Optimal Number of Measurements for Compressed Sensing with Quadratically Decreasing SNR

Yang Lu[1], Wei Dai[1] and Yonina C. Eldar[2]
[1]Department of Electrical and Electronic Engineering, Imperial College London, United Kingdom
[2]Department of Electrical Engineering, Israel Institute of Technology, Haifa, Israel

***Abstract*—In this paper, we consider a practical signal transmission application with fixed power budget such as radar/sonar. The system is modeled by a linear equation with the assumption that the signal energy per measurement decreases linearly and the noise energy per measurement increases approximately linearly with the increasing of the number of measurements. Thus the SNR decreases quadratically with the number of measurements. This model suggests an optimal operation point different from the common wisdom where more measurements always mean better performance. Our analysis shows that there is an optimal number of measurements, neither too few nor too many, to minimize the mean-squared error of the estimate. The analysis is based on a state evolution technique which is proposed for the approximate message passing algorithm. We consider the Gaussian, Bernoulli-Gaussian and least-favorite distributions in both real and complex domains. Numerical results justify the correctness of our analysis.**

***Index Terms*—Approximate message passing, compressed sensing, state evolution, signal recovery.**

I. Introduction

The problem of estimating a signal from its linear measurements has been studied for many decades. When the unknown signal is sparse, compressed sensing (CS) theory [1], [2] shows that the number of measurements can go below the dimension of the signal leading to an underdetermined linear measurement system. Low complexity algorithms to solve the resulting sparse recovery problem include greedy algorithms, for example orthogonal matching pursuit (OMP) [3], subspace pursuit (SP) [4] and compressive sampling matching pursuit (CoSaMP) [5], ℓ_1-norm minimization [6], and more recently approximate message passing (AMP) [7]. CS has been widely used in under-sampling [7], [8], imaging and localization [9]–[11], and sparse learning [12].

This paper focuses on a system design inspired by practical scenarios where the total energy budget of the linear measurements is fixed, the signal energy per measurement decreases linearly and the noise energy per measurement increases approximately linearly with the number of measurements. This scenario arises in many active sensing applications where measuring means observing the responses of a physical system to the stimulants that we actively put in. One example is radar systems. The number of measurements could correspond to the number of pulses per unit time (pulse frequency) or the number of sub-channels in the entire spectrum. When the number of measurements is increasing, the signal energy per measurement (per pulse/sub-channel) is decreasing linearly with the number of measurements. For the measurement noise, we adopt the commonly used additive white Gaussian noise model. Based on the famous thermal noise effect, the noise power (given by kTB [13]) increases approximately linearly with the number of measurements. With the above assumptions, the SNR per measurement should decrease quadratically if we add more measurements. Our goal in this paper is to address this trade-off and determine the optimal number of measurements. It is worth noting that although this paper focuses on sparse signals, the same trade-off exists for non-sparse signals as we show in Section IV.

The main contribution of this work is to find the exact optimal number of measurements required to minimize the mean-squared error (MSE) under certain mathematical assumptions. For the purpose of analysis, we assume a Gaussian measurement matrix, i.e., the elements in the measurement matrix are independently drawn from a Gaussian distribution. Let m be the number of measurements, n be the dimension of the unknown signal, and S be the number of non-zero elements. Further, let $m,\ n,\ S \to \infty$ with constant ratios $\delta := \frac{m}{n}$ and $\epsilon := \frac{S}{n}$ (sparsity level). We consider the MSE distortion metric. By characterizing the asymptotic distortion as a function of the normalized number of measurements δ, one can find the optimal number of measurements $\delta^\dagger$ that minimizes the distortion. The $\delta^\dagger$ may be directly achieved by a closed-form formula or by numerical calculation which depends on the statistics of the unknown signal. In order to provide intuition about the value of $\delta^\dagger$ for different unknown signals, we study upper bounds on $\delta^\dagger$ for three typical signal models: Gaussian, Bernoulli-Gaussian and least-favorite distributions in both real and complex domains. The first two signal models are commonly used for non-sparse and sparse signal analysis, respectively. The third model is used for worst case analysis meaning the resulting MSE performance is an upper bound on that of signals with arbitrary distribution with the same sparsity level. The worst case analysis result is pessimistic in general but at the same time universal. Analysis shows that for all three models, in both real and complex domains, the optimal value of $\delta^\dagger$ is upper bounded by 2.

Our results are based on the AMP algorithm and the associated state evolution analysis. It is noteworthy that though the rigorous derivation of state evolution of AMP requires a random Gaussian matrix, many works have demonstrated that the same results are relatively accurate for partial Fourier and Rademacher matrices [7], [14] when the sizes of these matrices are sufficiently large. Numerical results justify the correctness and the accuracy of the asymptotic analysis.

ISBN 978-0-9928626-7-1

II. Problem Formulation

A. System Model

Consider a signal transmission system with fixed power budget such as radar/sonar. If we transmit/receive a signal with more measurements m then the energy allocated to each single measurement is reduced. Thus, we assume the signal variance is proportional to m^{-1}. This effect can be modeled by multiplying each measurement by a factor of $1/\sqrt{m}$. In addition, based on [15], when considering the noise power of a signal receiving system such as radar, an adequate assumption is that the receiving system is ideal and only consider the thermal noise. A higher sampling rate means a larger bandwidth which results in a larger thermal noise with power increasing approximately linearly with m [13]. Let $\mathbb{H}$ represents the real ($\mathbb{R}$) or complex ($\mathbb{C}$) domain. The system is modeled as

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{w}, \tag{1}$$

where $\boldsymbol{y} \in \mathbb{H}^m$ denotes the observation vector; $\boldsymbol{A} \in \mathbb{H}^{m\times n}$ is the standard Gaussian random matrix with elements scaled by $1/\sqrt{m}$; $\boldsymbol{x} \in \mathbb{H}^n$ represents the unknown signal and $\boldsymbol{w} \in \mathbb{H}^m$ is additive Gaussian noise with mean zero and variance

$$\sigma_w^2 \coloneqq \delta\sigma_0^2, \tag{2}$$

where $\delta \coloneqq \frac{m}{n}$ and σ_0^2 is a constant. Let $\hat{\boldsymbol{x}}$ be the estimated signal. The performance of the system is given by the MSE

$$\mathrm{Err} \coloneqq \lim_{n\to\infty} \frac{1}{n} \|\boldsymbol{x} - \hat{\boldsymbol{x}}\|^2 . \tag{3}$$

In particular. we are interested in the value of δ that minimizes the MSE (3). We consider the system model (1) for both non-sparse and sparse signals.

B. Non-Sparse Setting

For the non-sparse setting, we consider the widely used Gaussian model as an example. The asymptotic MSE analysis for the traditional problem is well known. Assume that $\boldsymbol{A}$ is a Gaussian random matrix with i.i.d. elements drawn from $\mathcal{N}\left(0, \frac{1}{m}\right)$ when $\mathbb{H} = \mathbb{R}$ (or $\mathcal{CN}\left(0, \frac{1}{m}\right)$ when $\mathbb{H} = \mathbb{C}$), $\boldsymbol{x}$ is drawn from $\mathcal{N}\left(\boldsymbol{0}, \sigma_x^2\boldsymbol{I}\right)$ when $\mathbb{H} = \mathbb{R}$ (or $\mathcal{CN}\left(\boldsymbol{0}, \sigma_x^2\boldsymbol{I}\right)$ when $\mathbb{H} = \mathbb{C}$) and the noise $\boldsymbol{w}$ is drawn from $\mathcal{N}\left(\boldsymbol{0}, \sigma_w^2\boldsymbol{I}\right)$ when $\mathbb{H} = \mathbb{R}$ (or $\mathcal{CN}\left(\boldsymbol{0}, \sigma_w^2\boldsymbol{I}\right)$ when $\mathbb{H} = \mathbb{C}$). The asymptotic MSE of the minimum MSE (MMSE) estimator can be directly calculated based on random matrix theory [16]. Denoting $c = \frac{(1-\delta)}{\delta}$, we have

$$\lim_{n\to\infty} \frac{1}{n}\|\boldsymbol{x}-\hat{\boldsymbol{x}}\|^2 = \frac{\delta}{2}\left[\left(-\sigma_w^2 + c\sigma_x^2\right) + \sqrt{\left(\sigma_w^2 + c\sigma_x^2\right)^2 + 4\sigma_w^2\sigma_x^2}\right] . \tag{4}$$

By replacing the noise variance with our model (2), a trade-off between MSE and δ is achieved. Figure 1 plots an example, where we set $\sigma_x^2 = 1$ and vary the value of σ_0^2. For each given σ_0^2, by increasing the number of measurements, the MSE first decreases until reaches the optimal point; further increasing the number of measurements, the MSE becomes larger.

For sparse signals, we want to find a similar relationship, taking into account the nonlinear property of the sparse decoder. Instead, we use the state evolution technique of the well known AMP algorithm [7], [16]. Although AMP was originally proposed for solving CS problems in which the unknown signals are assumed to be S-sparse, we will show that the same analysis is also valid for non-sparse signals in Section IV-C.

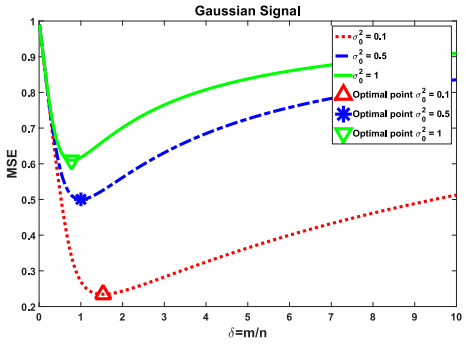

Figure 1: Trade-off for Gaussian signals.

III. Approximate Message Passing

The AMP algorithm was first proposed in [7] to solve (1) in the CS scenario. It, iteratively applyies the following equations:

$$\boldsymbol{x}^{t+1} = \eta\left(\boldsymbol{A}^*\boldsymbol{r}^t + \boldsymbol{x}^t\right), \tag{5}$$

$$\boldsymbol{r}^t = \boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}^t + \frac{1}{\delta}\left\langle \eta'\left(\boldsymbol{A}^*\boldsymbol{r}^{t-1} + \boldsymbol{x}^{t-1}\right)\right\rangle \boldsymbol{r}^{t-1}, \tag{6}$$

where $\boldsymbol{x}^t$ denotes the t-th estimation of the signal; $\eta(\cdot)$ is a component-wise estimator particularly designed based on the statistical information of its input argument; $\boldsymbol{A}^*$ stands for the (conjugate) transpose of $\boldsymbol{A}$; η' represents the first order derivative of η, and $\langle \boldsymbol{v} \rangle \coloneqq \frac{1}{n}\sum_{i=1}^{n} v_i$ computes the average. The last term of (6)

$$\text{Onsager} \coloneqq \frac{1}{\delta}\left\langle \eta'\left(\boldsymbol{A}^*\boldsymbol{r}^{t-1} + \boldsymbol{x}^{t-1}\right)\right\rangle \boldsymbol{r}^{t-1}, \tag{7}$$

is called the Onsager term. The input of $\eta\left(\boldsymbol{\beta}^t\right)$ can be written as the ground truth signal $\boldsymbol{x}$ plus an equivalent noise $\boldsymbol{w}_e^t$:

$$\boldsymbol{\beta}^t \coloneqq \boldsymbol{A}^*\boldsymbol{r}^t + \boldsymbol{x}^t = \boldsymbol{x} + \boldsymbol{w}_e^t. \tag{8}$$

By the effect of the Onsager term and the assumption of an i.i.d. Gaussian random matrix $\boldsymbol{A}$, the equivalent noise will always be approximately Gaussian and the statistical information can be calculated based on the formula:

$$\boldsymbol{w}_e^t \coloneqq \left(\boldsymbol{A}^*\boldsymbol{A} - \boldsymbol{I}\right)\left(\boldsymbol{x} - \boldsymbol{x}^t\right) + \boldsymbol{A}^*\boldsymbol{w}, \tag{9}$$

where $\left(\boldsymbol{A}^*\boldsymbol{A} - \boldsymbol{I}\right)$, $\left(\boldsymbol{x} - \boldsymbol{x}^t\right)$ and $\boldsymbol{A}^*\boldsymbol{w}$ are mutually independent.

At each iteration t, we first update the estimated signal $\boldsymbol{x}^t$ by a particularly designed η function which requires the knowledge of the equivalent noise $\boldsymbol{w}_e^{t-1}$. We than calculate MSE based on the current estimated signal $\boldsymbol{x}^t$. Next, we update the knowledge of the equivalent noise $\boldsymbol{w}_e^t$ according to (9) by

$$\left(\sigma_e^t\right)^2 = \frac{1}{\delta}\mathrm{Err}_t + \sigma_w^2 \tag{10}$$

where σ_e^t is the standard deviation of $\boldsymbol{w}_e^t$. The calculation of Err_t will be talked about later. This updated knowledge of $\boldsymbol{w}_e^t$ will be used to generate a new signal estimation $\boldsymbol{x}^{t+1}$ in the next iteration.

IV. Analysis in Real Domain

In this section, we analyze the relationship between MSE and δ (or equivalently m) in the real domain for both least-favorite and Bernoulli-Gaussian distributions. Then we extend the analysis to the complex domain in the next section. We

ISBN 978-0-9928626-7-1 © EURASIP 2017

first consider a realistic situation in which the only prior information on the unknown signal is that, it is sparse. The analysis in this case will provide a worst case universal solution. A designed decoder based on a given signal distribution should outperform the universal decoder. To study this case, we consider the Bernoulli-Gaussian distribution.

A. Least-Favorite Distribution (Worst Case Signal)

The worst case analysis of the AMP algorithm [7], [17] is applied by minimizing the MSE for the least-favorite distribution

$$p_x = \frac{\epsilon}{2}\Delta_{x=-\infty} + (1-\epsilon)\,\Delta_{x=0} + \frac{\epsilon}{2}\Delta_{x=+\infty} \tag{11}$$

where Δ denotes the Dirac delta function and $\epsilon \in (0,1]$ is the sparsity level. The MMSE estimator of the least-favorite distribution, based on (8) , is the soft thresholding function

$$\eta\left(\beta_i^t, \lambda^t\right) = \begin{cases} \beta_i^t - \lambda^t & \beta_i^t > \lambda^t \\ 0 & -\lambda^t \le \beta_i^t \le \lambda^t \\ \beta_i^t + \lambda^t & \beta_i^t < -\lambda^t. \end{cases} \tag{12}$$

The optimal threshold is defined by

$$\lambda^t \coloneqq \alpha^\dagger \sigma_e^t,\ \alpha^\dagger = \arg\min_{\alpha\in\mathbb{R}_+} M(\epsilon,\alpha)$$

$$M(\epsilon,\alpha) \coloneqq \epsilon\left(1+\alpha^2\right) + (1-\epsilon)\left[2\left(1+\alpha^2\right)\Phi(-\alpha) - 2\alpha\phi(\alpha)\right], \tag{13}$$

where $\phi(x)$ is the standard Gaussian density and $\Phi(x) = \int_{-\infty}^{x}\phi(t)dt$ is the corresponding cumulative distribution function. The MSE of estimation at each iteration can be calculated by

$$\mathrm{Err}_t = M\left(\epsilon, \alpha^\dagger\right)\left(\sigma_e^t\right)^2. \tag{14}$$

We apply the above results to our system model and achieve the following theorem

Theorem 1. *For a linear measurement system (1) with signal model (11) and additive white Gaussian noise with variance (2), apply AMP algorithm with estimator (12). By the convergence assumption of (10), we have*

$$\delta^\dagger = 2M\left(\epsilon, \alpha^\dagger\right), \tag{15}$$

which is independent of the noise variance.

B. Bernoulli-Gaussian Distribution

Next we consider the Bernoulli-Gaussian prior [8], [18], [19] with probability density given by

$$p_x = (1-\epsilon)\,\Delta_{x=0} + \epsilon p_G\left(x;\, 0,\, \sigma_x^2\right), \tag{16}$$

where $p_G\left(x;\, 0,\, \sigma_x^2\right)$ represents the Gaussian density with mean 0 and variance σ_x^2.

The η function can be designed based on the prior information of $\boldsymbol{x}$. Let $R^t \coloneqq \sigma_x^2/\left(\left(\sigma_e^t\right)^2 + \sigma_x^2\right)$ and define

$$I\left(R^t, \epsilon\right) \coloneqq \int \frac{\phi(x)}{1 + \frac{1-\epsilon}{\epsilon}\frac{1}{\sqrt{1-R^t}}\exp\left(-\frac{R^t}{1-R^t}\frac{x^2}{2}\right)} x^2 dx. \tag{17}$$

The component-wise function η can be chosen as the MMSE estimator, for each element of $\boldsymbol{\beta}^t$:

$$\eta\left(x_i^t|\beta_i^t\right) \coloneqq \frac{p_G\left(\beta_i^t; 0, \left(\sigma_e^t\right)^2 + \sigma_x^2\right)}{p\left(\beta_i^t\right)}\epsilon R^t \beta_i^t, \tag{18}$$

with $p\left(\beta_i^t\right) \coloneqq (1-\epsilon)\,p_G\left(\beta_i^t; 0, \left(\sigma_e^t\right)^2\right) + \epsilon p_G\left(\beta_i^t; 0, \left(\sigma_e^t\right)^2 + \sigma_x^2\right)$. The corresponding derivative of $\eta\left(x_i^t|\beta_i^t\right)$ is complicated, here we only give the final result. For simplicity of notation, define

$$v_1 \coloneqq \frac{1-\epsilon}{\epsilon}\sqrt{\frac{\left(\sigma_e^t\right)^2 + \sigma_x^2}{\left(\sigma_e^t\right)^2}},\ v_2 \coloneqq \frac{R^t}{\left(\sigma_e^t\right)^2},\ v_3 \coloneqq v_1 \exp\left(-\frac{1}{2}v_2\left(\beta_i^t\right)^2\right),$$

we have $\eta'\left(x_i^t|\beta_i^t\right) = R^t/(v_3+1) + R^t v_3 v_2\left(\beta_i^t\right)^2/(v_3+1)^2$ and

$$\mathrm{Err}_t \coloneqq \left[\frac{R^t\epsilon}{1-R^t}\left(1 - R^t I\left(R^t, \epsilon\right)\right)\right]\left(\sigma_e^t\right)^2. \tag{19}$$

Fast calculation of Err_t: We can increase the efficiency of AMP algorithm by avoiding the integration of (17). Based on Lemma 2, Err_t can be approximately calculated by

$$\mathrm{Err}_t \approx \left[\frac{1}{n}\sum_{i=1}^{n}\eta'\left(x_i^t|\beta_i^t\right)\right]\left(\sigma_e^t\right)^2. \tag{20}$$

Lemma 2. *[20, Lemma 2] Consider a random variable U with a conditional probability density function of the form $p_{U|V}(u|v) \coloneqq \frac{1}{Z(v)}exp\left(\phi_0(u) + uv\right)$, where $Z(v)$ is a normalization constant, Then,*

$$\frac{\partial}{\partial v}\log Z(v) = \mathbb{E}\left[U|V=v\right]$$
$$\frac{\partial^2}{\partial v^2}\log Z(v) = \frac{\partial}{\partial v}\mathbb{E}\left[U|V=v\right] = \mathrm{var}\left(U|V=v\right).$$

There is no closed form of Err_t, thus we can't directly achieve the optimal $\delta^\dagger$ as in Theorem 1. On the other hand, when AMP converges, Err_t and σ_e^t will converge to fixed points Err_∞ and σ_e^∞, respectively. Based on the relationship between Err_∞ and σ_e^∞ given in (19), the optimal $\delta^\dagger$ can be obtained by the following theorem.

Theorem 3. *For a linear measurement system (1) with signal model (16) and additive white Gaussian noise with variance (2), apply AMP algorithm with estimator (18). For any given set of parameters $\{\epsilon,\ \sigma_e^\infty,\ \sigma_x^2,\ \sigma_0^2\}$ such that $(\sigma_e^\infty)^4 - 4\sigma_0^2\mathrm{Err}_\infty \ge 0$, by the convergence assumption of (10), we have*

$$\delta = \frac{(\sigma_e^\infty)^2 \pm \sqrt{(\sigma_e^\infty)^4 - 4\sigma_0^2\mathrm{Err}_\infty}}{2\sigma_0^2}. \tag{21}$$

The optimal $\delta^\dagger$ is achieved when $(\sigma_e^\infty)^4 = 4\sigma_0^2\mathrm{Err}_\infty$.

C. Non-Sparse Case (Gaussian)

The state evolution analysis for sparse signals is also valid for non-sparse cases by considering the Bernoulli-Gaussian prior with $\epsilon = 1$. In this case, (17) will degenerate to the variance of a standard Gaussian distribution which is a constant with value equal to 1. The estimated error (19) then has a closed form

$$\mathrm{Err}_t = R^t\left(\sigma_e^t\right)^2. \tag{22}$$

Substituting (22) into (10) and setting $\sigma_e^t = \sigma_e^{t+1} = \sigma_e^\infty$, leads to the optimal value:

$$\sigma_e^\infty = \frac{\left(c\sigma_x^2 + \delta\sigma_0^2\right) + \sqrt{\left(c\sigma_x^2 + \delta\sigma_0^2\right)^2 + 4\sigma_x^2\delta\sigma_0^2}}{2},$$

where $c \coloneqq \frac{(1-\delta)}{\delta}$. We ignore the negative value due to the non-negative property of the error. The final estimation error at the fixed point will be

$$\mathrm{Err}_\infty = \delta\left((\sigma_e^\infty)^2 - \delta\sigma_0^2\right)$$

which is exactly the same as (4).

ISBN 978-0-9928626-7-1 © EURASIP 2017

V. Analysis in Complex Domain

The analysis in the complex domain follows the same line as it in the real domain but we need to take care of the modifications.

For least-favorite distribution: The complex AMP (CAMP) algorithm for least-favorite distribution has been analyzed in [21] providing a new Onsager term. The least-favorite distribution becomes $p_{|x|} = (1-\epsilon)\,\Delta_{|x|=0} + \epsilon\Delta_{|x|=+\infty}$ with the assumption that the phase of x is isotropic and based on [21], the η function will be,

$$\eta\left(\beta_i^t, \lambda\right) \coloneqq \left(\beta_i^t - \frac{\lambda\left(\beta_i^t\right)}{|\beta_i^t|}\right) 1_{\{|\beta_i^t|>\lambda\}} \tag{23}$$

where $1_{\{|\beta_i^t|>\lambda\}}$ denotes the indicator function. The formula of $\mathrm{Err}_{C,t}$ will be the same as in real case but with a new $M_C\left(\epsilon,\alpha\right)$ function:

$$M_C\left(\epsilon,\alpha\right) \coloneqq \epsilon\left(1+\alpha^2\right)+(1-\epsilon)\left[\sqrt{2\pi}\phi\left(\sqrt{2}\alpha\right)-2\alpha\sqrt{\pi}\Phi\left(-\sqrt{2}\alpha\right)\right]. \tag{24}$$

Compare (24) with (13), we can find the estimation error of non-zero components of signal are the same (first term). The difference between them comes from the de-noising for the zero components of signal (second term). For the complete derivation of new Onsager term and calculation of $\eta'\left(\beta_i^t,\lambda\right)$, please refer to [21].

For Bernoulli-Gaussian distribution: We assume that the real part and imaginary part of a complex variable share the same mean and variance and they are uncorrelated. For example, let $x \sim \mathcal{CN}\left(\mu,\sigma_x^2\right)$, then we have $(x)^R, (x)^I \sim \mathcal{N}\left(\mu, \frac{\sigma_x^2}{2}\right)$. Under this assumption, we have

$$\begin{aligned} p_{CG}\left(x;\mu,\sigma_x^2\right) &= p_G\left((x)^R;\mu,\frac{\sigma_x^2}{2}\right) p_G\left((x)^I;\mu,\frac{\sigma_x^2}{2}\right) \\ &= \frac{1}{\pi\sigma_x^2} exp\left(-\frac{|x-\mu|^2}{\sigma_x^2}\right), \end{aligned} \tag{25}$$

and the Bernoulli-Gaussian distribution in the complex domain becomes $p(x) = (1-\epsilon)\,\Delta_{|x|=0} + \epsilon p_{CG}\left(x\right)$. For the estimate function η, we just replace the p_G probability in (18) with p_{CG} defined above. Now let $p_{\beta,1}^t \coloneqq p_{CG}\left(\beta_j^t; 0, \left(\sigma_e^t\right)^2 + \sigma_x^2\right)$, $p_{\beta,2}^t \coloneqq p_{CG}\left(\beta_j^t; 0, \left(\sigma_e^t\right)^2\right)$, $p_{\beta,3}^t \coloneqq (1-\epsilon)\,p_{\beta,2}^t + \epsilon p_{\beta,1}^t$ and $p_o^t \coloneqq -\frac{2}{\sigma_x^2+(\sigma_e^t)^2}p_{\beta,3}^t + \frac{2(1-\epsilon)}{\sigma_w^2}p_{\beta,2}^t + \frac{2\epsilon}{\sigma_x^2+(\sigma_e^t)^2}p_{\beta,1}^t$, the four derivatives of η can be calculated based on the following formulas:

$$\frac{\partial\eta^R\left(\beta_j^t\right)}{\partial\left(\beta_j^t\right)^R} = \frac{p_o^t}{\left(p_{\beta,3}^t\right)^2} p_{\beta,1}^t \epsilon R\left(\left(\beta_j^t\right)^R\right)^2 + \frac{p_{\beta,1}^t}{p_{\beta,3}^t}\epsilon R \tag{26}$$

$$\frac{\partial\eta^R\left(\beta_j^t\right)}{\partial\left(\beta_j^t\right)^I} = \frac{\partial\eta^I\left(\beta_j^t\right)}{\partial\left(\beta_j^t\right)^R} = \frac{p_o^t}{\left(p_{\beta,3}^t\right)^2} p_{\beta,1}^t \epsilon R\left(\beta_j^t\right)^R\left(\beta_j^t\right)^I \tag{27}$$

$$\frac{\partial\eta^I\left(\beta_j^t\right)}{\partial\left(\beta_j^t\right)^I} = \frac{p_o^t}{\left(p_{\beta,3}^t\right)^2} p_{\beta,1}^t \epsilon R\left(\left(\beta_j^t\right)^I\right)^2 + \frac{p_{\beta,1}^t}{p_{\beta,3}^t}\epsilon R, \tag{28}$$

Finally, (19) will be replaced by

$$\mathrm{Err}_{C,t} = \left[\frac{R^t\epsilon}{1-R^t}\left(1 - R^t I_C\left(R^t,\epsilon\right)\right)\right]\left(\sigma_e^t\right)^2, \tag{29}$$

$$I_C\left(R^t,\epsilon\right) = \int_{x^R}\int_{x^I} \frac{\phi_C\left(x\right)}{1+\frac{1-\epsilon}{\epsilon}\frac{1}{1-R^t}\exp\left(-\frac{R^t}{1-R^t}|x|^2\right)} |x|^2\, dx^I dx^R \tag{30}$$

where $\phi_C\left(x\right) = p_{CG}\left(x;0,1\right)$ is the standard complex normal distribution. (29) and (19) are exactly the same except the integration terms.

Fast calculation of $\mathrm{Err}_{C,t}$: The same as in the real domain, we can efficiently calculate Equation (29) by focusing on the real part of the signal only,

$$\mathrm{Err}_{C,t}^R \approx \frac{1}{n}\sum_{i=1}^n \left(\frac{\partial\eta^R\left(\beta_j^t\right)}{\partial\left(\beta_j^t\right)^R}\right)\frac{\left(\sigma_e^t\right)^2}{2} \tag{31}$$

$$\mathrm{Err}_{C,t} \approx 2\mathrm{Err}_{C,t}^R = \frac{1}{n}\sum_{i=1}^n \left(\frac{\partial\eta^R\left(\beta_j^t\right)}{\partial\left(\beta_j^t\right)^R}\right)\left(\sigma_e^t\right)^2 \tag{32}$$

which based on the assumption that the real part and imaginary part of the complex random variable are i.i.d..

The optimal $\delta^\dagger$ can be achieved by using the same theorems in section IV, just replacing $M\left(\epsilon,\alpha^\dagger\right)$ and Err_t functions with $M_C\left(\epsilon,\alpha^\dagger\right)$ and $\mathrm{Err}_{C,t}$, respectively.

VI. Discussion and Numerical Justification

A. Discussion on the optimal of $\delta^\dagger$

The optimal $\delta^\dagger$ for least-favorite distribution can be directly achieved by (15) while for Bernoulli-Gaussian, $\delta^\dagger$ only can be achieved by numerically calculation. In order to find the common attribute of $\delta^\dagger$ and get an intuition about the possible outcome of $\delta^\dagger$ for different kinds of signals, we analysis the upper bounds for Gaussian, Bernoulli-Gaussian and least-favorite distributions.

Let's focus on the real domain first. For the Gaussian case, based on (4), we are able to achieve $\delta^\dagger = \left(\sqrt{\sigma_x^4/\sigma_0^4 + 16\sigma_x^2/\sigma_0^2} - \sigma_x^2/\sigma_0^2\right)/4$ which will monotonically decrease for increasing $\sigma_0 \in (0,\infty]$ and $\delta^\dagger$ is upper bounded by 2. For the least favorite distribution, based on Theorem 1 and the fact that $M\left(\epsilon,\alpha^\dagger\right) \in (0,1]$, a same upper bound can be achieved. For the Bernoulli-Gaussian case, based on Theorem 3, $\delta^\dagger$ will increase by the increasing of Err_∞ which is upper bounded by Gaussian case. Thus $\delta^\dagger$ is also upper bounded by 2. Same performance can be achieved in the complex domain. The detailed proof will be provided in our upcoming paper.

B. Numerical Justification

For the simulation, we set $n = 1000$, $\sigma_0^2 = 0.01$ as constants. Each simulation point is the average of 100 independent trials. The simulation results provided in Fig. 2 show the relationship between the MSE and the measurement ratio δ for a given sparsity level ϵ. From the figure, one can observe that when δ increases at the beginning, the MSE decreases dramatically until it reaches a minimum. After that, further increase in δ will enlarge the MSE. This phenomenon verifies our presumption that there exists an optimal $\delta^\dagger$ (or $m^\dagger$) for a power fixed signal transmission system. The overall performance of Bernoulli-Gaussian distribution is better than the one of least-favorite distribution which coincides with our explanation at the beginning of Section V. The numerical results of Bernoulli-Gaussian signals match the theoretical curves quite well but for the least-favorite distribution, the numerical results are slightly larger than the theoretical curves.

ISBN 978-0-9928626-7-1 © EURASIP 2017

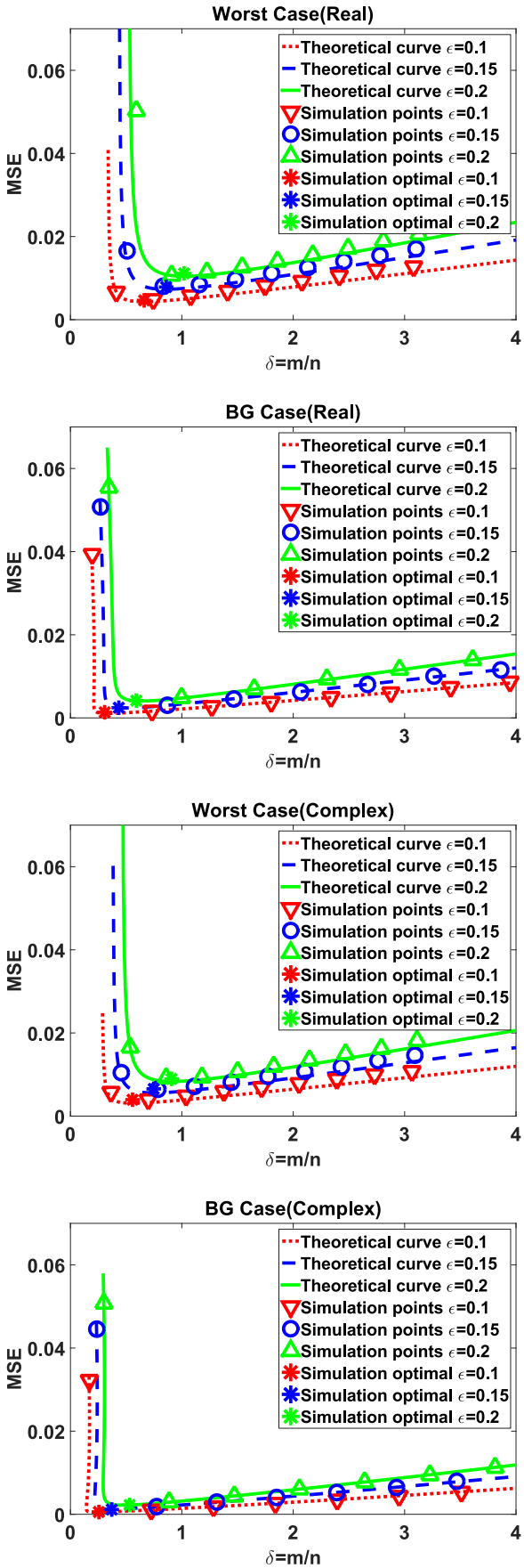

Figure 2: MMSE vs δ

The main reason is that for the theoretical analysis in this case, we assume that the values of the non-zero coefficients are $\pm\infty$, but in simulations, these values can only be set as certain large numbers which results in a lower SNR compared with the one in the theoretical case. For both signal distributions, the moving trends of $\delta^\dagger$ for different ϵ values coincide with above optimal analysis.

REFERENCES

[1] Y. C. Eldar, *"Sampling Theory: Beyond Bandlimited Systems"*, Cambridge University Press, April 2015.

[2] Y. C. Eldar and G. Kutyniok, *"Compressed Sensing: Theory and Applications"*, Cambridge University Press, May 2012.

[3] Y.C. Pati, R. Rezaiifar, and P.S. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," *Asilomar Conference on Signals, Systems and Computers*, pp. 40–44 vol.1, Nov, 1993.

[4] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Transactions on Information Theory*, vol. 55, no. 5, pp. 2230–2249, May, 2009.

[5] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis (ACHA)*, vol. 26, no. 3, pp. 301–321, 2009.

[6] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[7] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proceedings of the National Academy of Sciences (PNAS)*, vol. 106, no. 45, pp. 18914–18919, 2009.

[8] X. Zhao and W. Dai, "On joint recovery of sparse signals with common supports," *IEEE International Symposium on Information Theory (ISIT)*, pp. 541–545, June 2015.

[9] R. Baraniuk and P. Steeghs, "Compressive radar imaging," *IEEE Radar Conference*, pp. 128–133, Mar, 2017.

[10] C. R. Berger, B. Demissie, J. Heckenbach, P. Willett, and S. Zhou, "Signal processing for passive radar using ofdm waveforms," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 1, pp. 226–238, Feb 2010.

[11] S. Som, L. C. Potter, and P. Schniter, "Compressive imaging using approximate message passing and a markov-tree prior," pp. 243–247, Nov 2010.

[12] U. S. Kamilov, S. Rangan, A. K. Fletcher, and M. Unser, "Approximate message passing with consistent parameter estimation and applications to sparse learning," *IEEE Transactions on Information Theory*, vol. 60, no. 5, pp. 2969–2985, May 2014.

[13] J. R. Pierce, "Physical sources of noise," *Proceedings of the IRE*, vol. 44, no. 5, pp. 601–608, May 1956.

[14] D. L. Donoho, A. Maleki, and A. Montanari, "Supporting information to: Message-passing algorithms for compressed sensing," *Proc. Nat. Acad. Sci. USA*, 2009.

[15] Avionics Department, "Electronic warfare and radar systems engineering handbook," *A Comprehensive Handbook for Electronic Warfare and Radar Systems Engineers*, October 2013.

[16] M. Bayati and A. Montanari, "The dynamics of message passing on dense graphs, with applications to compressed sensing," *IEEE Transactions on Information Theory*, vol. 57, no. 2, pp. 764–785, Feb, 2011.

[17] A. Montanari, "Graphical models concepts in compressed sensing," *arXiv preprint arXiv:1011.4328*, 2010.

[18] Y. Lu and W. Dai, "Extended amp algorithm for correlated distributed compressed sensing model," *IEEE International Conference on Digital Signal Processing(DSP)*, Oct, 2016, in press.

[19] Y. Lu and W. Dai, "Independent versus repeated measurements: a performance quantification via state evolution," *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, Mar, 2016, in press.

[20] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," *IEEE International Symposium on Information Theory Proceedings (ISIT)*, pp. 2168–2172, July 2011.

[21] A. Maleki, L. Anitori, Z. Yang, and R. G. Baraniuk, "Asymptotic analysis of complex lasso via complex approximate message passing (camp)," *IEEE Transactions on Information Theory*, vol. 59, no. 7, pp. 4290–4308, July 2013.

ISBN 978-0-9928626-7-1 © EURASIP 2017