Bayesian Cramér-Rao Bound for Noisy Non-Blind and Blind Compressed Sensing

H. Zayyani¹, M. Babaie-Zadeh¹ Member, and C. Jutten² Fellow
EDICS:SAS-ICAB or SAS-STAT

Abstract

In this paper, we address the theoretical limitations in reconstructing sparse signals (in a known complete basis) using compressed sensing framework. We also divide the CS to non-blind and blind cases. Then, we compute the Bayesian Cramer-Rao bound for estimating the sparse coefficients while the measurement matrix elements are independent zero mean random variables. Simulation results show a large gap between the lower bound and the performance of the practical algorithms when the number of measurements are low.

Index Terms—Compressed sensing, Sparse component analysis, Blind source separation, Cramer-Rao bound.

I. INTRODUCTION

Compressed Sensing or Compressive Sampling (CS) [1], [2] is an emerging field in signal processing. The theory of CS suggests to use only a few random linear measurement of a sparse signal (in a basis) for reconstructing the original signal. The mathematical model of noise free CS is:

\[ y = \Phi x \]  

where \( x = \Psi w \) is the original signal with length \( m \) and is sparse in the basis \( \Psi \) (i.e., \( \|w\|_0 < K \) and \( K \) is defined as sparsity level) and \( \Phi \) is an \( n \times m \) random measurement matrix where \( n < m \). For near perfect

¹Electrical engineering department, Sharif university of technology, Tehran, Iran.
²GIPSA-LAB, Grenoble, and Institut Universitaire de France, France.

This work has been partially funded by Iran NSF (INSF) under contract number 86/994, by Iran Telecom Research Center (ITRC), and also by center for International Research and Collaboration (ISMO) and French embassy in Tehran in the framework of a GundiShapour collaboration program.

First author: Hadi Zayyani, email: zayyani@ee.sharif.edu, Tel: +98 21 66164125, Fax: +98 21 66023261.
Second author: Masoud Babaie-zadeh, email: mbzadeh@sharif.edu, Tel: +98 21 66165925, Fax: +98 21 66023261.
Fourth author: Christian Jutten, email: christian.jutten@gipsa-lab.grenoble-inp.fr, Tel: +33 (0)4 76574351, Fax: +33 (0)4 76574790.
recovery, in addition to the signal sparsity, the incoherence of the random measurement matrix $\Phi$ with the basis $\Psi$ is needed. The incoherence is satisfied with high probability for some types of random matrices such as i.i.d Gaussian elements or i.i.d Bernoulli $\pm 1$ elements. Recent theoretical results show that under these two conditions (sparsity and incoherence), the original signal can be recovered from only a few linear measurements of the signal within a controllable error, even in the case of noisy measurements [1], [2], [3], [4].

In [3], some error bounds are introduced for reconstructing the original sparse (or compressible) signal in the noisy CS framework. In [4], the performance limits of noisy CS is investigated by definition of some performance metrics which are of Shannon Theoretic spirit. [1] considers the no noise CS and finds an upper bound on reconstruction error in terms of Mean Square Error (MSE) only for $\ell^1$-minimization recovery algorithm. But, [3] finds some upper bounds in the noisy CS and for general recovery algorithms. [4] is also investigated its own decoder which is derived based on joint typicality. Moreover, some information theoretic bounds are derived in [5].

In this paper, we derive a Bayesian Cramer-Rao Bound (BCRB) ([6], [7]), which is a lower bound, for noisy CS by a statistical view to the CS problem. This BCRB bounds the performance of any parametric estimator (whether biased or unbiased) of the sparse coefficient vector in terms of mean square estimation error [6], [7]. We also introduce the notion of blind CS in contrast to the traditional CS to whom we refer on the non-blind CS. We compute BCRB for both non-blind and blind CS, where in the latter, we do not know the measurement matrix in advance. In a related direction of research, a CRB is obtained for mixing matrix estimation in Sparse Component Analysis (SCA) [8].

II. NON-BLIND AND BLIND NOISY CS

Consider the noisy CS problem:

$$y = \Phi \Psi w + e = Dw + e$$

(2)

where $D = \Phi \Psi$, $w$ is a sparse vector and $e$ is a Gaussian zero-mean noise vector with the covariance $\sigma^2 I$. In CS framework, we want to estimate $w$, from which, $x = \Psi w$ can be reconstructed from the measurement vector $y$.

We nominate the traditional CS problem as non-blind CS since we know the basis $\Psi$ and the measurement matrix $\Phi$ and hence $D$ in advance. In some cases, we have no prior information about the signals in addition to their sparsity. As such, we do not know the basis $\Psi$, in which the signals are sparse. One application is a blind interceptor who intercepts the signals. The only information is that the signals have
been received are sparse in some unknown domain. In these cases, we nominate the problem as blind
CS which is inspired from the well known problem of Blind Source Separation (BSS). As such, each
measurement will be:

\[ y = \varphi^T \Psi w + e = d^T w + e \]  \hspace{1cm} (3)

where \( \varphi^T \) is the random measurement vector and a row of \( \Phi \) and \( d^T = \varphi^T \Psi \) is the corresponding row
in \( D \) and an unknown random vector.

III. BAYESIAN CRAMER-RAO BOUND

The Posterior Cramer-Rao Bound (PCRB) or Bayesian Cramer-Rao Bound (BCRB) of a vector of
parameters \( \theta \) estimated from data vector \( y \) is the inverse of the Fisher information matrix, and bounds
the estimation error in the following form [7]:

\[ E \left[ (\theta - \hat{\theta})(\theta - \hat{\theta})^T \right] \geq J^{-1} \]  \hspace{1cm} (4)

where \( \hat{\theta} \) is the estimate of \( \theta \) and \( J \) is the Fisher information matrix with the elements [7]:

\[ J_{ij} = E_{y,\theta} \left[ -\frac{\partial^2 \log p(y, \theta)}{\partial \theta_i \partial \theta_j} \right], \]  \hspace{1cm} (5)

where \( p(y, \theta) \) is the joint probability between the observations and the parameters. Unlike CRB, the
BCRB (4) is satisfied for any estimator (even for biased estimators) under some mild conditions [6], [7]
which we assume that are fulfilled in our problem. Using Bayes rule, the Fisher information matrix can
be decomposed into two matrices [7]:

\[ J = J_D + J_P, \]  \hspace{1cm} (6)

where \( J_D \) represents data information matrix and \( J_P \) represents prior information matrix which their
elements are [7]:

\[ J_{D,ij} = E_{y,\theta} \left[ -\frac{\partial^2 \log p(y|\theta)}{\partial \theta_i \partial \theta_j} \right] = E_{\theta}(J_{s,ij}) \]  \hspace{1cm} (7)

\[ J_{P,ij} = E_\theta \left[ -\frac{\partial^2 \log p(\theta)}{\partial \theta_i \partial \theta_j} \right] \]  \hspace{1cm} (8)

where \( J_s \triangleq E_{y|\theta}[ -\frac{\partial^2 \log p(y|\theta)}{\partial \theta_i \partial \theta_j} ] \) is the standard Fisher information matrix [9] and \( p(\theta) \) is the prior
distribution of the parameter vector.

In this paper, we use this BCRB for our problem because we have a sparse prior information about
the parameter which is estimated. We compute BCRB for two blind and non-blind cases.
A. Computing BCRB in non-blind CS

In the non-blind CS case, the matrices $\Phi$ and $\Psi$ are assumed to be known and $\Phi$ is a random matrix while $\Psi$ is a fixed basis matrix. Similar to [10], since $\Phi$ is assumed to be known and random, $\Phi$ can be added as an additional observation. Hence, the data information matrix elements $J_{Dij}$ from model (2) are of the form:

$$J_{Dij} = E_{y,w,\Phi} \left[ \frac{\partial^2 \log p(y, \Phi | w)}{\partial w_i \partial w_j} \right].$$

since $p(y, \Phi | w) = p(\Phi)p(y | \Phi, w)$, $p(\Phi)$ is independent of $w$ and $p(y | \Phi, w) = (2\pi \sigma_e^2)^{-n/2} \exp\left(-\frac{1}{2\sigma_e^2} \|y - Dw\|^2_2\right)$, we can write $\frac{\partial \log p(y, \Phi | w)}{\partial w} = -\frac{1}{2\sigma_e^2} (-2y^TD + 2D^TDw)$. So, we have $\frac{\partial \log p(y, \Phi | w)}{\partial w} = \frac{1}{\sigma_e^2} (y^TD)_{ij} - \frac{1}{\sigma_e^2} \sum_{r=1}^m g_{ir}w_r$ where $g_{ij}$ denotes the elements of the matrix $G = D^TD$. Hence, we have $\frac{\partial \log p(y, \Phi | w)}{\partial w} = -\frac{1}{\sigma_e^2} g_{ij}$. So, the expectation (9) will be $J_{Dij} = E_{y,w,\Phi} \left[ \frac{1}{\sigma_e^2} g_{ij} \right] = \frac{1}{\sigma_e^2} E_{\Phi} \{g_{ij}\} = J_{Dij} = \frac{1}{\sigma_e^2} \sum_{r=1}^n E_{\Phi} \{d_{ir}d_{rj}\}$. Some simple manipulations show that under assumption that the elements of $\Phi$ are zero mean and independent random variables, the data information matrix will be:

$$J_D = n \frac{\sigma_e^2}{\sigma_e^2} \Psi^T \Psi$$

where $\sigma_e^2 = E(\varphi_{ij}^2)$ is the variance of the random measurement matrix elements. If $\Psi$ is an orthonormal basis then $\Psi^T \Psi = I$ and hence $J_D = n \sigma_e^2 I$.

To compute the prior information matrix $J_P$ from (8), we should assume a sparse prior distribution for our parameter vector elements $w_i$. Similarly to [11], we assume $w_i$’s are independent and have a parameterized Gaussian distribution:

$$p(w_i) = \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left(-\frac{w_i^2}{2\sigma_i^2}\right).$$

In (11), the variance $\sigma_i^2$ enforce the sparsity of the corresponding coefficient: a small variance means that the coefficient is inactive and a large value means the activity of the coefficient. It can be easily seen that in this case, the prior information matrix is $J_P = \text{diag}\left(\frac{1}{\sigma_i^2}\right)$. Finally, for orthonormal bases for $\Psi$ and for prior distribution (11), the BCRB results in:

$$E \left[(w_i - \hat{w}_i)^2\right] \geq \left(\frac{\sigma_e^2}{\sigma_i^2} + \frac{1}{\sigma_i^2}\right)^{-1}.$$  

B. Computing BCRB in blind CS

In the blind CS case, the matrix $\Psi$ is not known in advance and hence the elements of matrix $D$ are random and unknown with zero mean. If we restrict ourselves to Gaussian measurements matrix elements $(\varphi_{ij}$ is a zero-mean Gaussian) then different measurement samples of $y$ are also Gaussian and independent
of each other. Hence, we can compute the data information matrix from only one measurement (3). Then, the information matrix elements \(J_{D_{ij}} = E_{y,w} \left[ \frac{\partial^2 \log p(y|w)}{\partial w_i \partial w_j} \right] \) will be equal to (refer to [9]):

\[
J_{D_{ij}} = E_{y,w} \left[ \frac{\partial \log p(y|w)}{\partial w_i} \frac{\partial \log p(y|w)}{\partial w_j} \right].
\]  

(13)

If the elements of \(\varphi\) are assumed to be random with a Gaussian distribution of zero mean and variance \(\sigma^2\) and the columns of the basis matrix \(\Psi\) have unit norms, then:

\[
p(y|w) = \frac{1}{\sqrt{2\pi\sigma^2(w)}} \exp(-\frac{y^2}{2\sigma^2(w)})
\]

(14)

where \(\sigma^2(w) \triangleq \sigma_e^2 + \sigma_\varphi^2 \|w\|^2_2\). Simple manipulations show:

\[
\frac{\partial \log p(y|w)}{\partial w_i} = -\frac{w_i \sigma^2}{\sigma^4(w)} (\sigma^2(w) - y^2)
\]

(15)

and from (13) we should compute:

\[
J_{D_{ij}} = \sigma^4 \int_w \frac{w_i w_j}{\sigma^2(w)} \left[ \int_y (\sigma^2(w) - y^2)^2 p(y|w) dy \right] p(w) dw
\]

(16)

where the internal integral is \(\int_y (\sigma^2(w) - y^2)^2 p(y|w) dy = m_4 - 2\sigma^2(w)m_2 + \sigma^4(w)\) in which \(m_2\) and \(m_4\) are the second and fourth order moments equal to \(m_2 = \sigma^2(w)\) and \(m_4 = 3\sigma^4(w)\). So, we have \(\int_y (\sigma^2(w) - y^2)^2 p(y|w) dy = 2\sigma^4(w)\) and then:

\[
J_{D_{ij}} = 2\sigma^4 \int_w \frac{w_i w_j}{\sigma^4(w)} p(w) dw
\]

(17)

where the off diagonal terms are zeros \(J_{D_{ij}} = 0, j \neq i\) because the integrand is an odd function. The diagonal terms are:

\[
J_{D_{ii}} = 2\sigma^4 \int_w \frac{w_i^2}{\left(\sigma^2 + \sigma_\varphi^2 \|w\|^2_2\right)} p(w) dw
\]

(18)

Following Appendix I, the diagonal elements are simplified as:

\[
J_{D_{ii}} = \frac{2\sigma^2}{m} (A_1 - \sigma_e^2 A_2)
\]

(19)

where \(A_1\) and \(A_2\) are defined and calculated in Appendix II.

The prior information matrix for BG distribution \(p(w_i) = p\delta(w_i) + (1-p)\frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{w_i^2}{2\sigma^2})\) is calculated in Appendix II:

\[
J_P = \frac{1 - p}{\sigma^2} I
\]

(20)

Finally, the Blind BCRB is calculated as:

\[
E \left[ (w_i - \hat{w}_i)^2 \right] \geq \left( 2 \frac{\sigma_e^2}{m} (A_1 - \sigma_e^2 A_2) + \frac{1 - p}{\sigma^2} \right)^{-1}
\]

(21)
IV. SIMULATION RESULTS

In this section, we compare the CRB’s with the results of some of the state-of-the-art algorithms for signal reconstruction in CS. In our simulations, we used sparse signals with the length $m = 512$ in the time domain where $\Psi = I$. We used a BG distribution with the probability of being nonzero equal to $1 - p = 0.1$ and the variance for nonzero coefficients is equal to $\sigma^2 = (0.5)^2$. So, in average there were 51 active coefficients. We used a Gaussian random measurement matrix with elements drawn from zero mean Gaussian distribution with variance equal to $\sigma_r^2 = 1$. The number of measurements are varied between 60 to 200. We computed the Mean Square Error (MSE) for sparse coefficient vector over 100 different runs of the experiment:

$$\text{MSE} \triangleq 10 \log_{10} \left( \frac{1}{100} \sum_{r=1}^{100} \| w_r - \hat{w}_r \|_2^2 \right)$$  \hspace{1cm} (22)

where $r$ is the experiment index. We compared this measure for various algorithms with the average value of BCRB for non-blind case which is equal to $\frac{1}{m} \text{trace}(J^{-1})$. The algorithms used for our simulation are Orthogonal Matching Pursuit (OMP) [12], Basis Pursuit (BP) [13], Bayesian Compressive Sampling (BCS) [14] and Smoothed-L0 (SL0) \footnote{We used the OMP code from \url{http://sparselab.stanford.edu} with 50 iterations, the BP code from \url{http://www.acm.caltech.edu/l1magic/l1eq-pd.m} with pdtol=1e-6 and its default parameters, the BCS code from \url{http://people.ee.duke.edu/~lihan/cs} with its default parameters and the SL0 code from \url{http://ee.sharif.edu/~SLzero} with parameters sigma-min=0.001 and sigma-decrease-factor=0.9.} [15]. We also computed the BCRB for blind case \footnote{May 25, 2010 DRAFT 21} to compare the BCRB’s in both blind and non-blind case. Figure 1 shows the results of the simulation. It can be seen that in the low number of measurements, there is a gap between the BCRB and the performance of algorithms while one of the algorithms approximately reaches the BCRB for large number of measurements. Moreover, the difference between the BCRB’s for the non-blind and blind cases are very large. It shows that the blind case needs much more linear measurements than the non-blind case.

V. CONCLUSIONS

In this paper, the CS problem is divided into non-blind and blind cases and the Bayesian Cramer-Rao bound for estimating the sparse vector of the signal was calculated in the two cases. The simulation results show a large gap between the lower bound and the performance of the practical algorithms when
the number of measurements are low. There was also a large gap between the BCRB in both non-blind
and blind cases. It also shows that in the blind CS framework, much more blind linear measurements of
the sparse signal are needed for perfect recovery of the signal.

APPENDIX I

COMPUTING THE INTEGRAL

Let define \( I_i = \int \frac{w^2}{(\sigma_e^2 + \sigma_r^2 ||w||^2)^2} p(w)dw \) and assume an equal prior distribution for all coefficients \( w_i \),
then all \( I_i \)'s are the same because of the symmetry of the integral. So, we can add all the integrals and
write:

\[
m \sigma_i^2 I_i = \int \frac{\sigma_e^2 ||w||^2}{(\sigma_e^2 + \sigma_r^2 ||w||^2)^2} p(w)dw =
\int p(w) \frac{p(w)}{(\sigma_e^2 + \sigma_r^2 ||w||^2)^2} d\mathbf{w} - \sigma_e^2 \int \frac{p(w)}{(\sigma_e^2 + \sigma_r^2 ||w||^2)^2} d\mathbf{w}
\]

Then, if we nominate the two above integrals as \( A_1 = \int \frac{p(w)}{(\sigma_e^2 + \sigma_r^2 ||w||^2)^2} d\mathbf{w} \) and \( A_2 = \int \frac{p(w)}{(\sigma_e^2 + \sigma_r^2 ||w||^2)^2} d\mathbf{w} \),
the integral \( I_i \) is computed as \( I_i = \frac{A_1}{m \sigma_e^2} (A_1 - \sigma_e^2 A_2) \). To compute \( A_1 \) and \( A_2 \), we approximate the joint probability distribution of coefficients as:

\[
p(w) = \prod_{i=1}^{m} p(w_i) \approx p^m \prod_{i=1}^{m} \delta(w_i) + p^{m-1} (1 - p) \sum_{r=1}^{m} \prod_{i=1,i \neq r}^{m} \frac{\delta(w_i)}{\sigma \sqrt{2\pi}} \exp \left(-\frac{w_i^2}{2\sigma^2} \right)
\]
This approximation is based on the assumption that the value of \((1-p)\) which is the activity probability is very small and so we can neglect the higher order powers of \((1-p)\). By this approximation, the two integrals will be approximately:

\[
A_1 = \frac{p^m}{\sigma_e^2} + \frac{mp^{m-1}(1-p)}{\sigma\sqrt{2\pi}} B_1
\]

\[
A_2 = \frac{p^m}{\sigma_i^2} + \frac{mp^{m-1}(1-p)}{\sigma\sqrt{2\pi}} B_2
\]

where the two integrals are

\[
B_1 = \int_{w}^{\infty} \frac{\exp(-w^2)}{(\sigma_e^2 + \sigma_i^2 w^2)} \, dw
\]

and

\[
B_2 = \int_{0}^{\infty} \frac{\exp(-x^2)}{(a^2 + x^2)^2} \, dx
\]

By change of variable

\[
x = \frac{w}{\sigma \sqrt{2}}
\]

the two integrals are equal to:

\[
B_1 = \frac{1}{\sqrt{2\sigma\sigma_e^2}} \int_{-\infty}^{\infty} \frac{\exp(-x^2)}{a^2 + x^2} \, dx = \frac{1}{\sqrt{2\sigma\sigma_e^2}} C_1
\]

\[
B_2 = \frac{1}{2\sqrt{2\sigma^3\sigma_i^4}} \int_{-\infty}^{\infty} \frac{\exp(-x^2)}{(a^2 + x^2)^2} \, dx = \frac{1}{2\sqrt{2\sigma^3\sigma_i^4}} C_2
\]

where \(a^2 = \frac{\sigma_i^2}{2\sigma^2\sigma_e^2}\). The above integrals are equal to

\[
C_1 = \frac{\pi}{a} \exp(a^2) [1 - \text{erf}(a)]
\]

\[
C_2 = \frac{\pi \exp(a^2)}{2a^3} \left[ 1 - 2a^2 + \frac{2\sqrt{\pi}}{\pi \exp(a^2)} - \text{erf}(a) + 2a^2 \text{erf}(a) \right]
\]

where \(\text{erf}(x)\) is the error function, defined as \(\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_{0}^{x} \exp(-t^2) \, dt\).

**APPENDIX II**

**PRIOR INFORMATION MATRIX FOR BG DISTRIBUTION**

Since the coefficients \(w_i\)'s are independent, the off diagonal terms \(J_{P_{ij}}, i \neq j\) are zero. Because of the independence of \(w_i\)'s, we can write \(J_{P_{ii}} = E_{w_i} \{ -\frac{\partial^2 \log p(w_i)}{\partial^2 w_i} \}\). To calculate this term, we use a Gaussian distribution with small variance \(\sigma_0^2\) instead of delta function \(\delta(w_i)\). So, the prior is:

\[
p(w_i) = A \exp \left( - \frac{w_i^2}{2\sigma_0^2} \right) + B \exp \left( - \frac{w_i^2}{2\sigma^2} \right)
\]

where \(A = \frac{p}{\sigma_0\sqrt{2\pi}}\), \(B = \frac{1-p}{\sigma\sqrt{2\pi}}\) and \(\sigma_0 \rightarrow 0\). The partial derivative can be calculated as:

\[
\frac{\partial^2 \log p(w_i)}{\partial w_i^2} = \frac{1}{p(w_i)} \frac{\partial^2 p(w_i)}{\partial w_i^2} - \frac{1}{p^2(w_i)} \left( \frac{\partial p(w_i)}{\partial w_i} \right)^2
\]

Hence, we have:

\[
J_{P_{ii}} = - \int_{-\infty}^{+\infty} \frac{\partial^2 p(w_i)}{\partial w_i^2} \, dw_i + \int_{-\infty}^{+\infty} \frac{1}{p(w_i)} \left( \frac{\partial p(w_i)}{\partial w_i} \right)^2 \, dw_i
\]

\(^2\text{We used Maple software to compute the integrals analytically.}\)
To compute the above integrals, the partial derivatives are 
\[
\frac{\partial p(w_i)}{\partial w_i} = -A w_i \exp(-\frac{w_i^2}{2\sigma^2}) - B w_i \exp(-\frac{w_i^2}{2\sigma^2})
\]
and 
\[
\frac{\partial^2 p(w_i)}{\partial w_i^2} = -\frac{A}{\sigma^2} \exp(-\frac{w_i^2}{2\sigma^2}) + \frac{A w_i^2}{\sigma^2} \exp(-\frac{w_i^2}{2\sigma^2}) - \frac{B}{\sigma^2} \exp(-\frac{w_i^2}{2\sigma^2}) + \frac{B w_i^2}{\sigma^2} \exp(-\frac{w_i^2}{2\sigma^2}).
\]
Simple calculations show that \(\int \frac{\partial^2 p(w_i)}{\partial w_i^2} dw_i = 0\) and hence:
\[
J_{P_{ii}} = \int_{-\infty}^{+\infty} \frac{[-A w_i \exp(-\frac{w_i^2}{2\sigma^2}) - B w_i \exp(-\frac{w_i^2}{2\sigma^2})]^2}{A \exp(-\frac{w_i^2}{2\sigma^2}) + B \exp(-\frac{w_i^2}{2\sigma^2})} dw_i \tag{28}
\]
where the above integral can be decomposed to three integrals which are \(D_1 = \int_{-\infty}^{+\infty} \frac{-A w_i^2 \exp(-\frac{w_i^2}{2\sigma^2})}{A \exp(-\frac{w_i^2}{2\sigma^2}) + B \exp(-\frac{w_i^2}{2\sigma^2})} dw_i\), \(D_2 = \int_{-\infty}^{+\infty} \frac{-B w_i^2 \exp(-\frac{w_i^2}{2\sigma^2})}{A \exp(-\frac{w_i^2}{2\sigma^2}) + B \exp(-\frac{w_i^2}{2\sigma^2})} dw_i\) and \(D_3 = \int_{-\infty}^{+\infty} \frac{-A w_i^2 \exp(-\frac{w_i^2}{2\sigma^2}) - B w_i^2 \exp(-\frac{w_i^2}{2\sigma^2})}{A \exp(-\frac{w_i^2}{2\sigma^2}) + B \exp(-\frac{w_i^2}{2\sigma^2})} dw_i\). Since we have a term \(w_i^2\) in the numerator of the above integrals and the Gaussian term with small variance is large near zero, we can neglect the Gaussian term with small variance (delta function) in the denominator. So, the integrals \(D_1\) and \(D_2\) with neglecting this term will be approximately zero. We verify this approximation in the simulation results by computing these integrals numerically. Finally, the third integral will be approximately \(D_3 \approx \int_{-\infty}^{+\infty} \frac{-A w_i^2 \exp(-\frac{w_i^2}{2\sigma^2})}{B \exp(-\frac{w_i^2}{2\sigma^2})} dw_i\). Calculating this integral results is \(J_{P_{ii}} \approx D_3 \approx \frac{1-p^2}{\sigma^2}\).

REFERENCES

[1] E. Candès and T. Tao, “Near optimal signal recovery from random projections: Universal encoding strategies?,” IEEE Trans. Inform. Theory., vol. 52, no. 12, pp. 5406–5425, Dec 2006.
[2] D. L. Donoho, “Compressed Sensing,” IEEE Trans. Info. Theory, vol. 52, no. 4, pp. 1289–1306, April 2006.
[3] J. Haupt and R. Nowak, “Signal reconstruction from noisy random projections,” IEEE Trans. Inform. Theory., vol. 52, no. 9, pp. 4036–4048, Sep 2006.
[4] M. Akcakaya and V. Tarokh, “Shannon theoretic limits on noisy compressive sampling,” Submitted to IEEE Trans. Inform. Theory., 2007.
[5] S. Aeron, M. Zhao, and V. Saligrama, “Information theoretic bounds to sensing capacity of sensor networks under fixed SNR,” in IEEE Information Theory Workshop, pp. 84–89, Lake Tahoe, CA, Sep 2007.
[6] H. L. van Trees, Detection, estimation and modulation theory, Wiley, 1968.
[7] P. Tichavsky, C. H. Muravchik and A. Nehorai, “Posterior Cramer-Rao bounds for discrete-time nonlinear filtering,” IEEE Trans. Signal Process., vol. 46, no. 5, pp. 1386–1395, May 1998.
[8] H. Zayyani, M. Babaie-Zadeh, F. Haddadi, and C. Jutten, “On the Cramer Rao Bound for estimating the mixing matrix in noisy Sparse Component Analysis,” IEEE Signal Process Letters., vol. 15, pp. 609–612, 2008.
[9] S. M. Kay, Fundamentals of statistical signal processing: Estimation theory, Prentice-Hall, 1993.
[10] A. Wiesel, Y. C. Eldar and A. Yeredor, “Linear Regression with Gaussian Model Uncertainty: Algorithms and Bounds,” IEEE Trans. on Signal Proc., vol. 56, no. 6, pp. 2194–2205, June 2008.
[11] D. Wipf and B. D. Rao, “Sparse Bayesian Learning for basis selection,” IEEE Trans. on Signal Proc., vol. 52, no. 8, pp. 2153–2164, 2004.
[12] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, “Orthogonal matching pursuit: recursive function approximation with application to wavelet decomposition,” proceeding of the 27th Annual Asilomar Conf. Signals, systems, and Computers, vol. 1, pp. 40–44. 1993.

[13] S. S. Chen, D. L. Donoho, and M. A. Saunders, “Atomic decomposition by basis pursuit,” SIAM Journal on Scientific Computing, vol. 20, no. 1, pp. 33–61. 1998.

[14] S. Ji, Y. Xue, and L. Carin, “Bayesian Compressive Sensing,” IEEE Trans. on Signal Proc., vol. 56, no. 6, pp. 2346–2356, June 2008.

[15] G. H. Mohimani, M. Babaie-Zadeh, and C. Jutten, “A fast approach for overcomplete sparse decomposition based on smoothed $\ell^0$norm,” IEEE Trans. on Signal Proc., vol. 57, no. 1, pp. 289–301, January 2009.