Proceedings of the World Congress on Engineering and Computer Science 2012 Vol I
WCECS 2012, October 24-26, 2012, San Francisco, USA

Minimax MMSE Estimator for Sparse System

Hongqing Liu, Mandar Chitre

***Abstract*—In this work, we consider a minimum mean square error (MMSE) estimator utilizing compressed sensing (CS) idea when the system is underdetermined. First, we attempt to directly solve the nonconvex problem in an alternating way. However, this does not guarantee the optimality. Second, in a more efficient way, we reformulate the problem in the context of minimax framework using the worst case optimization technique. And later on, based on duality theory, we transform this minimax problem into a semidefinite programming problem (SDP). Numerical results show that utilizing CS idea indeed improves MMSE estimator when the signal is sparse.**

***Index Terms*—minimum mean square error (MMSE), compressed sensing (CS), minimax optimization, semidefinite programming (SDP).**

I. Introduction

PARAMETER estimation is the most fundamental problem in variety of engineering applications, such as signal processing, communication, imaging, etc. Consider a linear system of $\mathbf{y} = \mathbf{Hx} + \mathbf{n}$, where $\mathbf{H}$ is observation matrix, $\mathbf{x}$ is the parameter to be estimated, $\mathbf{n}$ is measurement noise and $\mathbf{y}$ is the received data, the objective is to estimate the unknown parameter $\mathbf{x}$ based on the received data $\mathbf{y}$. Generally, methods of parameter estimation can be categorized into two groups [1]. One is so-called classic approaches, which mean that estimators are derived considering parameter $\mathbf{x}$ is deterministic. A well known estimator in this category is least square (LS), i.e., $\hat{\mathbf{x}} = \text{minimize } \|\mathbf{y} - \mathbf{Hx}\|_2$. Another one is assuming the availability of the distribution of noise $\mathbf{n}$ maximum likelihood estimator (MLE), which is obtained by maximizing the likelihood function, i.e., $\hat{\mathbf{x}} = \text{maximize } p(\mathbf{y}|\mathbf{x})$. The second category is so-called Bayesian methods, which assume that the parameter $\mathbf{x}$ is not deterministic, but a random variable and the prior information on it is available. The famous one in this category is minimum mean square error (MMSE) by calculating the posterior mean, i.e., $\hat{\mathbf{x}}_{\text{MMSE}} = \int \mathbf{x} p(\mathbf{x}|\mathbf{y}) d\mathbf{x}$, where $p(\mathbf{x}|\mathbf{y})$ is posterior distribution. In most cases, MMSE is difficult to obtain in closed-form solution due to the potential multi-dimension integration and dependency on the unknown parameter $\mathbf{x}$. In case of the linear and Gaussian system, the closed-form expression is obtainable by so-called linear MMSE (LMMSE) estimator. Another one is called maximum a posteriori (MAP) estimator. The MAP can be calculated by maximizing the posterior distribution, i.e., $\hat{\mathbf{x}}_{\text{MAP}} = \text{maximize } (\log p(\mathbf{y}|\mathbf{x}) + \log p(\mathbf{x}))$, where $p(\mathbf{x})$ is prior distribution on parameter $\mathbf{x}$.

The above discussions are based on the fact that the system we consider is complete. In case of a underdetermined system, it is impossible to find the unique solution without any prior information. In recent years of research development, compressed sensing (CS) [2], [3] has become a powerful technique to reconstruct signal for the underdertimend system when the signal is sparse. For a underdetermiend system of $\mathbf{y} = \mathbf{H}_{N\times M}\mathbf{x} + \mathbf{n}$ where $N \ll M$, CS states that it only needs minimum $N = K\log(M/K)$ measurements to reconstruct the signal correctly by using ℓ_1-norm constraint convex optimization. Sparse signals exit in many research areas, such in image processing, seismic signal processing and underwater acoustic communications, etc. This idea will be utilized in MMSE estimator when the signal is sparse to handle the underdetermiend system.

After parameter estimation, to measure the estimation performance, one usually employs the MSE criterion. Therefore, MMSE estimator is the optimal one in the MSE sense to be used to do the estimation. However, like we mentioned earlier, the problem with MMSE estimator is that it usually dependents on the unknown parameter $\mathbf{x}$, which prevents us from using it [1]. In this work, first, we consider the signal is bounded, i.e. $\|\mathbf{x}\|_2^2 < L$, and then we reformulate the problem in worst case MSE optimization under minimax optimization framework. The duality theory is utilized to transform the minimax optimization into a solvable semidefinite programming (SDP). Second, we consider another prior information on signal, namely sparsity, is available to us. By sparse, we mean that the most of entries of $\mathbf{x}$ is zero. Based on compressed sensing (CS) concept [2], [3], ℓ_1-norm on parameter i.e., $\|\mathbf{x}\|_1$ is utilized to explore the sparsity. Therefore, we reformulate the problem under the framework of minimax technique by taking both ℓ_2-norm and ℓ_1-norm into account. And later on, this problem can be as well transformed into a SDP by using duality approach.

The rest of paper is organized as follows. The problem formulation is given in Section II. In Section III, the proposed methods are presented to solve the problem. Section IV, the proposed algorithm based on compressed sensing is developed. In Section V, the simulation and filed trial experiments are conducted to verify the effectiveness of the proposed method. In Section IV, the conclusions are drawn.

II. Problem Formulation

Suppose we have a linear estimator $\mathbf{G}$ to estimate the unknown parameter $\mathbf{x}$ to the following linear system

$$\mathbf{y} = \mathbf{Hx} + \mathbf{n} \tag{1}$$

that means the estimate can be obtained by a linear transformation, i.e. $\hat{\mathbf{x}} = \mathbf{Gy}$. The most common measure to evaluate estimation performance is to calculate mean square error (MSE) [4], [5]. In our case, the MSE can be computed as

$$\begin{aligned}
&\mathbf{E}\left\{\|\mathbf{x} - \hat{\mathbf{x}}\|^2\right\} \\
&= \mathbf{E}\left\{[(\mathbf{I} - \mathbf{GH})\mathbf{x} + \mathbf{Gn}]^{\mathrm{H}}[(\mathbf{I} - \mathbf{GH})\mathbf{x} + \mathbf{Gn}]\right\} \\
&= \mathbf{x}^{\mathrm{H}}(\mathbf{I} - \mathbf{GH})^{\mathrm{H}}(\mathbf{I} - \mathbf{GH})\mathbf{x} + \mathbf{E}\left\{\mathbf{n}^{\mathrm{H}}\mathbf{G}^{\mathrm{H}}\mathbf{Gn}\right\} \\
&= \mathbf{x}^{\mathrm{H}}(\mathbf{I} - \mathbf{GH})^{\mathrm{H}}(\mathbf{I} - \mathbf{GH})\mathbf{x} + \mathrm{Tr}(\mathbf{GC}_w\mathbf{G}^{\mathrm{H}})
\end{aligned} \tag{2}$$

Manuscript received May 28, 2012; revised July 02, 2012.

Hongqing Liu and Mandar Chitre are with ARL, Tropical Marine Science Institute, National University of Singapore, 12A Kent Ridge Road, Singapore 119222. e-mails: hqliuster@gmail.com,mandar@arl.nus.edu.sg.

where $\mathbf{E}(\cdot)$ $\mathrm{Tr}(\cdot)$ are the expectation operation and the trace operator, respectively. The derivation in (2) has used the fact of $\mathbf{s}^{\mathrm{H}}\mathbf{s} = \mathrm{Tr}(\mathbf{s}\mathbf{s}^{\mathrm{H}})$. To compute $\mathbf{G}$, the MMSE estimator can be written as follows

$$\text{minimize} \quad \mathrm{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^{\mathrm{H}}) + \mathbf{x}^{\mathrm{H}}(\mathbf{I}-\mathbf{G}\mathbf{H})^{\mathrm{H}}(\mathbf{I}-\mathbf{G}\mathbf{H})\mathbf{x} \tag{3}$$

The solution of this direct optimization is usually not attainable since it involves the unknown parameter $\mathbf{x}$. In the following section, we will detail the procedures on how to minimize (3) by using worst case optimization technique.

III. Proposed Approaches

A. Compressed Sensing

Prior to going to how solve problem (3), we would like to present a brief introduction on compressed sensing (CS) whose idea will be utilized later. CS explores the signal sparsity to estimate the signal when the system is underdetermined. Denote the received signal $\mathbf{y}$, an $N \times 1$ vector, and a measurement matrix $\mathbf{H}_{N\times M}$, such that

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \tag{4}$$

where $\mathbf{x}_{M\times 1}$ is the signal we want to estimate and we assume the system is underdetermined, namely $N \ll M$. The signal $\mathbf{x}$ is called K-$sparse$ if coefficients of $\mathbf{x}$ only have K nonzeros and $M-K$ zeros. The goal is to reconstruct the $\mathbf{x}$ based on received data $\mathbf{y}$ given that the signal is sparse. The traditional methods would fail because we have less measurements than the number of variables we would like to estimate. However, CS exploits the sparsity to obtain the solution though the following optimization problem [3], [6]

$$\begin{array}{ll} \underset{\mathbf{x}}{\text{minimize}} & \|\mathbf{x}\|_1 \\ \text{subject to} & \|\mathbf{y}-\mathbf{H}\mathbf{x}\|_2 < \epsilon \end{array} \tag{5}$$

where $\|\cdot\|_1$ and $\|\cdot\|_2$ represent the ℓ_1-norm and ℓ_2-norm, respectively. This is a convex optimization problem and studies have shown that we only need minimum $N = K\log(M/K)$ measurements to estimate the signal correctly. Sometimes, one rewrites (5) as form of $\|\mathbf{y}-\mathbf{H}\mathbf{x}\|_2 + \varepsilon\|\mathbf{x}\|_1$. If we replace the second norm by a ℓ_2-norm, it becomes $\|\mathbf{y}-\mathbf{H}\mathbf{x}\|_2 + \varepsilon\|\mathbf{x}\|_2$, the solution of which is the least norm solution. Studies have shown that the latter one will not work well for this underdetermined system, but the former one will. The core concept behind CS is that ℓ_1-norm is utilized to exploit the sparsity to make the estimation possible. Other 'norm', like ℓ_p-norm $(p < 1)$ can be used as well. This core concept will be utilized to develop a new MMSE estimator to improve the performance when signal is sparse and the system is underdetermined.

B. Alternating Convex Optimization

One obvious way to exploit the sparsity in MMSE estimator is that we simply add the ℓ_1-norm regularization term into (3) as follows

$$\underset{\mathbf{G},\mathbf{x}}{\text{minimize}} \left\{\mathrm{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^{\mathrm{H}}) + \mathbf{x}^{\mathrm{H}}(\mathbf{I}-\mathbf{G}\mathbf{H})^{\mathrm{H}}(\mathbf{I}-\mathbf{G}\mathbf{H})\mathbf{x} + \|\mathbf{x}\|_1\right\} \tag{6}$$

Solving this problem (6) is a challenge since it is a nonconvex optimization. However, we can see that the problem is convex in $\mathbf{x}$ when $\mathbf{G}$ is fixed and vice versa. Therefore, in an intuitive way, we can minimize (6) in an alternative way, which works as follows: First, we optimize (6) over $\mathbf{x}$ fixing $\mathbf{G}$ at the current value, and we optimize (6) over $\mathbf{G}$ fixing $\mathbf{x}$ at the current value, and then we iterate the same procedure. The alternating convex optimization [7] to solve (6) proceeds in an iterative fashion as follows:

- Step 1: $\mathbf{x}_k \longleftarrow \text{argmim}(6)|_{\mathbf{G}_{k-1}}$,
- Step 2: $\mathbf{G}_k \longleftarrow \text{argmim}(6)|_{\mathbf{x}_k}$,
- Step 3: iterate Steps 1 and 2.

At each iteration, we solve a convex optimization. Therefore, we can say that objective function will go down at each step, and then solution will be bounded. This approach is named MMSE AL in this work.

C. Worst Case Optimization

In alternating optimization, we cannot always guarantee the global convergence since it directly deals with noncovex problem. The worst case optimization has been widely used in robust optimization problem. In this section, we reformulate the problem in the context of minimax optimization and show how to transform it into a solvable problem, namely SDP, by using duality theory.

1) Worst case over ℓ_2-norm constraint: First we consider that $\mathbf{x}$ is bounded, namely, $\|\mathbf{x}\|_2^2 < L$. With this constraint, the original problem of (3) can be reexpressed as

$$\underset{\mathbf{G}}{\text{minimize}} \left\{\max_{\|\mathbf{x}\|_2^2<L} \mathrm{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^{\mathrm{H}}) + \mathbf{x}^{\mathrm{H}}(\mathbf{I}-\mathbf{G}\mathbf{H})^{\mathrm{H}}(\mathbf{I}-\mathbf{G}\mathbf{H})\mathbf{x}\right\}. \tag{7}$$

Focusing on inner problem of (7), we have the following Lagrangian

$$L(\mathbf{x},\lambda) = -\mathbf{x}^{\mathrm{H}}(\mathbf{I}-\mathbf{G}\mathbf{H})^{\mathrm{H}}(\mathbf{I}-\mathbf{G}\mathbf{H})\mathbf{x} + \lambda^{\mathrm{H}}(\mathbf{x}^{\mathrm{H}}\mathbf{x}-L). \tag{8}$$

Lemma 1 (Minimal point of a quadratic function): The optimal value of the following quadratic optimization problem

$$\text{minimize} \;\; \mathbf{x}^{\mathrm{H}}\mathbf{A}\mathbf{x} + 2\mathbf{q}^{\mathrm{H}}\mathbf{x} + r$$

is

$$\mathbf{x}^* = \begin{cases} r - \mathbf{q}^{\mathrm{H}}\mathbf{A}^{-1}\mathbf{q}, & \mathbf{A} \succeq \mathbf{0}, \\ -\infty, & \text{otherwise} \end{cases}$$

The dual function of (8) can be expressed as

$$\begin{aligned} g(\lambda) &= \inf_{\mathbf{x}} L(\mathbf{x},\lambda) \\ &= \inf_{\mathbf{x}} \left\{-\mathbf{x}^{\mathrm{H}}(\mathbf{I}-\mathbf{G}\mathbf{H})^{\mathrm{H}}(\mathbf{I}-\mathbf{G}\mathbf{H})\mathbf{x} + \lambda^{\mathrm{H}}(\mathbf{x}^{\mathrm{H}}\mathbf{x}-L)\right\} \\ &= \begin{cases} -\lambda L, & \boldsymbol{\lambda}\mathbf{I}-\mathbf{W} \succeq \mathbf{0} \\ -\infty, & \text{otherwise} \end{cases} \end{aligned} \tag{9}$$

where $\mathbf{W} = (\mathbf{I}-\mathbf{G}\mathbf{H})^{\mathrm{H}}(\mathbf{I}-\mathbf{G}\mathbf{H})$ and $'\succeq'$ means the matrix is positive semidefinite. The solution of (9) is obtained based on Lemma 1.

ISBN: 978-988-19251-6-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

Lemma 2 (Schur Complement): Let

$$\mathbf{M} = \begin{bmatrix} \mathbf{A} & \mathbf{B^H} \\ \mathbf{B} & \mathbf{C} \end{bmatrix}$$

be a Hermitian matrix with $\mathbf{C} \succeq \mathbf{0}$. Then, $\mathbf{M} \succeq \mathbf{0}$ if only if the Schur complement is positive semidefinite, i.e., $\mathbf{A} - \mathbf{B^H}\mathbf{C}^{-1}\mathbf{B} \succeq \mathbf{0}$.

Based on Lemma 2, the dual optimization can be formulated as

$$\begin{aligned} &\text{maximize } g(\lambda) = \text{minimize } \lambda L \\ &\text{subject to } \lambda \geq 0, \\ &\begin{bmatrix} \lambda\mathbf{I} & (\mathbf{I}-\mathbf{GH})^{\mathrm{H}} \\ (\mathbf{I}-\mathbf{GH}) & \mathbf{I} \end{bmatrix} \succeq \mathbf{0}. \end{aligned} \tag{10}$$

Based on (10), the original minimax problem in (7) now can be cast into the following SDP

$$\begin{aligned} &\text{minimize } \lambda L + \mathbf{g}^{\mathrm{H}}\mathbf{g} \\ &\text{subject to } \lambda \geq 0, \\ &\begin{bmatrix} \lambda\mathbf{I} & (\mathbf{I}-\mathbf{GH})^{\mathrm{H}} \\ (\mathbf{I}-\mathbf{GH}) & \mathbf{I} \end{bmatrix} \succeq \mathbf{0}. \end{aligned} \tag{11}$$

with variables $\mathbf{G}, \lambda$, and where $\mathbf{g} = \mathrm{vec}(\mathbf{GC}^{1/2})$, and $\mathrm{vec}(\cdot)$ means vectorization operation. This result is the simialr to that reported in [4]. This SDP reformulation is easy to solve in polynomial time using interior point method [7]. In the end, the estimate can be obtained by $\hat{\mathbf{x}} = \mathbf{Gy}$. This approach is named MMSE with 2-norm constraint (MMSE-2).

2) Worst case over both ℓ_1 and ℓ_2-norms constraint: We would also like to take advantage of the fact that the signal is sparse. That means ℓ_1-norm constraint needs to be considered. In that regard, we have the following worst case optimization problem

$$\begin{aligned} &\underset{\mathbf{G}}{\text{minimize}} \max_{\|\mathbf{x}\|_1<\beta, \|\mathbf{x}\|_2^2<L} \\ &\left\{\mathrm{Tr}(\mathbf{GC}_w\mathbf{G}^{\mathrm{H}}) + \mathbf{x}^{\mathrm{H}}(\mathbf{I}-\mathbf{GH})^{\mathrm{H}}(\mathbf{I}-\mathbf{GH})\mathbf{x}\right\} \end{aligned} \tag{12}$$

First let us consider the inner optimization in (12), which is equivalent to the following problem

$$\underset{\|\mathbf{x}\|_2^2<L}{\text{maximize}} \ \mathbf{x}^{\mathrm{H}}(\mathbf{I}-\mathbf{GH})^{\mathrm{H}}(\mathbf{I}-\mathbf{GH})\mathbf{x} - \|\mathbf{x}\|_1. \tag{13}$$

By introducing a new variable $\mathbf{t}$, we have the equivalent problem to (13) as follows

$$\begin{aligned} &\underset{\mathbf{x},\mathbf{t}}{\text{maximize}} \ \mathbf{x}^{\mathrm{H}}(\mathbf{I}-\mathbf{GH})^{\mathrm{H}}(\mathbf{I}-\mathbf{GH})\mathbf{x} - \mathbf{1}^{\mathrm{H}}\mathbf{t} \\ &\text{subject to } -\mathbf{t} \preceq \mathbf{x} \preceq \mathbf{t}, \mathbf{x}^{\mathrm{H}}\mathbf{x} < L. \end{aligned} \tag{14}$$

The Lagrangian of (14) can be written as

$$\begin{aligned} L(\mathbf{x},\mathbf{t},\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2,\lambda_3) = &-\mathbf{x}^{\mathrm{H}}(\mathbf{I}-\mathbf{GH})^{\mathrm{H}}(\mathbf{I}-\mathbf{GH})\mathbf{x} + \mathbf{1}^{\mathrm{H}}\mathbf{t} \\ &+\boldsymbol{\lambda}_1^{\mathrm{H}}(\mathbf{x}-\mathbf{t}) - \boldsymbol{\lambda}_2^{\mathrm{H}}(\mathbf{x}+\mathbf{t}) + \lambda_3^{\mathrm{H}}(\mathbf{x}^{\mathrm{H}}\mathbf{x}-L). \end{aligned} \tag{15}$$

Lemma 3 (Minimal point of an affine function): The optimal value of the following affine optimization problem

$$\text{minimize } \mathbf{a}^{\mathrm{H}}\mathbf{x} + b$$

is

$$\mathbf{x}^* = \begin{cases} b, & \mathbf{a} = \mathbf{0}, \\ -\infty, & \text{otherwise} \end{cases}$$

Therefore the dual function is

$$\begin{aligned} &g(\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2,\lambda_3) \\ &= \inf_{\mathbf{x},\mathbf{t}} L(\mathbf{x},\mathbf{t},\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2,\lambda_3) \\ &= \inf_{\mathbf{x},\mathbf{t}}\{-\mathbf{x}^{\mathrm{H}}(\mathbf{I}-\mathbf{GH})^{\mathrm{H}}(\mathbf{I}-\mathbf{GH})\mathbf{x} + \mathbf{1}^{\mathrm{H}}\mathbf{t} + \boldsymbol{\lambda}_1^{\mathrm{H}}(\mathbf{x}-\mathbf{t}) \\ &- \boldsymbol{\lambda}_2^{\mathrm{H}}(\mathbf{x}+\mathbf{t}) + \lambda_3^{\mathrm{H}}(\mathbf{x}^{\mathrm{H}}\mathbf{x}-L)\} \\ &= \begin{cases} -\lambda_3 L - \frac{1}{4}(\boldsymbol{\lambda}_1-\boldsymbol{\lambda}_2)^{\mathrm{H}}(\lambda_3\mathbf{I}-\mathbf{W})^{-1}(\boldsymbol{\lambda}_1-\boldsymbol{\lambda}_2), \\ \quad \mathbf{1}-\boldsymbol{\lambda}_1-\boldsymbol{\lambda}_2 = \mathbf{0}, \lambda_3\mathbf{I}-\mathbf{W} \succeq \mathbf{0} \\ -\infty, \\ \quad \text{otherwise} \end{cases} \end{aligned} \tag{16}$$

where $\mathbf{W} = (\mathbf{I}-\mathbf{GH})^{\mathrm{H}}(\mathbf{I}-\mathbf{GH})$. The solution of (16) is obtained by both Lemmas 1 & 3. Finally, the dual problem can be expressed as

$$\begin{aligned} &\text{maximize } g(\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2,\lambda_3) = \text{maximize } \gamma \\ &\text{subject to} \\ &\mathbf{1}-\boldsymbol{\lambda}_1-\boldsymbol{\lambda}_2 = \mathbf{0}, \\ &\lambda_3\mathbf{I}-\mathbf{W} \succeq \mathbf{0}, \\ &-\lambda_3 L - (\boldsymbol{\lambda}_1-\boldsymbol{\lambda}_2)^{\mathrm{H}}(\lambda_3\mathbf{I}-\mathbf{W})^{-1}(\boldsymbol{\lambda}_1-\boldsymbol{\lambda}_2) \geq \gamma, \\ &\boldsymbol{\lambda}_1 \succeq \mathbf{0}, \boldsymbol{\lambda}_2 \succeq \mathbf{0}, \lambda_3 \geq 0. \end{aligned} \tag{17}$$

Based on Lemma 2, the dual problem (17) can be transformed into the following SDP problem

$$\begin{aligned} &\text{minimize } -\gamma \\ &\text{subject to} \\ &\mathbf{1}-\boldsymbol{\lambda}_1-\boldsymbol{\lambda}_2 = \mathbf{0}, \\ &\boldsymbol{\lambda}_1 \succeq \mathbf{0}, \boldsymbol{\lambda}_2 \succeq \mathbf{0}, \lambda_3 \geq 0, \\ &\begin{bmatrix} \lambda_3 L - \gamma & (\boldsymbol{\lambda}_1-\boldsymbol{\lambda}_2)^{\mathrm{H}} & \mathbf{0} \\ (\boldsymbol{\lambda}_1-\boldsymbol{\lambda}_2) & \lambda_3\mathbf{I} & (\mathbf{I}-\mathbf{GH})^{\mathrm{H}} \\ \mathbf{0} & (\mathbf{I}-\mathbf{GH}) & \mathbf{I} \end{bmatrix} \succeq \mathbf{0}. \end{aligned} \tag{18}$$

Therefore, the original minimax problem of (12) can be now expressed as

$$\begin{aligned} &\text{minimize } -\gamma + \mathbf{g}^{\mathrm{H}}\mathbf{g} \\ &\text{subject to} \\ &\mathbf{1}-\boldsymbol{\lambda}_1-\boldsymbol{\lambda}_2 = \mathbf{0}, \\ &\boldsymbol{\lambda}_1 \succeq \mathbf{0}, \boldsymbol{\lambda}_2 \succeq \mathbf{0}, \lambda_3 \geq 0, \\ &\begin{bmatrix} \lambda_3 L - \gamma & (\boldsymbol{\lambda}_1-\boldsymbol{\lambda}_2)^{\mathrm{H}} & \mathbf{0} \\ (\boldsymbol{\lambda}_1-\boldsymbol{\lambda}_2) & \lambda_3\mathbf{I} & (\mathbf{I}-\mathbf{GH})^{\mathrm{H}} \\ \mathbf{0} & (\mathbf{I}-\mathbf{GH}) & \mathbf{I} \end{bmatrix} \succeq \mathbf{0}. \end{aligned} \tag{19}$$

with variables $\mathbf{G}, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \lambda_3$. Compared to the formulation (11), the dual variables $\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2$ will be exploring the sparsity to make estimation possible when the system is underdetermined. In the end, the estimate can be obtained again by $\hat{\mathbf{x}} = \mathbf{Gy}$. This approach is named MMSE with 1&2-norm constraint (MMSE-1-2).

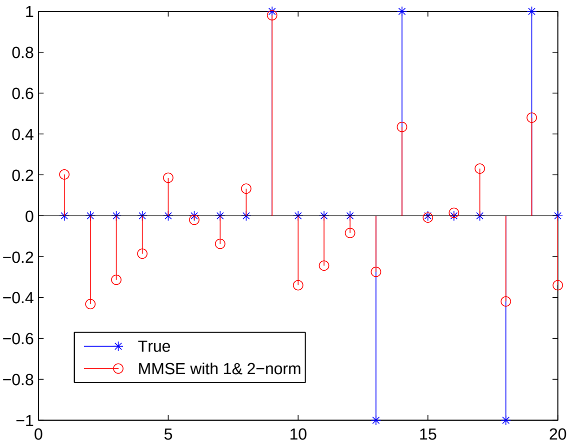
IV. Numerical Studies

In this section, we conduct studies on simulated and recorded data to demonstrate the performance of the proposed methods.

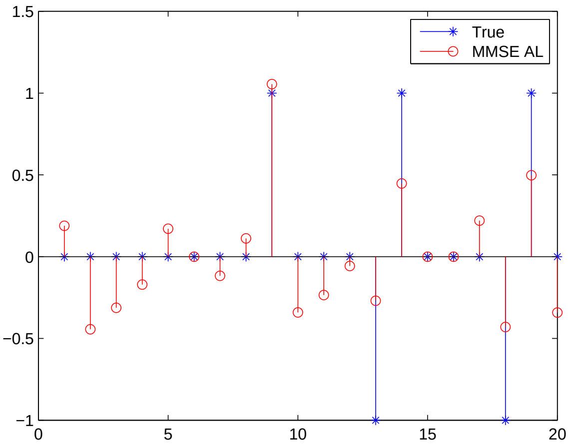
A. Simulated example

We simulate data from system model of (4), where $\mathbf{H}$ is generated by Gaussian random matrix with $N = 10, M = 20$

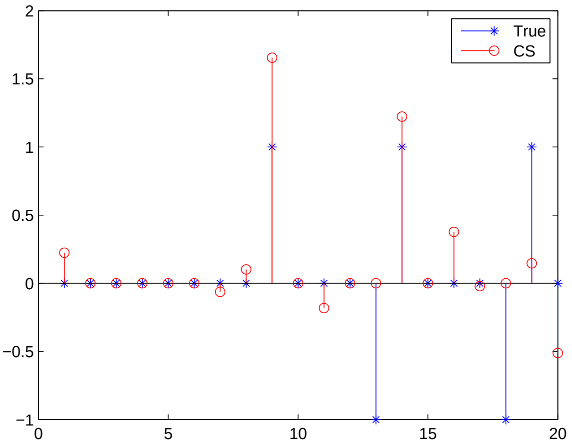
ISBN: 978-988-19251-6-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)



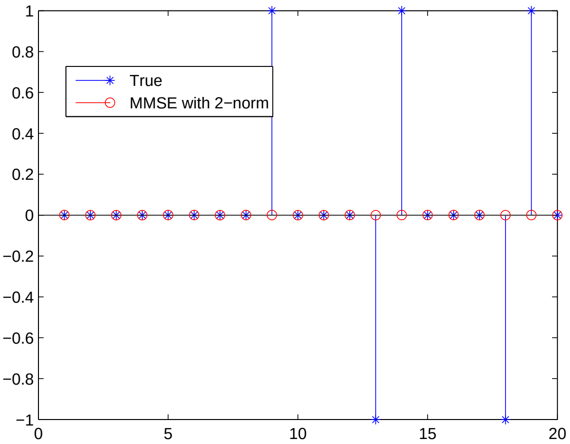
(a) MMSE estimator with ℓ_1 and ℓ_2-norm constraints



(b) MMSE estimator with Alternating optimization



(c) Compressed sensing estimation



(d) MMSE with ℓ_2-norm constraint

Fig. 1: Estimates

to indicate that the system is underdetermined. The signal $\mathbf{x}$ is generated randomly with sparsity $K = 5$ shown in Figure 1. The noise is Gaussian distributed. From Figure 1, it is observed that MMSE-1-2 and MMSE AL have almost the same performance, which is confirmed later by MSE in Figure 2. The MSEs are obtained on average of 100 independent runs. However, the problem with alternating optimization is that first there is no optimality guaranteed and in this case, the initialization of $\mathbf{G}$ is smartly chosen, which is the general inverse of the measurement matrix. If $\mathbf{G}$ is randomly selected, it will not converge in our test. It is also observed that MMSE with 2-norm dose not work due to the underdeterminability.

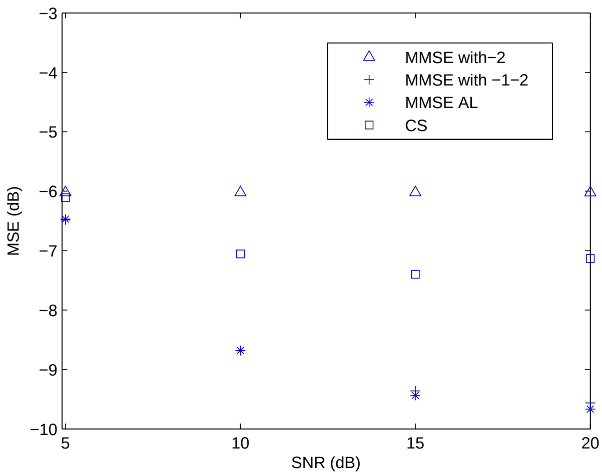


Fig. 2: MSEs versus SNR (dB).

B. Experimental data

For this test, we perform the algorithms on channel estimation, which can be stated as follows. Let $s(n), n = 1, 2, \cdots, N$ be the training sequence, that is transmitted through a sparse channel. The received signal samples can be modeled as

$$y(t) = \sum_{i=1}^{L} s^*(t-i)h(i) + n(t), t = 1, 2, \cdots, N \quad (20)$$

where $h(i), i = 1, 2, \cdots, L$ represents the channel response. The $n(t)$ is the observation Gaussian noise with zero mean and variance σ_n^2.
Rewrite (20) in a matrix form as

$$y(t) = \mathbf{S}^{\mathrm{H}}\mathbf{h} + n(t) \tag{21}$$

where $\mathbf{S} = [s_t, \cdots, s_{t-L+1}]^T$ is training symbols vector formed by the training signal acting like measurement matrix, $\mathbf{y}$ is the observation vector and $\mathbf{h} = [h_1\ h_2\ \cdots\ h_L]^T$ is the channel vector of interest. In this test, the channel response was recorded by a single transmitter and receiver communication link. The data were collected in Singapore water on April 21st, 2010, shown in Figure 3, in which the sparsity of channel is present. The transmitted signal was a BPSK signal. The sampling frequency was 20kHz and bandwidth was 5kHz. The length of channel in this test is chosen to be $L = 20$ and only 6 symbols are used as training symbols to indicate that the system is underdetermined. As we can see that CS cannot reconstruct the signal well, but on contrary MMSE-1-2 and MMSE AL estimate the channel response pretty well.

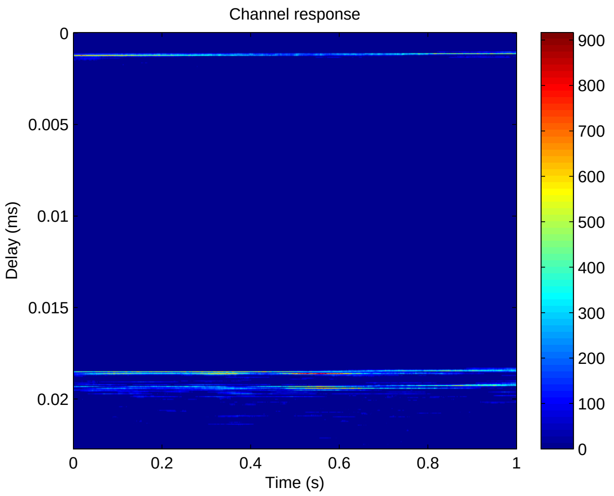


Fig. 3: Recorded channel response.

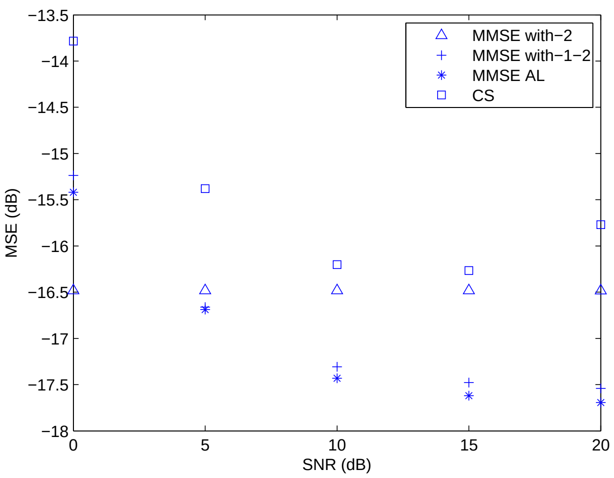


Fig. 5: MSEs versus SNR (dB).

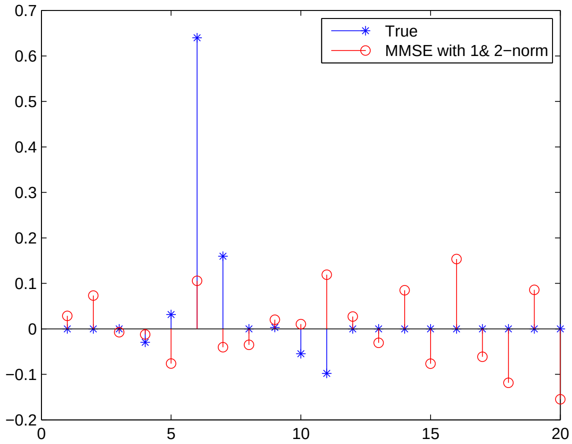
V. Conclusion

In this work, we reformulate the minimum mean square error (MMSE) estimator under minimax framework considering ℓ_1 and ℓ_2-norm constraints. By using duality approach, we then transform the problem into semidefinite programming (SDP), which can be solved efficiently. Numerical studies on both simulated and experimental data demonstrate the promising results from proposed approaches when signal sparsity is utilized.

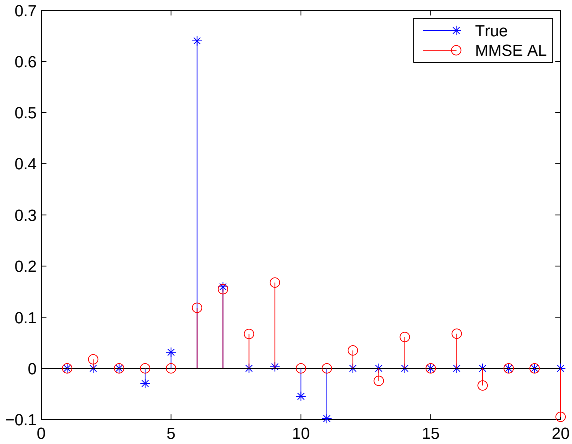
References

[1] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice-Hall, Englewood Cliffs, NJ, 1998.
[2] D. Donoho, "Compressed sensing," *IEEE Trans. Information Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
[3] E. J. Candès and T. Tao, "The dantzig selector: statistical estimation when p is much larger than n," *Annals of Statistics*, vol. 35, no. 35, pp. 2313–2351, 2007.
[4] Y. C. Eldar, A. Ben-Tal, and A. Nemirovski, "Robust mean-squared error estimation in the presence of model uncertainties," *IEEE Trans. Signal Processing*, vol. 53, no. 1, pp. 168–181, Jan. 2005.
[5] Z.-Q. Luo, T. N. Davidson, G. B. Giannakis, and K. M. Wong, "Transceiver optimization for block-based multiple access through ISI channels," *IEEE Trans. Signal Processing*, vol. 52, no. 4, pp. 1037–1052, Apr. 2004.
[6] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, Mar. 2008.
[7] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

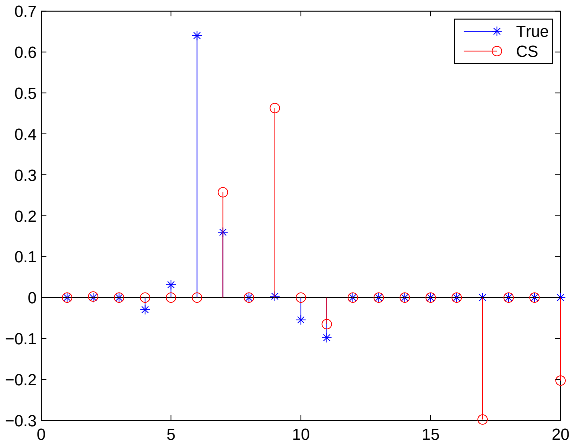
ISBN: 978-988-19251-6-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)



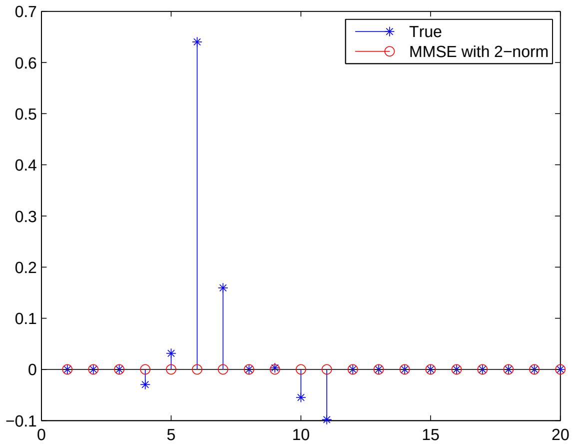
(a) MMSE estimator with ℓ_1 and ℓ_2-norm constraints



(b) MMSE estimator with Alternating optimization



(c) Compressed sensing estimation



(d) MMSE with ℓ_2-norm constraint

Fig. 4: Estimates on recorded channel.

ISBN: 978-988-19251-6-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)