

# Sur les bornes minimales pour l'estimation de paramètres déterministes

Philippe FORSTER<sup>1</sup> et Pascal LARZABAL<sup>2,3</sup>

<sup>1</sup>GEA, Pôle scientifique et technologique de Paris 10, 1 chemin Desvallières, 92410 Ville d'Avray, 01-47-09-45-20

philippe.forster@cva.u-paris10.fr

<sup>2</sup>LESIR/ENS Cachan 61 avenue du président Wilson 94235 Cachan Cedex, 01-47-40-27-09 larzabal @ lesir.ens-cachan.fr

<sup>3</sup>CRHIP, IUT de Cachan, Université Paris Sud, 9 avenue de la division Leclerc, BP134, 94234 Cachan Cedex, 01-41-24-11-56

*Abstract*— Nous avons revisité et résolu le problème de recherche de bornes d'estimation sous l'angle d'un problème contraint de minimisation d'erreur. Nous montrons que les différentes bornes d'estimation de paramètres déterministes (Cramer-Rao, Barankin, Battacharyya) s'obtiennent aisément comme le résultat d'une optimisation en imposant le biais de l'estimateur.

de ces bornes.

Le but de cet article est de présenter une approche unifiée de l'obtention des bornes sous la forme d'un problème d'optimisation sous contraintes. Cette approche permet de développer dans certains cas l'estimateur contraint optimal.

## I. INTRODUCTION

### A. Insuffisance de la Borne de Cramer Rao (BCR)

L'étude des performances statistiques d'un estimateur combine généralement 3 sources d'information : une borne inférieure de la variance, une évaluation théorique de la variance associée à l'estimateur, et des résultats de simulation de type Monte Carlo. La connaissance d'une borne inférieure est d'une importance primordiale. Elle permet de savoir si dans un contexte donné il est possible de respecter certaines exigences en performances définies par un cahier des charges. Elle permet aussi de savoir si, à partir d'un algorithme particulier, une amélioration est possible. Dans les problèmes d'estimation de paramètres déterministes, la borne minimale qui est le plus souvent retenue est la borne de Cramer Rao [1]. Elle présente 2 avantages :

- un calcul simple;
- sous des conditions peu restrictives elle est atteinte par la méthode du MV asymptotiquement en :
  - le nombre d'observations indépendantes ;
  - le RSB dans le cas d'un signal déterministe noyé dans un bruit gaussien.

Toutefois, lorsque les conditions d'estimation se durcissent (faible RSB, faible nombre d'échantillons,...), elle s'avère trop optimiste. De nombreux chercheurs ont proposé de meilleures bornes, au sens de minorants plus fins de la variance sur toute la plage d'évolution du scénario d'estimation (nombre d'échantillons, RSB, corrélation,...). On peut citer les travaux de Battacharyya [2], puis ceux de Barankin [3]. Ces derniers travaux englobent tous les précédents. Ces différentes bornes, bien qu'utilisées par certains chercheurs (ex [4]), ne connaissent pas la popularité de la BCR dans la communauté des traiteurs de signaux. Ceci s'explique vraisemblablement par :

- une complexité numérique jugée importante dans les travaux originaux,
- une méconnaissance et une incompréhension des phénomènes et hypothèses mis en jeu lors de l'établissement

### B. Approche unifiée pour l'obtention des bornes d'estimation

Pour simplifier la présentation de nos travaux nous allons traiter le cas de l'estimation d'un paramètre scalaire. Ceci ne nuit en rien à la généralité des résultats présentés. Soit  $\mathbf{x}$  une variable aléatoire vectorielle continue à valeurs dans  $\mathfrak{R}^N$  dont la densité de probabilité est  $p(\mathbf{x}; \theta)$  où  $\theta$  est un paramètre inconnu déterministe appartenant à l'intervalle  $[a, b]$ : les résultats qui suivent se généralisent sans problème au cas de paramètres vectoriels. Soit  $F$  l'ensemble des estimateurs  $\hat{\theta}(\mathbf{x})$  de  $\theta$ . Le problème est d'établir la valeur minimale que peut prendre l'*EQM* (Erreur Quadratique Moyenne) des estimateurs appartenant à  $F$  en un point  $\theta_0 \in [a, b]$ :

$$EQM = \int (\hat{\theta}(\mathbf{x}) - \theta_0)^2 p(\mathbf{x}; \theta_0) d\mathbf{x} . \quad (1)$$

Les estimateurs appartenant à  $F$  doivent posséder certaines propriétés jugées désirables. La propriété la plus généralement attendue est un biais identiquement nul sur tout l'intervalle  $[a, b]$ , ou de façon moins exigeante, sur un sous-ensemble  $\Theta \subset [a, b]$  :

$$F = \left\{ \hat{\theta}(\mathbf{x}) \text{ tel que } \int \hat{\theta}(\mathbf{x}) p(\mathbf{x}; \theta) d\mathbf{x} = \theta \quad \forall \theta \in \Theta \right\} . \quad (2)$$

La minimisation de l'*EQM* pour  $\hat{\theta}(\mathbf{x}) \in F$  est un problème d'optimisation quadratique sous des contraintes linéaires, qui est résolu dans ce papier par des techniques de calcul variationnel.

Dans le cas le plus exigeant où  $\Theta = [a, b]$ , la résolution conduit à une équation intégrale dont la solution fournit la borne minimale la plus fine sur la variance en  $\theta_0$  des estimateurs non biaisés sur  $[a, b]$ . Certains cas peuvent être résolus de manière analytique [4], et d'autres nécessitent une résolution numérique. Une telle procédure fournit, par ailleurs, l'estimateur optimal contraint qui dans certains cas ne dépend pas de  $\theta_0$  présentant ainsi un intérêt pratique [4].

Les autres bornes classiques rentrent dans ce cadre par un choix approprié de  $\Theta$  se traduisant par un allègement des contraintes qui permet d'aboutir à une expression analytique de la borne.

- $\Theta = \{\theta_0, \dots, \theta_K\}$  : cela revient à annuler le biais seulement en certains points de  $[a, b]$ . Le calcul de la borne correspondante est simple et conduit aux résultats de Barankin [3].

- $\Theta = \{\theta_0, \theta_0 + \epsilon\}$  : si  $\epsilon \rightarrow 0$ , le biais et sa dérivée sont nuls en  $\theta_0$ . On retrouve là les hypothèses sous lesquelles a été développée la borne de Cramer Rao.

- $\Theta = \{\theta_0, \theta_0 + \epsilon, \dots, \theta_0 + K\epsilon\}$  : si  $\epsilon \rightarrow 0$ , le biais et ses  $K$  dérivées successives sont nuls en  $\theta_0$ . On retrouve les hypothèses sous lesquelles ont été développées les bornes de Battacharyya [2] dont la borne de Cramer Rao est un cas particulier ( $K = 1$ ). Cela revient à calculer le minimum de l'EQM sur la famille d'estimateurs :

$$F = \{\hat{\theta}(\mathbf{x}) \text{ tel que } \int \hat{\theta}(\mathbf{x}) p(\mathbf{x}; \theta_0) d\mathbf{x} = \theta_0 \text{ et} \\ \int \hat{\theta}(\mathbf{x}) \left( \frac{\partial^k}{\partial \theta^k} p(\mathbf{x}; \theta) \right)_{\theta_0} d\mathbf{x} = \delta_{k1} \text{ pour } 1 \leq k \leq K\}. \quad (3)$$

Finalement ces bornes s'obtiennent comme des cas particuliers de la minimisation de l'EQM en  $\theta_0 \in [a, b]$  pour des estimateurs satisfaisant un jeu de contraintes linéaires défini par

$$F = \{\hat{\theta}(\mathbf{x}) \text{ tel que } \int \hat{\theta}(\mathbf{x}) g_i(\mathbf{x}; \theta) d\mathbf{x} = h_i(\theta) \\ \forall \theta \in \Theta, i = 0, \dots, K\} \quad (4)$$

où  $g$  et  $h$  sont des fonctions dépendant des contraintes retenues.

Le papier est organisé comme suit: dans le paragraphe II nous exposons la solution la plus exigeante du point de vue des contraintes. Le paragraphe III donne la solution générale du problème allégé correspondant à des contraintes discrétisées. Dans le paragraphe IV cette solution est particularisée aux contraintes correspondant aux bornes classiques.

## II. RECHERCHE DE L'ESTIMATEUR À VARIANCE MINIMALE EN $\theta_0$ ET À BIAIS IDENTIQUEMENT NUL

Un tel estimateur optimal minimise l'EQM (1) sous le continuum de contraintes suivant :

$$\int \hat{\theta}(\mathbf{x}) p(\mathbf{x}; \theta) d\mathbf{x} = \theta \quad \forall \theta \in [a, b]. \quad (5)$$

Les travaux de Barankin ont déjà apporté une solution à ce problème [3]. Nous proposons dans le théorème 1 une alternative dont l'intérêt est de fournir dans certains cas une solution analytique.

*Theorem 1:* L'estimateur  $\hat{\theta}(\mathbf{x})$  à variance minimale en  $\theta_0$  et à biais identiquement nul sur  $[a, b]$  est donné par

$$\hat{\theta}(\mathbf{x}) = \theta_0 + \int \mu(\theta) \frac{p(\mathbf{x}; \theta)}{p(\mathbf{x}; \theta_0)} d\theta \quad (6)$$

où  $\mu(\theta)$  est solution de l'équation intégrale:

$$\int \mu(\theta') K(\theta, \theta') d\theta' = \theta - \theta_0 \text{ avec} \\ K(\theta, \theta') = \int \frac{p(\mathbf{x}; \theta) p(\mathbf{x}; \theta')}{p(\mathbf{x}; \theta_0)} d\mathbf{x}. \quad (7)$$

*Proof:* En utilisant les multiplicateurs de Lagrange, la

solution s'obtient en minimisant la fonction coût suivante

$$C(\hat{\theta}(\mathbf{x}), \lambda(\theta)) = \int (\hat{\theta}(\mathbf{x}) - \theta_0)^2 p(\mathbf{x}; \theta_0) d\mathbf{x} + \\ \int \lambda(\theta) \left[ \int \hat{\theta}(\mathbf{x}) p(\mathbf{x}; \theta) d\mathbf{x} - \theta \right] d\theta. \quad (8)$$

En définissant  $f(\epsilon) = C(\hat{\theta}(\mathbf{x}) + \epsilon h(\mathbf{x}), \lambda(\theta))$ ,  $f(\epsilon)$  doit être minimale pour  $\epsilon = 0$ , et ce,  $\forall h(\mathbf{x})$ . L'annulation de la dérivée de  $f(\epsilon)$  en 0  $\forall h(\mathbf{x})$  conduit à:

$$2(\hat{\theta}(\mathbf{x}) - \theta_0) p(\mathbf{x}; \theta_0) + \int \lambda(\theta) p(\mathbf{x}; \theta) d\theta = 0. \quad (9)$$

Ainsi  $\hat{\theta}(\mathbf{x}) = \theta_0 + \int \mu(\theta) \frac{p(\mathbf{x}; \theta)}{p(\mathbf{x}; \theta_0)} d\theta$  avec  $\mu(\theta) = -1/2 \lambda(\theta)$ . Le report de cette dernière expression de  $\hat{\theta}(\mathbf{x})$  dans l'équation (5) conduit au résultat annoncé. ■

## III. BORNES LOCALES AVEC UN JEU DISCRET DE

### CONSTRAINTES

#### A. Résultat préliminaire

Soit  $f(\mathbf{x})$  et  $g(\mathbf{x})$  deux fonctions de  $\mathbb{R}^N$  dans  $\mathbb{R}$ , de carré sommable. On note  $\|f(\mathbf{x})\| = \left( \int f^2(\mathbf{x}) d\mathbf{x} \right)^{1/2}$  et  $(f, g)$  le produit scalaire  $\int f(\mathbf{x}) g(\mathbf{x}) d\mathbf{x}$ .

*Theorem 2:* Soient  $u(\mathbf{x}), g_0(\mathbf{x}), \dots, g_K(\mathbf{x})$  un ensemble de fonctions. La valeur minimale de  $\|u(\mathbf{x})\|^2$  lorsque l'on impose à  $u(\mathbf{x})$  les  $K + 1$  contraintes :

$$(u(\mathbf{x}), g_k(\mathbf{x})) = c_k \text{ pour } k = 0, \dots, K \quad (10)$$

est donnée par

$$\min \|u(\mathbf{x})\|^2 = \mathbf{c}^T \mathbf{G}^{-1} \mathbf{c} \text{ où } \mathbf{c} = (c_0 \dots c_K)^T \text{ et} \\ \mathbf{G} = [G_{mn}] \text{ avec } G_{mn} = (g_m, g_n) \quad (11)$$

*Proof:* La démonstration est similaire à celle du paragraphe précédent. ■

### B. Formulation générales des bornes locales

Particularisons maintenant les résultats du théorème précédent aux problèmes qui nous intéressent. L'EQM en  $\theta_0$  définie en (1) peut se réécrire

$$EQM = \int u^2(\mathbf{x}) d\mathbf{x} \quad (12)$$

avec

$$u(x) = (\hat{\theta}(\mathbf{x}) - \theta_0) \sqrt{p(\mathbf{x}; \theta_0)} \quad (13)$$

Le biais en  $\theta$  est donné par

$$b(\theta) = \int (\hat{\theta}(x) - \theta) p(x; \theta) dx = \left( u(x), \frac{p(\mathbf{x}; \theta)}{\sqrt{p(\mathbf{x}; \theta_0)}} \right) + (\theta_0 - \theta). \quad (14)$$

et sa dérivée  $n^{\text{ième}}$  par

$$\frac{d^n b(\theta)}{d\theta^n} = \left( u(x), \frac{1}{\sqrt{p(\mathbf{x}; \theta_0)}} \frac{\partial^n p(\mathbf{x}; \theta)}{\partial \theta^n} \right) - \delta_{n1} \text{ pour } n \geq 1. \quad (15)$$

Quand nous introduirons des contraintes spécifiques sur le biais et ses dérivées pour certaines valeurs du paramètre, elles se traduiront, d'après les expressions (14) et (15), par des égalités du type (10). Le théorème 2 nous fournira alors directement la valeur minimale recherchée de l'EQM.

## IV. BORNES CLASSIQUES REVISITÉES

On s'intéresse dans ce paragraphe à l'application du théorème 2 afin de retrouver les bornes les plus connues: Cramer Rao, Battacharyya, et Barankin.

### A. Bornes de Cramer Rao et de Battacharyya

La borne de Battacharyya à l'ordre  $K$  s'obtient en imposant au biais  $b(\theta)$  et à ses  $K$  dérivées successives d'être nuls en  $\theta_0$ . Le cas particulier  $K = 1$  correspond à la BCR. On définit ainsi d'après (14) et (15) un jeu de contraintes de la forme (10) avec

$$c_k = \delta_{k1} \text{ et } \begin{cases} g_0(\mathbf{x}) = p^{1/2}(\mathbf{x}; \theta_0) \text{ et} \\ g_k(\mathbf{x}) = p^{-1/2}(\mathbf{x}; \theta_0) p^{(k)}(\mathbf{x}; \theta_0) \\ \text{pour } 1 \leq k \leq K \end{cases} \quad (16)$$

où  $.^{(n)}$  représente la dérivée  $n^{\text{ième}}$  par rapport à  $\theta$ . L'application du théorème 2 nécessite le calcul de la matrice  $\mathbf{G}$  des produits scalaires  $(g_m, g_n)$  qui admet l'expression :

$$\mathbf{G} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & \mathbf{B} & \\ 0 & & & \end{pmatrix} \quad (17)$$

avec

$$B_{mn} = (g_m(\mathbf{x}), g_n(\mathbf{x})) = \int \frac{p^{(m)}(\mathbf{x}; \theta_0) p^{(n)}(\mathbf{x}; \theta_0)}{p(\mathbf{x}; \theta_0)} d\mathbf{x} \quad \text{pour } m, n \geq 1. \quad (18)$$

L'EQM minimale, notée  $EQM_{Bat}$ , est fournie par (11) qui s'écrit encore, en tenant compte de la forme particulière de  $\mathbf{G}$  (17) et du vecteur  $\mathbf{c}$  de contraintes (16)

$$EQM_{Bat} = (1, 0, \dots, 0) \mathbf{B}^{-1} (1, 0, \dots, 0)^T. \quad (19)$$

### B. Bornes de Barankin

Barankin [3] a proposé une méthode pour obtenir l'EQM minimale, notée  $EQM_{Bar}$ , d'un estimateur sans biais pour toutes les valeurs admissibles du paramètre, qui se réduit dans ce papier à l'intervalle  $[a, b]$ . Cette approche passe entre autres, par la recherche de la variance minimale, notée  $EQM(\Theta)$ , d'un estimateur dont le biais est nul sur  $\Theta = \{\theta_0, \dots, \theta_K\}$ . Il n'existe pas à notre connaissance, dans la littérature, de travaux permettant d'obtenir  $EQM(\Theta)$  de manière analytique. L'absence d'une telle expression analytique explique en partie la difficulté d'utilisation de la borne de Barankin en traitement du signal. Nous montrons dans ce paragraphe comment l'expression analytique de  $EQM(\Theta)$  s'obtient aisément à l'aide du théorème 2. Les contraintes de biais nul sur  $\Theta = \{\theta_0, \dots, \theta_K\}$  définissent d'après (14) un jeu de contraintes de la forme (10) avec

$$\begin{cases} c_0 = 0 \\ c_k = \theta_k - \theta_0 \triangleq h_k \quad \text{et} \\ \begin{cases} g_0(\mathbf{x}) = p^{1/2}(\mathbf{x}; \theta_0) \text{ et} \\ g_k(\mathbf{x}) = p^{-1/2}(\mathbf{x}; \theta_0) p(\mathbf{x}; \theta_k) \\ \text{pour } 1 \leq k \leq K \end{cases} \end{cases} \quad (20)$$

D'après le théorème 2 l'EQM minimale recherchée est donnée par  $\mathbf{c}^T \mathbf{G}^{-1} \mathbf{c}$  avec

$$\mathbf{G} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & & & \\ \vdots & & \mathbf{D} & \\ 1 & & & \end{pmatrix} \quad (21)$$

où la matrice  $\mathbf{D}$  est définie par:

$$D_{mn} = \int \frac{p(\mathbf{x}; \theta_m) p(\mathbf{x}; \theta_n)}{p(\mathbf{x}; \theta_0)} d\mathbf{x} \quad \text{pour } m, n \geq 1. \quad (22)$$

L'inversion par bloc de  $\mathbf{G}$  dans l'expression (11) de l'EQM minimale conduit alors à:

$$EQM(\Theta) = (h_1 \cdots h_K) (\mathbf{D} - \mathbf{11}^T)^{-1} (h_1 \cdots h_K)^T. \quad (23)$$

Grace à l'équation (23), il est maintenant aisé d'obtenir la borne de Barankin comme

$$EQM_{Bar} = \lim_{K \rightarrow \infty} \sup_{h_1, \dots, h_K} \{EQM(\Theta)\}. \quad (24)$$

**Remarque :** Afin de pouvoir exploiter les travaux de Barankin avec une complexité numérique réduite Chapman et Robbins ([5]) ont proposé de travailler avec  $K = 1$ . On retrouve, grâce à l'équation (23), le minorant suivant de  $EQM_{Bar}$ :

$$EQM_{ChapRob} = \sup_h \left\{ \frac{h^2}{\int \frac{p^2(\mathbf{x}; \theta_0 + h)}{p(\mathbf{x}; \theta_0)} d\mathbf{x} - 1} \right\}. \quad (25)$$

## V. EXEMPLE D'APPLICATION

A titre d'exemple, considérons le problème d'analyse spectrale classique suivant. L'observation vectorielle  $\mathbf{x}$  est complexe et admet le modèle:

$$\mathbf{x} = \mathbf{s}(\theta) + \mathbf{b} \quad (26)$$

où :  $\mathbf{s}(\theta) = (1, e^{2i\pi\theta}, \dots, e^{2i\pi(N-1)\theta})^T$ ,  $\theta \in [-0.5, 0.5]$  est le paramètre déterministe inconnu à estimer et  $\mathbf{b}$  est un bruit gaussien, complexe circulaire centré de matrice de covariance connue  $\sigma^2 \mathbf{I}$ . L'application des résultats précédents conduit aux expressions suivantes des bornes.

**Cramer Rao.**  $EQM_{CR} = \frac{\sigma^2}{2 \|\mathbf{s}^{(1)}\|^2}$ .

**Battacharyya.** La borne de Battacharyya à l'ordre 1 ou 2 redonne dans ce cas particulier la borne de Cramer Rao. A l'ordre 3, après de longs calculs que nous ne pouvons pas reporter ici, nous obtenons

$$EQM_{Bat3} = \frac{\sigma^2 (48 \|\mathbf{s}^{(1)}\|^6 + 36 \sigma^2 \|\mathbf{s}^{(1)}\|^2 \|\mathbf{s}^{(2)}\|^2 + 2 \sigma^4 \|\mathbf{s}^{(3)}\|^2)}{[96 \|\mathbf{s}^{(1)}\|^8 + 72 \sigma^2 \|\mathbf{s}^{(1)}\|^4 \|\mathbf{s}^{(2)}\|^2 + 4 \sigma^4 (\|\mathbf{s}^{(3)}\|^2 \|\mathbf{s}^{(1)}\|^2 - \|\mathbf{s}^{(2)}\|^4)]}. \quad (27)$$

Cette expression redonne à fort RSB ( $\sigma^2$  petit) la borne de Cramer Rao.

**Barankin.** Elle est donnée par les expressions (23) et (24) dans lesquelles la matrice  $\mathbf{D}$  est donnée par l'équation (28) :

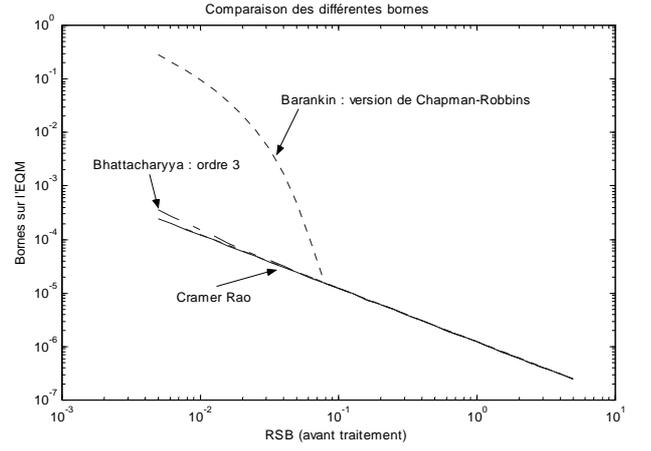


Fig. 1. Comparaison des différentes bornes dans le cas de l'estimation de la fréquence d'une exponentielle d'amplitude complexe connue.  $N=32$  échantillons.

$$D_{mn} = \exp \left( \frac{2 \operatorname{Re} \left[ (\mathbf{s}(\theta_m) - \mathbf{s}(\theta_0))^H (\mathbf{s}(\theta_n) - \mathbf{s}(\theta_0)) \right]}{\sigma^2} \right). \quad (28)$$

La figure 1 représente l'évolution des différentes bornes en fonction du rapport signal sur bruit dans le cas de  $N = 32$  échantillons. La borne dite de Barankin représentée sur la figure correspond à la version de Chapman-Robbins. Nous pouvons noter que la borne de Battacharyya à l'ordre 3 n'apporte que peu d'amélioration par rapport à la borne de Cramer Rao. Il faudrait aller à des ordres élevés pour observer une différence notable avec la borne de Cramer Rao. La borne de Chapman-Robbins, quant à elle rend bien compte du phénomène de "SNR Threshold" observé en pratique.

## VI. CONCLUSION

La présentation unifiée des différentes bornes exposée dans cet article contribue à une meilleure compréhension des hypothèses et mécanismes à l'oeuvre dans leur établissement. Elle permet l'obtention de l'estimateur optimal comme solution d'une équation intégrale. Le théorème 2 fournit de plus l'expression analytique de  $EQM(\Theta)$ . Enfin, cette approche autorise le développement d'autres bornes par le choix d'un jeu de contraintes appropriées.

## REFERENCES

- [1] H. Cramér, *Mathematical Methods of Statistics*. Princeton, NJ: Princeton Univ. Press, 1949.
- [2] A. Battacharyya, On some analogues of the amount of information and their use in statistical estimation, *Sankhya*, Vol. 8 (1946), pp. 1-14.
- [3] E. W. Barankin, Locally best unbiased estimates, *Annals of Math. Stat.*, Vol. 20, pp. 477-501, 1949.
- [4] T. L. Marzetta, Computing the Barankin bound by solving an unconstrained quadratic optimization problem, in *Proc. ICASSP'97*, pp. 3829-3832.
- [5] D.G. Chapman and H. Robbins "Minimum variance estimation without regularity assumptions" *Ann. Math. Statis.* vol. 22, pp 581-586. 1951.