet de thèse.

⁷ Dans le cas d'un comité d'encadrement réparti sur plusieurs établissements, la plus grande partie de l'encadrement est effectuée par des membres de l'établissement. Si l'encadrement de la thèse implique des membres hors de l'ED EEA, la part de l'encadrement des membres ED doit être très supérieure à 50%.

⁸ Un certain nombre de commissions type CNU ne reconnaissent un co-encadrement qu'au-delà d'un certain pourcentage. Souvent l'encadrement est considéré comme effectif si > 30%.



EEA
ÉLECTRONIQUE
ÉLECTROTECHNIQUE
ET AUTOMATIQUE
UNIVERSITÉ DE LYON



Le doctorant rejoindra le département Automatique pour l'Ingénierie des Systèmes (AIS) et la priorité Conception Intégrée des Systèmes Dynamiques du laboratoire Ampère. Il travaillera sur le site de l'Ecole Centrale de Lyon.

Financement de la thèse : Contrat doctoral de l'établissement d'inscription

Profil du candidat recherché (prérequis) :

Ingénieur ou titulaire d'un master à dominante Automatique avec des compétences en Probabilités et statistiques, Optimisation et Traitement du signal. Le candidat doit également posséder une bonne aisance en programmation (Matlab) et une bonne maîtrise de l'anglais.

Objectifs de valorisation des travaux de recherche :

Ces travaux relevant plutôt de recherche fondamentale, la valorisation se fera au travers essentiellement de publications dans des revues et des conférences nationales et internationales.

Compétences qui seront développées au cours du doctorat :

Le candidat recruté développera des compétences pointues en théorie de l'estimation et en optimisation mathématique. D'autre part il/elle aura l'opportunité de se former au métier d'enseignant-chercheur durant sa thèse à travers des activités d'enseignement comme des travaux dirigés et des travaux pratiques.

Perspectives professionnelles après le doctorat :

Ce sujet de thèse a une forte composante théorique. Les perspectives professionnelles après le doctorat sont idéalement, un poste de chercheur ou d'enseignant-chercheur dans des domaines comme l'Automatique, le traitement du signal ou les Mathématiques Appliquées. Mais des possibilités de poursuite de carrière en industrie existent et sont tout à fait nombreuses, notamment dans le domaine de l'intelligence artificielle ou de l'apprentissage statistique qui recrute de plus en plus des profils pointus de théoriciens.

Références bibliographiques sur le sujet de thèse :

[1] L. Bako. On a Class of Optimization-Based Robust Estimators. IEEE Transactions on Automatic Control, vol. 62, pp. 5990-5997, 2017.

[2] E. Candès and P. A. Randall, "Highly robust error correction by convex programming," IEEE Trans. Inf. Theory, vol. 54, no. 7, pp. 2829–2840, 2006.

[3] P. M. Esfahani and D. Kuhn. Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations, Mathematical Programming, vol. 171, pp 115-166, 2018.

- [4] R. Chen and I. Ch. Paschalidis. A Robust Learning Approach for Regression Models Based on Distributionally Robust Optimization, *Journal of Machine Learning Research*, vol. 19, pp 1-48, 2018.
- [5] S. Shafieezadeh-Abadeh, V. A. Nguyen, D. Kuhn, P. M. Esfahani. Wasserstein Distributionally Robust Kalman Filtering. *Neural Information Processing Systems (NIPS)*, Montreal, Canada, 2018.
- [6] S. Boucheron, G. Lugosi, P. Massart. Concentration inequalities : a nonasymptotic theory of independence. OUP Oxford Ed., 2013.
- [7] L. Ljung. System identification : theory for the user. Prentice Hall, 2nd Ed., 1999.
- [8] S.N. Negahban, P. Ravikumar, M.J. Wainwright, and B. Yu. A Unified Framework for High-Dimensional Analysis of M-Estimators with Decomposable Regularizers. *Statistical Science*, vol. 27, pp 538-557, 2012.
- [9] H. Ha, J. S. Welsh, C. R. Rojas, B. Wahlberg. An analysis of the SPARSEVA estimate for the finite sample data case. *Automatica*, vol 96, pp 141-149, 2018.