



HAL
open science

Oracle performance estimation of Bernoulli-distributed sparse vectors

Remy Boyer, Pascal Larzabal, Bernard Fleury

► **To cite this version:**

Remy Boyer, Pascal Larzabal, Bernard Fleury. Oracle performance estimation of Bernoulli-distributed sparse vectors. 2016 IEEE Statistical Signal Processing Workshop (SSP), Jun 2016, Palma de Majorque, Spain. 10.1109/ssp.2016.7551780 . hal-01313460

HAL Id: hal-01313460

<https://hal.science/hal-01313460>

Submitted on 11 May 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ORACLE PERFORMANCE ESTIMATION OF BERNOULLI-DISTRIBUTED SPARSE VECTORS

Rémy Boyer⁽¹⁾, Pascal Larzabal⁽²⁾, Bernard-Henri Fleury⁽³⁾

⁽¹⁾ Université Paris-Sud, L2S, CNRS, CentraleSupélec, France

⁽²⁾ ENS-Cachan, SATIE, France

⁽³⁾ Aalborg University, Department of Electronic Systems, Denmark

ABSTRACT

Compressed Sensing (CS) is now a well-established research area and a plethora of applications has emerged in the last decade. In this context, assuming N available noisy measurements, lower bounds on the Bayesian Mean Square Error (BMSE) for the estimated entries of a sparse amplitude vector are derived in the proposed work for (i) a Gaussian overcomplete measurement matrix and (ii) for a random support, assuming that each entry is modeled as the product of a continuous random variable and a Bernoulli random variable indicating that the current entry is non-zero with probability P . A closed-form expression of the Expected CRB (ECRB) is proposed. In the second part, the BMSE of the Linear Minimum MSE (LMMSE) estimator is derived and it is proved that the LMMSE estimator tends to be statistically efficient in asymptotic conditions, i.e., if product $(1 - P)^2$ SNR is maximized. This means that in the context of the Gaussian CS problem, the LMMSE estimator gathers together optimality in the low noise variance regime and a simple derivation (as opposed to the derivation of the MMSE estimator). This result is original because the LMMSE estimator is generally sub-optimal for CS when the measurement matrix is a single realization of a given random process.

1. INTRODUCTION

In the Compressed Sensing (CS) framework [1,2], it is assumed that the signal of interest can be linearly decomposed into few basis vectors. By exploiting this property, CS allows for using sampling rates lower [3] than Shannon's sampling rate. As a result, CS technics have found a plethora of applications in numerous areas, e.g. array processing [4] or the published contributions during the special session [5], wireless communications, video processing or in MIMO radar.

Mean-square error (MSE) lower bounds [6] for the estimation of the non-zero entries of a sparse vector have been proposed in [7–11]. In particular, the existing literature has studied MSE lower bounds when the support set (i.e., the indexes of the non-zero entries) is provided by an oracle or a genie. At first glance, this assumption seems unrealistic and too optimistic. However, it has been proved in [7–9] that for a sufficiently low noise variance, it exists sparse-based estimators, (e.g., [12,13]) unaware of the support which reach

the oracle MSE lower bound. This existence property explains why the oracle framework is adopted in this work.

Unlike the existing approaches, our performance analysis is provided for (i) a Gaussian measurement matrix and (ii) for a random support of random cardinality. More precisely, we assume that each amplitude of interest has the probability P to be non-zero. This statistical model is called a Bernoulli prior [14] and implies that the support set is randomly generated with a Binomial-distributed cardinality.

2. CS MODEL AND RECOVERY REQUIREMENTS

Let \mathbf{y} be a $N \times 1$ noisy measurement vector in the (real) Compressed Sensing (CS) model [1–3]:

$$\mathbf{y} = \Psi \mathbf{s} + \mathbf{n}, \quad (1)$$

where \mathbf{n} is a (zero-mean) white Gaussian noise vector with component variance σ^2 and Ψ is the $N \times K$ sensing/measurement matrix with $N < K$. The vector \mathbf{s} is given by $\mathbf{s} = \Phi \boldsymbol{\theta}$, where Φ is a $K \times K$ orthonormal matrix and $\boldsymbol{\theta}$ is a $K \times 1$ amplitude vector. With this definition eq. (1) can be recast as

$$\mathbf{y} = \mathbf{H} \boldsymbol{\theta} + \mathbf{n} \quad (2)$$

where the overcomplete $N \times K$ matrix $\mathbf{H} = \Psi \Phi$ is commonly referred to as the dictionary. The amplitude vector is assumed to be ℓ -sparse, with $\ell < N$. With \mathcal{S} denoting the set of indices of the non-zero entries in $\boldsymbol{\theta}$ and we have $\ell = |\mathcal{S}|$. Under this assumption, we can rewrite the first summand in eq. (2) as $\mathbf{H} \boldsymbol{\theta} = \mathbf{H}_{\mathcal{S}} \boldsymbol{\theta}_{\mathcal{S}}$ with $\mathbf{H}_{\mathcal{S}} = \Psi \Phi_{\mathcal{S}}$, where $\boldsymbol{\theta}_{\mathcal{S}}$ is the vector composed of the entries in $\boldsymbol{\theta}$ with indices in \mathcal{S} , i.e. the ℓ non-zero entries of this vector, and the $N \times \ell$ matrix $\Phi_{\mathcal{S}}$ is built up with the ℓ columns of Φ having their indices in \mathcal{S} . The entries of the measurement matrix, Ψ , are i.i.d. Gaussian distributed with zero mean and variance $1/N$. With this choice the Restricted Isometry Property (RIP) [1–3] is satisfied with high probability [15]. The RIP ensures that practical algorithms can successfully recover any sparse amplitude vector from noisy measurements. Classical sampling theory states that to ensure no loss of information the number of measurements, N , should be at least equal to K . By contrast, CS shows that this holds true if $N = \ell O(\log(K/L)) < K$ if the vector $\boldsymbol{\theta}$ is ℓ -sparse in a given basis Φ [1].

Remark 2.1 (Identifiability constraint) *Given N available measurements, we cannot accurately estimate more than $N - 2$ parameters whose indexes are collected in set $\mathcal{I} \subset \{1, \dots, K\}$ of cardinality $|\mathcal{I}| < N$. In the estimation point of view, considering more*

This work was partly supported by the European Commission in the framework of the FP7 Network of Excellence in Wireless COMMUNICATIONS NEWCOM# (Grant agreement no. 318306). Rémy Boyer and Pascal Larzabal are involved in projects MAGELLAN (ANR-14-CE23-0004-01) and MI-CNRS TITAN.

parameters of interest than the number of available measurements leads to a singular Fisher Information Matrix (FIM). In this scenario, it cannot exist an estimator with finite variance [16,17].

3. LOWER BOUNDS FOR THE BMSE

We define the conditional BMSE given $\mathcal{S} \in \mathcal{P}$ and $\mathcal{I} \in \Omega$ to be

$$\text{BMSE}_{\mathcal{L}} = \frac{1}{N} \mathbb{E}_{\mathbf{y}, \mathbf{H}_{\mathcal{S}}, \theta_{\mathcal{S}} | \mathcal{S}, |\mathcal{S}|} \left\| \boldsymbol{\theta}_{\mathcal{S}} - \hat{\boldsymbol{\theta}}_{\mathcal{S}}(\mathbf{y}, \mathbf{H}_{\mathcal{S}}) \right\|^2$$

where $\hat{\boldsymbol{\theta}}_{\mathcal{S}}(\mathbf{y}, \mathbf{H}_{\mathcal{S}})$ is an estimate of $\boldsymbol{\theta}_{\mathcal{S}}$ that knows the support \mathcal{S} and respect the identifiability constraint in Remark 2.1.

Remark 3.1 The $\text{BMSE}_{\mathcal{L}}$ is a natural limit performance criterion on the BMSE of an estimator unaware of the support.

Averaging the local $\text{BMSE}_{\mathcal{L}}$ over the random quantities, i.e., \mathcal{S} and $|\mathcal{S}|$ yields the global BMSE:

$$\text{BMSE} = \mathbb{E}_{|\mathcal{S}| | \mathcal{I}} \mathbb{E}_{\mathcal{S} | |\mathcal{S}|} \text{BMSE}_{\mathcal{L}}$$

As proved in [18], the conditional ECRB is the tightest Bayesian lower bound in the low noise regime regarding the considered model. Consequently, we focus our effort on this lower bound.

The global BMSE is lower bounded by the global ECRB, denoted by ECRB, according to

$$\text{BMSE} \geq \text{ECRB} = \mathbb{E}_{|\mathcal{S}| | \mathcal{I}} \mathbb{E}_{\mathcal{S} | |\mathcal{S}|} \text{ECRB}_{\mathcal{L}} \quad (3)$$

where $\text{ECRB}_{\mathcal{L}}$ is the conditional ECRB derived in Appendix 6.1.

3.1. conditional ECRB in case of a Gaussian measurement matrix

Lemma 3.2 For a Gaussian measurement matrix, the local $\text{ECRB}_{\mathcal{L}}$ given in eq. (7) reads

$$\text{ECRB}_{|\mathcal{S}|} = \sigma^2 \frac{|\mathcal{S}|}{N - |\mathcal{S}|}.$$

Proof The proof is straightforward by using eq. (7) and the property of the Wishart matrices [19] for $|\mathcal{S}| < N$. Let $\mathbf{Z}_{\mathcal{S}} = \sqrt{N} \mathbf{H}_{\mathcal{S}}$. Observe that the entries of matrix $\mathbf{Z}_{\mathcal{S}}$ have now a unit variance. Thus $\text{ECRB}_{\mathcal{L}}$ given in eq. (7) reads

$$\text{ECRB}_{|\mathcal{S}|} = \sigma^2 \mathbb{E}_{\mathbf{H}_{\mathcal{S}} | \mathcal{S}} \text{Tr} \left[\left(\mathbf{Z}_{\mathcal{S}}^T \mathbf{Z}_{\mathcal{S}} \right)^{-1} \right] = \sigma^2 \frac{|\mathcal{S}|}{N - |\mathcal{S}|}.$$

□

Remark 3.3 For a Gaussian measurement matrix, the conditional ECRB is a function of the random cardinality $|\mathcal{S}|$. Thus, we note $\text{ECRB}_{\mathcal{L}} = \text{ECRB}_{|\mathcal{S}|}$.

3.2. Global ECRB for a random support of Binomial cardinality

Thanks to the above remark, the global ECRB in eq. (3) is

$$\text{ECRB} = \mathbb{E}_{|\mathcal{S}| | \mathcal{I}} \text{ECRB}_{|\mathcal{S}|} = \sum_{\ell=1}^{|\mathcal{I}|} \Pr(|\mathcal{S}| = \ell; |\mathcal{I}|) \text{ECRB}_{|\mathcal{S}|=\ell}.$$

3.2.1. Random cardinality with an identifiability constraint

At this point, for $1 \leq i \leq K$, we have two possible cases:

$$\begin{aligned} \theta_{i \in \mathcal{I}} &\neq 0 && \text{with probability } P, \\ \theta_{i \in \mathcal{I}} &= 0 && \text{with probability } 1 - P. \end{aligned}$$

This is equivalent to modeled the i -th entry according to $1_{\mathcal{S}}(i) 1_{\mathcal{I}}(i) \theta_i$ where $1_{\mathcal{S}}(i) \sim \mathcal{B}(P)$. By doing this, the identifiability constraint is taken into account thanks to $1_{\mathcal{I}}(i)$. By definition the cardinality of \mathcal{S} conditionally to a given set \mathcal{I} is

$$|\mathcal{S}| \Big| \mathcal{I} = \sum_{i=1}^K 1_{\mathcal{S}}(i) 1_{\mathcal{I}}(i) = \sum_{i \in \mathcal{I}} 1_{\mathcal{S}}(i) = \sum_{i=1}^{|\mathcal{I}|} 1_{\mathcal{S}}(i)$$

with $|\mathcal{I}| = N - 1$. So, $|\mathcal{S}| \Big| \mathcal{I}$ is the sum of $|\mathcal{I}|$ i.i.d. Bernoulli-distributed variables and is in fact not a function of set \mathcal{I} but of its non-random cardinality $|\mathcal{I}|$. As a consequence, $|\mathcal{S}|; |\mathcal{I}| \sim \mathcal{BN}(|\mathcal{I}|, P)$. We will note $\Pr(|\mathcal{S}| \Big| \mathcal{I}) = \Pr(|\mathcal{S}|; |\mathcal{I}|)$.

3.2.2. Closed-form expression of the global ECRB

Result 3.1 For any amplitude vector prior, for $L < N - 1$ where $L = \mathbb{E}[|\mathcal{S}|]$ and for a probability of success given by $P = L/(N - 1)$, the global (i.e. on set \mathcal{P}) ECRB verifies the following inequality:

$$\text{BMSE} \geq \text{ECRB} = \sigma^2 \frac{P}{1 - P} \left(1 - P^{N-1} \right). \quad (4)$$

Proof See Appendix 6.2.

Remark 3.4 For a large number of measurements, i.e., $N \gg 1^1$, we can give the following approximation:

$$\text{ECRB} \approx \sigma^2 \frac{P}{1 - P} = \sigma^2 \frac{L}{N - L}.$$

4. ANALYSIS OF THE LMMSE ESTIMATOR IN THE LOW NOISE VARIANCE REGIME

4.1. Global BMSE of the LMMSE estimator

Result 4.1 In the low noise variance regime and for uncorrelated amplitudes of variance σ_{θ}^2 , the BMSE of the LMMSE estimator is given by

$$\text{BMSE} \approx \sigma^2 \left(\eta_1 - \frac{N}{\text{SNR}} \eta_2 \right) \quad (5)$$

where η_1 and η_2 are given in eq. (6) and eq. (7) with $P' = L/(N - 2)$.

Proof See Appendix 6.3.

4.2. Discussion on the asymptotic statistical efficiency of the LMMSE

Result 4.2 In the low noise variance regime, the ratio between the global BMSE of the LMMSE estimator and the global ECRB on set \mathcal{P} for $N \gg 1$ is given by $\frac{\text{BMSE}}{\text{ECRB}} \approx 1 - \frac{1}{(1-P)^2 \text{SNR}}$.

¹Note that it is assumed that L is not neglected with respect to N .

$$\begin{aligned}\eta_1 &= \frac{(N-1)NP' - NP'(1-P') - (NP')^2 + NP'^N}{(1-P')^2 N(N-1)} \\ \eta_2 &= \frac{(N+1)P' - (N+1)P'^{N+1} - N(N+1)P'^N(1-P') - \frac{N(N+1)(N-1)}{2} P'^{N-1}(1-P')^2}{(1-P')^3(N-1)(N+1)}\end{aligned}\quad (6)$$

Proof For $N \gg 1$, the following approximations hold: $P' \approx P$, $\eta_1 \approx \text{ECRB}/\sigma^2$ and $\eta_2 \approx \frac{P}{N(1-P)^3}$. Using the expression of ECRB and the Result 4.1 provide the desired result. \square

Remark 4.1 Based on the above result, in the low noise regime and for a large number of measurements, the LMMSE estimator tends to be statistically efficient according to the following criterion:

$$\max \{(1-P)^2 \text{SNR}\}.$$

This means that we can identify the key quantities to characterize the performance of the LMMSE estimator. More precisely, the LMMSE estimator tends to be efficient if at least one of the following conditions is satisfied.

- \mathcal{C}_1 . The SNR goes to infinity with finites N and L . Note that if $P \rightarrow 1$ and a large L can mitigate the statistical efficiency of the LMMSE estimator even for a high SNR.
- \mathcal{C}_2 . The number of measurements goes to infinity. In this last case, $P \rightarrow 0$. So, in the case of highly sparse amplitude vector, the statistical efficiency of the LMMSE estimator is governed by the SNR.

In our view, this result is original and important. Indeed, it is well-known that the LMMSE estimator for deterministic matrix \mathbf{H}_S is sub-optimal, i.e., does not reach the BCRB or the ECRB (indifferently for deterministic matrix \mathbf{H}_S) for any amplitude vector priors excepted for the Gaussian prior [20]. In the Gaussian case, it is well-known that the LMMSE estimator is in fact the MMSE estimator and is thus statistically efficient. In the CS framework where \mathbf{H}_S is time-varying/stochastic, the situation is rather different and to the best of our knowledge has not been fully investigated. This fact explains why the LMMSE estimator optimality for any priors seems to us an original contribution. This result is also of great practical interest because the LMMSE estimator has been introduced in the statistical signal community thanks to its easy computation (recall that only the second-order statistics of the amplitude source vector are needed). At contrary the optimal MMSE estimator needs to know the posterior distribution which is generally mathematically intractable unless for the Gaussian prior. So, when \mathbf{H}_S is time-varying/stochastic, the LMMSE estimator gathers together asymptotic (in number of measurements or/and in SNR) statistical optimality and a simple computation. These two characteristics being generally contradictory.

5. CONCLUSION

The lowest accuracy for the estimation of a sparse amplitude vector given N available measurements has been derived for (i) a Gaussian overcomplete measurement matrix and (ii) for a random support, meaning that each entry of the vector of interest is modeled as the product of a continuous random variable and a Bernoulli random variable indicating if the current entry is non-zero with probability P . As the Expected CRB is the tightest Bayesian lower bound in

the low noise regime, the conditions for that the oracle LMMSE estimator tends to be statically efficient are derived thanks to a simple criterion. It is proved in this work that the oracle LMMSE estimator gathers together asymptotic statistical optimality and simple computation relatively to the MMSE estimator. These two characteristics being generally contradictory.

6. APPENDIX

6.1. Derivation of the conditional ECRB

The conditional ECRB is based on the deterministic/Bayesian connexion [21]. Following this principle, remark that an alternative expression of the local BMSE given \mathcal{S} can be obtained by rewriting it as an expected local MSE criterion according to $\text{BMSE}_{\mathcal{L}} = \frac{\mathbb{E}_{\mathbf{H}_S, \boldsymbol{\theta}_S | \mathcal{S}} \text{MSE}_{\mathcal{L}}}{N}$ where the MSE conditioned to \mathbf{H}_S and $\boldsymbol{\theta}_S$ is defined as

$$\text{MSE}_{\mathcal{L}} = \mathbb{E}_{\mathbf{y} | \mathbf{H}_S, \boldsymbol{\theta}_S | \mathcal{S}} \|\boldsymbol{\theta}_S - \hat{\boldsymbol{\theta}}_S(\mathbf{y}, \mathbf{H}_S)\|^2.$$

The local MSE given \mathbf{H}_S , $\boldsymbol{\theta}_S$ and \mathcal{S} verifies inequality $\text{MSE}_{\mathcal{L}} \geq \text{Tr}[\mathbf{F}_S^{-1}]$ where the Fisher Information Matrix (FIM) is given by

$$[\mathbf{F}_S]_{ij} = \mathbb{E}_{\mathbf{y} | \mathbf{H}_S, \boldsymbol{\theta}_S, \mathcal{S}} \left\{ -\frac{\partial^2 \log p(\mathbf{y} | \mathbf{H}_S, \boldsymbol{\theta}_S)}{\partial [\boldsymbol{\theta}_S]_i \partial [\boldsymbol{\theta}_S]_j} \right\} = \frac{1}{\sigma^2} [\mathbf{H}_S^T \mathbf{H}_S]_{ij}$$

since $\mathbf{y} | \mathbf{H}_S, \boldsymbol{\theta}_S, \mathcal{S} \sim \mathcal{N}(\mathbf{H}_S \boldsymbol{\theta}_S, \sigma^2 \mathbf{I})$ [22]. Finally, the trace of the conditional ECRB takes the following simple expression:

$$\text{ECRB}_{\mathcal{L}} = \frac{\sigma^2}{N} \mathbb{E}_{\mathbf{H}_S | \mathcal{S}} \text{Tr} \left[(\mathbf{H}_S^T \mathbf{H}_S)^{-1} \right]. \quad (7)$$

6.2. Proof of Result 3.1

Using eq. (3), Lemma 3.2, Remark 3.3, the ECRB is given by

$$\text{ECRB} = \sigma^2 \sum_{\ell=1}^{N-1} \frac{\ell}{N-\ell} \Pr(|\mathcal{S}| = \ell) = \frac{\sigma^2}{N(1-P)} \left(\mathbb{E}G - NP^N \right)$$

where $G \sim \mathcal{BN}(N, P)$. Using the first moment of the Binomial variable G [23] given by $\mathbb{E}G = NP$, we obtain eq. (4).

6.3. Proof of Result 4.1

The conditional BMSE of the LMMSE based on the Bayesian Gauss-Markov Theorem [20] is given by

$$\text{BMSE} = \sum_{\ell \in \mathcal{V}} \sum_{\mathcal{S}} \Pr(\mathcal{S}, |\mathcal{S}| = \ell) \mathbb{E}_{\mathbf{H}_S | \mathcal{S}} (\text{BMSE}_{\mathcal{S}})$$

where $\text{BMSE}_{\mathcal{S}} = \frac{1}{N} \text{Tr} \left[\left(\frac{1}{\sigma^2} \mathbf{H}_S^T \mathbf{H}_S + \mathbf{R}_{\boldsymbol{\theta}_S}^{-1} \right)^{-1} \right]$. Exploiting the decorrelation of the amplitudes and using \mathbf{Z}_S , we obtain $\text{BMSE} = \sigma^2 \text{Tr} \left[\left(\mathbf{Z}_S^T \mathbf{Z}_S + \frac{N}{\text{SNR}} \mathbf{I} \right)^{-1} \right]$ where $\text{SNR} = \sigma_a^2 / \sigma^2$. In the low

noise variance regime or for a large SNR, the conditional BMSE can be approximated thanks to the Neumann expansion according to

$$\text{BMSE}_{\mathcal{S}} = \sigma^2 \text{Tr} \left[\left(\mathbf{Z}_{\mathcal{S}}^T \mathbf{Z}_{\mathcal{S}} \right)^{-1} \right] - \frac{N\sigma^2}{\sigma_{\theta}^2} \text{Tr} \left[\left(\mathbf{Z}_{\mathcal{S}}^T \mathbf{Z}_{\mathcal{S}} \right)^{-2} \right] + O(\sigma^6).$$

Thus, the global BMSE is given by eq. (5) where

$$\begin{aligned} \eta_1 &= \sum_{\ell \in \mathcal{V}} \sum_{\mathcal{S}} \Pr(\mathcal{S}, |\mathcal{S}| = \ell) \mathbb{E}_{\mathbf{H}_{\mathcal{S}} | \mathcal{S}} \text{Tr} \left[\left(\mathbf{Z}_{\mathcal{S}}^T \mathbf{Z}_{\mathcal{S}} \right)^{-1} \right] \\ \eta_2 &= \sum_{\ell \in \mathcal{V}} \sum_{\mathcal{S}} \Pr(\mathcal{S}, |\mathcal{S}| = \ell) \mathbb{E}_{\mathbf{H}_{\mathcal{S}} | \mathcal{S}} \text{Tr} \left[\left(\mathbf{Z}_{\mathcal{S}}^T \mathbf{Z}_{\mathcal{S}} \right)^{-2} \right]. \end{aligned}$$

According to [19], we have

$$\mathbb{E}_{\mathbf{H}_{\mathcal{S}} | \mathcal{S}} \text{Tr} \left[\left(\mathbf{Z}_{\mathcal{S}}^T \mathbf{Z}_{\mathcal{S}} \right)^{-2} \right] = \frac{N\ell}{(N-\ell)(N-\ell-1)(N-\ell+1)}$$

for $\ell < N-1$. This technical difficulty justifies $|\mathcal{S}| \sim \mathcal{BN}(N-2, P')$ with $P' = L/(N-2)$. Let $V \sim \mathcal{BN}(N+1, P')$, then, η_2 is given by

$$\eta_2 = \frac{\sum_{\ell=0}^{N-2} \ell \Pr(V = \ell)}{(1-P')^3 (N-1)(N+1)} = \frac{\mathbb{E}V - \sum_{\ell=N-1}^{N+1} \ell \Pr(V = \ell)}{(1-P')^3 (N-1)(N+1)}$$

where $\mathbb{E}V = (N+1)P'$ which leads to eq. (7). Due to the new constraint on the identifiability constraint, η_1 cannot be confuse with the global ECRB given in eq. (4). But, let $G' \sim \mathcal{BN}(N, P')$, we have

$$\eta_1 = \frac{(N-1) \sum_{\ell=0}^{N-2} \ell \Pr(G' = \ell) - \sum_{\ell=0}^{N-2} \ell^2 \Pr(G' = \ell)}{(1-P')^2 N(N-1)}$$

where $\sum_{\ell=0}^{N-2} \ell \Pr(G' = \ell) = \mathbb{E}G' - \sum_{\ell=N-1}^N \ell \Pr(G' = \ell)$ and $\sum_{\ell=0}^{N-2} \ell^2 \Pr(G' = \ell) = \mathbb{E}G'^2 - \sum_{\ell=N-1}^N \ell^2 \Pr(G' = \ell)$ with $\mathbb{E}G' = NP'$ and $\mathbb{E}G'^2 = \text{Var}G' + (NP')^2 = NP'(1-P') + (NP')^2$ [23] which leads to eq. (6).

7. REFERENCES

- [1] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [2] R. Baraniuk, "Compressive sensing [lecture notes]," *Signal Processing Magazine, IEEE*, vol. 24, no. 4, pp. 118–121, 2007.
- [3] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling," *Signal Processing Magazine, IEEE*, vol. 25, no. 2, pp. 21–30, 2008.
- [4] D. Malioutov, M. Cetin, and A. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 3010–3022, 2005.
- [5] R. Boyer and P. Larzabal, "Sparsity in array processing: methods and performances," in *IEEE Sensor Array and Multichannel (SAM) conference*. Special Session, <http://www.gtec.udc.es/sam2014/>, 2014.
- [6] H. L. Van Trees and K. L. Bell, "Bayesian bounds for parameter estimation and nonlinear filtering/tracking," *AMC*, vol. 10, p. 12, 2007.
- [7] B. Babadi, N. Kalouptsidis, and V. Tarokh, "Asymptotic achievability of the Cramér–rao bound for noisy compressive sampling," *IEEE Transactions on Signal Processing*, vol. 57, no. 3, pp. 1233–1236, 2009.
- [8] R. Boyer, B. Babadi, N. Kalouptsidis, and V. Tarokh, "Corrections to Asymptotic achievability of the Cramér–Rao bound for noisy compressive sampling," *IEEE Transactions on Signal Processing*, p. submitted, 2016.
- [9] R. Niazadeh, M. Babaie-Zadeh, and C. Jutten, "On the achievability of Cramér–Rao bound in noisy compressed sensing," *IEEE Transactions on Signal Processing*, vol. 60, no. 1, pp. 518–526, 2012.
- [10] Z. Ben-Haim and Y. Eldar, "The cramèr-rao bound for estimating a sparse parameter vector," *IEEE Transactions on Signal Processing*, vol. 58, no. 6, pp. 3384–3389, June 2010.
- [11] A. Wiesel, Y. Eldar, and A. Yeredor, "Linear regression with gaussian model uncertainty: Algorithms and bounds," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2194–2205, June 2008.
- [12] Y. C. Pati, R. Rezaifar, and P. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Conference Record of The Twenty-Seventh Asilomar Conference*. IEEE, 1993, pp. 40–44.
- [13] D. Needell and R. Vershynin, "Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 310–316, 2010.
- [14] N. Dobigeon and J.-Y. Tourneret, "Bayesian orthogonal component analysis for sparse representation," *IEEE Transactions on Signal Processing*, vol. 58, no. 5, pp. 2675–2685, 2010.
- [15] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, 2008.
- [16] P. Stoica and T. L. Marzetta, "Parameter estimation problems with singular information matrices," *IEEE Transactions on Signal Processing*, vol. 49, no. 1, pp. 87–90, 2001.
- [17] J. D. Gorman and A. O. Hero, "Lower bounds for parametric estimation with constraints," *IEEE Transactions on Information Theory*, vol. 36, no. 6, pp. 1285–1301, 1990.
- [18] J. H. Goulart, M. Boizard, R. Boyer, G. Favier, and P. Comon, "Tensor CP decomposition with structured factor matrices: Algorithms and Performance," *IEEE Transactions on Journal of Selected Topics in Signal Processing*, p. accepted, 2016.
- [19] R. Couillet and M. Debbah, *Random matrix methods for wireless communications*, 1st ed. New York, NY, USA: Cambridge University Press, 2011.
- [20] S. M. Kay, *Fundamentals of statistical signal processing: estimation theory*. Prentice-Hall, Inc., 1993.
- [21] Z. Ben-Haim and Y. Eldar, "A lower bound on the bayesian mse based on the optimal bias function," *IEEE Transactions on Information Theory*, vol. 55, no. 11, pp. 5179–5196, 2009.
- [22] P. Stoica and R. L. Moses, *Spectral analysis of signals*. Pearson/Prentice Hall Upper Saddle River, NJ, 2005.
- [23] N. L. Johnson, A. W. Kemp, and S. Kotz, *Univariate discrete distributions*. John Wiley & Sons, 2005, vol. 444.