Improved AMP (IAMP) for Non-Ideal Measurement Matrices

Yang Lu and Wei Dai
Department of Electrical and Electronic Engineering, Imperial College London, United Kingdom
{yang.lu111, wei.dai1}@imperial.ac.uk

***Abstract*—This paper studies the sparse recovery problem. Of particular interest is the well known approximate message passing (AMP) algorithm. AMP enjoys low computational complexity and good performance guarantees. However, the algorithm and performance analysis heavily rely on the assumption that the measurement matrix is a standard Gaussian random matrix. The main contribution of this paper is an improved AMP (IAMP) algorithm that works better for non-ideal measurement matrices. The algorithm is equivalent to AMP for standard Gaussian random matrices but provides better recovery when the correlations between columns of the measurement matrix deviate from those of the standard Gaussian random matrices. The derivation is based on a modification of the message passing mechanism that removes the conditional independence assumption. Examples are provided to demonstrate the performance improvement of IAMP where both a particularly designed matrix and a matrix from real applications are used.**

***Index Terms*—AMP, compressed sensing, message passing, sparse signal processing, standard Gaussian random matrix.**

I. Introduction

This paper studies the sparse recovery problem. Let

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

where $\boldsymbol{y} \in \mathbb{R}^m$ is the observation vector, $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ stands for the measurement matrix, $\boldsymbol{x} \in \mathbb{R}^n$ is the signal vector, and $\boldsymbol{w} \in \mathbb{R}^m$ denotes the additive white Gaussian noise. Assume that $\boldsymbol{x}$ is sparse, that is, the number (k) of nonzero entries in $\boldsymbol{x}$ is much smaller than n the dimension of $\boldsymbol{x}$. Sparse recovery problem concerns the reconstruction of the unknown sparse signal $\boldsymbol{x}$ from the noisy observations $\boldsymbol{y}$. It finds its wide applications in compressed sensing, denoising, recovery of missing data, and sparse linear regression [1]–[5].

There has been a huge literature in low complexity algorithms to solve the sparse recovery problem. From the information theoretical point of view, solving this problem requires exhaustive search of which the complexity increases exponentially with the problem dimension. One way to reduce the complexity is to cast the recovery problem as a convex optimisation where the ℓ_1-norm replaces the sparsity constraint. Theoretic performance guarantees in terms of phase transition, restrict isometry property (RIP), etc. have been proved [6], [7]. Another low-complexity alternative is the greedy algorithms include OMP, SP, CoSaMP, etc. [6], [8], [9] where the performance guarantees are often based on RIP. For both approaches, it has been proved that for noiseless case exact recovery is possible even when the linear system in (1) is severely under-determinined, i.e., $m = O\left(k \cdot \log \frac{n}{k}\right) \ll n$.

In this paper, of our particular interest is the approximate message passing (AMP) algorithm. It has received wide attention due to its two nice properties: low complexity and good performance guarantees. It only involves matrix-vector products and scalar operations and therefore the complexity is $O\left(n^2\right)$. At the same time, if the measurement matrix is a standard Gaussian random matrix, it has been rigorously proved that AMP achieves the same phase transition curve as ℓ_1-minimisation does. Furthermore, AMP allows complicated statistical models for both the unknown sparse signal and the noise [10]–[14]. It has been proved in [11] that any signal model can be applied as long as the corresponding denoiser is Lipschitz continuous. This extends the applicability of AMP.

One drawback of AMP is that both the algorithm and the performance analysis heavily rely on the standard Gaussian random matrix. It has been numerically observed that the performance of AMP may severely deteriorate if the measurement matrix is significantly different from the standard Gaussian random matrix. A particularly designed example is given in Section IV to highlight this phenomenon. This drawback limits the applicability of AMP. There have been methods proposed to address this issue, including Damped GAMP [15] which linearly combines the results from two adjacent iterations, SwAMP [16] which updates components in $\boldsymbol{x}$ sequentially, and ADMM-GAMP [17] which considers the inference problem of generalized linear models (GLM) as a large-system-limit approximation of the Bethe Free Energy (LSL-BFE) minimization problem and using ADMM method to solve it.

The major contribution of this paper is an improved AMP (IAMP) algorithm that works better for non-ideal measurement matrices. The derivation of AMP is based on a factor graph representation of the system and Gaussian approximations of the passed messages on the factor graph. We observe that the conditional independence assumption used in the message computation is not valid any more when the measurement matrix is not ideal. It turns out that the correlation profile of the columns of the measurement matrix needs to be taken into consideration. Based on this observation, a new message passing mechanism is derived where all messages are computed at the variable nodes. This is quite different form previous approaches in [15]–[17]. The developed IAMP algorithm reduces to AMP when the measurement matrix is

This work was supported in part by Defence Science and Technology Laboratory (Dstl) under Grant No: DSTLX-1000081291

978-0-9928626-3-3/15/$31.00 ©2015 IEEE

standard Gaussian; at the same time, substantial performance improvements of IAMP are demonstrated for non-ideal measurement matrices. It is noteworthy that IAMP involves extra computations. However the extra computations can be made offline so that the 'operational' complexity of IAMP is in the same order as that of AMP.

II. Approximate Message Passing (AMP)

AMP is an efficient and powerful technique to solve the inverse problem (1). Let $\boldsymbol{A}$ be the standard Gaussian random matrix containing i.i.d. Gaussian entries, and $\boldsymbol{w}$ be the additive white Gaussian noise. AMP iteratively computes

$$\boldsymbol{x}^{t+1} = \eta\left(\boldsymbol{A}^T\boldsymbol{r}^t + \boldsymbol{x}^t\right), \tag{2}$$

$$\boldsymbol{r}^t = \boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}^t + \frac{1}{\delta}\boldsymbol{r}^{t-1}\left\langle \eta'\left(\boldsymbol{A}^T\boldsymbol{r}^{t-1} + \boldsymbol{x}^{t-1}\right)\right\rangle, \tag{3}$$

where by convention define $\boldsymbol{r}^{-1} = \boldsymbol{0}$ and $\boldsymbol{x}^0 = \boldsymbol{0}$, $\eta(\cdot)$ is a scalar function often referred to as denoiser, $\eta(\boldsymbol{v})$ means applying the $\eta(\cdot)$ to each component of the vector $\boldsymbol{v}$, the superscript T denotes the matrix transpose, $\delta = \frac{m}{n}$ is a constant, $\langle \boldsymbol{v}\rangle = \frac{1}{n}\sum_{i=1}^n v_i$ computes the average of the vector $\boldsymbol{v}$, and $\eta'(\cdot)$ denotes the derivative of the function $\eta(\cdot)$. In [11], it has been shown that if the prior distribution of $\boldsymbol{x}$ is assumed to be Laplacian, i.e., $p(\boldsymbol{x}) \propto \exp\left(-\beta\|\boldsymbol{x}\|_1\right)$, then $\eta(\cdot)$ takes the form of the soft thresholding function and AMP achieves the same phase transition curve as the famous LASSO problem

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \lambda\|\boldsymbol{x}\|_1 \tag{4}$$

for an appropriately chosen constant $\lambda \in \mathbb{R}^+$.

For the purpose of this paper, we briefly review the derivation of the AMP algorithm. AMP is based on the well known belief propagation (BP) mechanism. Describe the probability model of the system using the factor graph in Figure 1, where a variable node $i \in [n]$ denotes x_i and a factor node $a \in [m]$ specifies the conditional probability $p(y_a|\boldsymbol{x})$. The message from a factor node a to a variable node i, denoted by $m_{a\to i}(x_i)$, is given by

$$m_{a\to i}(x_i) = p(x_i|y_a) \propto \int p(y_a|x_i, \boldsymbol{x}_{\sim i})\, p(\boldsymbol{x}_{\sim i})\, d\boldsymbol{x}_{\sim i} \tag{5}$$

$$= \int p(y_a|x_i, z_{a\to i})\, p(z_{a\to i})\, dz_{a\to i}, \tag{6}$$

where $\boldsymbol{x}_{\sim i}$ denotes all the components in $\boldsymbol{x}$ except x_i, and $z_{a\to i} := \sum_{j\neq i} A_{aj}x_j$. Note that generally speaking, $p(z_{a\to i})$ is complicated and it is computationally expensive to compute the integral involved in $m_{a\to i}(x_i)$. However, when $\boldsymbol{A}$ is a standard Gaussian random matrix, $z_{a\to i}$ is a *Gaussian* random variable. The message $m_{a\to i}(x_i)$ can be easily obtained and is also Gaussian. Now consider the message from a variable node i to a factor node a. Let $\hat{x}_{a\leftarrow i} = \arg\max_{x_i} p(x_i|\boldsymbol{y}_{\sim a})$, where

$$p(x_i|\boldsymbol{y}_{\sim a}) \overset{(a)}{\propto} p(x_i)\prod_{b\neq a} p(y_b|x_i) \propto p(x_i)\prod_{b\neq a} m_{b\to i}(A_{bi}x_i), \tag{7}$$

where the relation (a) is based on the *conditional independence* assumption. When each $m_{b\to i}(A_{bi}x_i)$ is in Gaussian form, the computation of $p(x_i|\boldsymbol{y}_{\sim a})$ is highly simplified. In summary, the Gaussian approximation and conditional independence assumption are the two key elements in the derivation.

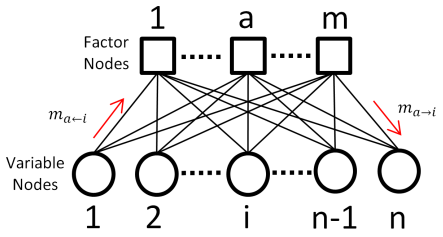


Figure 1: Factor Graph and Message Passing

III. Improved Approximate Message Passing (IAMP)

A. Modification of Message Passing

The main difference between AMP and IAMP is the message passing mechanism to handle general measurement matrix $\boldsymbol{A}$. When the matrix $\boldsymbol{A}$ is sufficiently dense, $z_{a\to i}$ can be approximated by a Gaussian random variable so the Gaussian assumption for AMP is still valid. However, when the columns of $\boldsymbol{A}$ are highly correlated, the independence assumption (among $m_{b\to i}$'s, $b \in\sim a$) is not true any more and neither is (7). To address this issue, a new message passing mechanism has to be designed. In particular, due to the dependence between $m_{b\to i}$'s, the computation of them at the factor nodes becomes un-necessary. We focus on the message at the variable node

$$m_{\boldsymbol{y}\to i}(x_i) = p(x_i|\boldsymbol{y}) \propto \int p(\boldsymbol{y}|x_i, \boldsymbol{x}_{\sim i})\, p(\boldsymbol{x}_{\sim i})\, d\boldsymbol{x}_{\sim i}, \tag{8}$$

where we stick to the common assumption that $p(\boldsymbol{x}_{\sim i}) = \prod_{j\neq i} p(x_j)$. With the assumption that the measurement noise $\boldsymbol{w} \sim \mathcal{N}\left(\boldsymbol{0}, \sigma_w^2\boldsymbol{I}\right)$, the following Lemma suggests that $p(x_i|\boldsymbol{y})$ can be approximated by a simple Gaussian pdf.

Lemma 1. *Let $\boldsymbol{A}_i$ be the i^{th} column of $\boldsymbol{A}$ and $\boldsymbol{A}_i^T$ is the transpose of $\boldsymbol{A}_i$. Define $\tilde{y}_i := \boldsymbol{A}_i^T\boldsymbol{y}$, $z_i := \boldsymbol{A}_i^T\left(\sum_{j\neq i}\boldsymbol{A}_j x_j\right)$, $\mathrm{E}[x_j] = 0$, $\sigma_{x_j}^2 := \mathrm{E}\left[x_i^2\right]$, and*

$$\sigma_{z_i}^2 := \mathrm{var}(z_i) = \sum_{j\neq i}\left(\boldsymbol{A}_i^T\boldsymbol{A}_j\right)^2\sigma_{x_j}^2 \tag{9}$$

Assume that $\|\boldsymbol{A}_i\|_2 = 1$, $\forall i \in [n]$, and $\boldsymbol{A}\boldsymbol{x}$ is jointly Gaussian. Then $p(x_i|\boldsymbol{y})$ can be approximated by $\mathcal{N}\left(\tilde{y}_i, \sigma_w^2 + \sigma_{z_i}^2\right)$.

Proof Sketch. We first calculate $\mathrm{P}(\boldsymbol{y}|\, x_i,\ \boldsymbol{x}_{\sim i})$ by treating $\boldsymbol{y}$ and $\boldsymbol{x}_{\sim i}$ as constant vectors:

$$\ln p(\boldsymbol{y}|\, x_i,\, \boldsymbol{x}_{\sim i}) \tag{10}$$

$$= -\frac{1}{2\sigma_w^2}\left(\sum_a\left(y_a - A_{ai}x_i - \sum_{j\neq i}A_{aj}x_j\right)^2\right) + c \tag{11}$$

$$= -\frac{1}{2\sigma_w^2}\left(\|\boldsymbol{A}_i\|^2x_i^2 - 2\tilde{y}_ix_i + 2z_ix_i\right) + c + c' \tag{12}$$

$$= -\frac{1}{2\sigma_w^2}\left(\tilde{y}_i - z_i - x_i\right)^2 + c + c' + c'', \tag{13}$$

where c is a constant only dependent on $\boldsymbol{y}$, and c' and c'' are two constants and their sum is given by

$$c'+c''=-\frac{1}{2\sigma_w^2}\left(\|\boldsymbol{y}\|_2^2+\left\|\boldsymbol{A}_i^{\perp T}\boldsymbol{A}_{\sim i}\boldsymbol{x}_{\sim i}\right\|_2^2-\tilde{y}_i^2 +2\tilde{y}_i z_i-2\left(\boldsymbol{A}_{\sim i}\boldsymbol{x}_{\sim i}\right)^T\boldsymbol{y}\right), \tag{14}$$

where $\boldsymbol{A}_i^{\perp}$ is the orthogonal complement of $\boldsymbol{A}_i$. Now consider the Gaussian vector $\boldsymbol{z}_i^{\perp}:=\boldsymbol{A}_i^{\perp T}\boldsymbol{A}_{\sim i}\boldsymbol{x}_{\sim i}$. It is straightforward to verify that the correlation between z_i and $\boldsymbol{z}_i^{\perp}$ is zero and therefore z_i and $\boldsymbol{z}_i^{\perp}$ are independent. As a result, the integral in (8) becomes

$$\int p\left(\boldsymbol{y}|x_i,\boldsymbol{x}_{\sim i}\right)p\left(\boldsymbol{x}_{\sim i}\right)d\boldsymbol{x}_{\sim i} \tag{15}$$

$$=\int p\left(\boldsymbol{y}|x_i,z_i,\boldsymbol{z}_i^{\perp}\right)p\left(z_i\right)p\left(\boldsymbol{z}_i^{\perp}\right)dz_i d\boldsymbol{z}_i^{\perp} \tag{16}$$

$$\overset{(a)}{=}\int c_1\exp\left(-\frac{1}{2\sigma_w^2}\left(\tilde{y}_i-z_i-x_i\right)^2-\frac{z_i^2}{2\sigma_{z_i}^2}+c_2 z_i\right)dz_i \tag{17}$$

$$=c_3\exp\left(-\frac{1}{2\left(\sigma_w^2+\sigma_{z_i}^2\right)}\left(\tilde{y}_i-x_i\right)^2\right), \tag{18}$$

where Equation (a) is obtained by integrating $\boldsymbol{z}_i^{\perp}$ out, and c_1, c_2 and c_3 are three constants. This lemma is therefore proved.

B. Algorithm Description

At the variable nodes, the operation of IAMP is the same as that of AMP: each signal component is denoised individually from its noisy observation

$$\tilde{y}_i=x_i+\tilde{w}_i, \tag{19}$$

where $\tilde{w}_i$ is additive Gaussian noise with distribution $\mathcal{N}\left(0,\sigma_{\tilde{w}_i}^2\right)$. Based on Lemma 1, $\sigma_{\tilde{w}_i}^2=\sigma_w^2+\sigma_{z_i}^2$. To make the notation more intuitive, we also denote $\sigma_{\tilde{w}_i}^2$ by $\sigma_{\mathrm{in},i}^2$. Now consider the popular denoiser of the form [11]

$$\hat{x}_i=\eta\left(\tilde{y}_i;\theta_i\right)=\frac{\left(|\tilde{y}_i|-\theta_i\right)_+}{|\tilde{y}_i|}\tilde{y}_i, \tag{20}$$

where $(x)_+=\max(x,0)$, and θ_i is a threshold. Define the mean squared error of this denoiser by $\sigma_{\mathrm{out},i}^2:=\mathrm{E}\left[|\hat{x}_i-x_i|^2\right]$. It can be shown that for a given sparsity probability $\epsilon>0$, the worst case prior distribution (maximising the $\sigma_{\mathrm{out},i}^2$) of x_i is given by $p(x_i)=\frac{\epsilon}{2}\delta_{x_i=+\infty}+\frac{\epsilon}{2}\delta_{x_i=-\infty}+(1-\epsilon)\delta_{x_i=0}$, where δ is the Dirac function. Under this assumption, the optimal threshold θ_i (to minimise $\sigma_{\mathrm{out},i}^2$) and the corresponding mean squared error $\sigma_{\mathrm{out,i}}^2$ are given by

$$\theta_i=\alpha^*\sigma_{\mathrm{in},i}, \tag{21}$$

$$\alpha^*:=\arg\min_{\alpha}M^{\#}(\alpha,\epsilon) \tag{22}$$

$$M^{\#}(\alpha,,\epsilon):=\epsilon\left(1+\alpha^2\right)+(1-\epsilon)\cdot\left[2\left(1+\alpha^2\right)\Phi(-\alpha)-2\alpha\phi(\alpha)\right], \tag{23}$$

$$\sigma_{\mathrm{out},i}^2=M\left(\sigma_{\mathrm{in},i}^2,\epsilon\right):=M^{\#}\left(\alpha^*,\epsilon\right)\sigma_{\mathrm{in},i}^2. \tag{24}$$

See [11] for more details on the derivations and discussions of this minimax approach.

Algorithm 1 IAMP

Input:
$\boldsymbol{y}$: the observation vector.
$\boldsymbol{A}=[\boldsymbol{A}_1,\boldsymbol{A}_2,\boldsymbol{A}_3,...,\boldsymbol{A}_n]$: the measurement matrix.
σ_w^2: the noise variance.
ϵ: the nonzero probability (defined as the ratio between the number of nonzero elements in $\boldsymbol{x}$ and n the dimension of $\boldsymbol{x}$).
Output:
$\hat{\boldsymbol{x}}$: the estimated signal.
Initialization:
Let $\boldsymbol{r}^0=\boldsymbol{y}$, $\boldsymbol{x}^0=0$ and $t=0$.
Set $\sigma_{\mathrm{out},i}^2=\sigma_x^2=\left(\|\boldsymbol{y}\|_2^2-m\sigma_w^2\right)/\|\boldsymbol{A}\|_F^2$.
Iteration: In the t-th iteration, do

1) Compute
$$\sigma_{\mathrm{in},i}^2=\sum_{j\neq i}\left(\boldsymbol{A}_i^T\boldsymbol{A}_j\right)^2\sigma_{\mathrm{out},j}^2+\sigma_w^2,\ \forall i\in[n]. \tag{25}$$
2) Let $\tilde{y}_i^t=x_i^t+\sum_a A_{ai}r_a^t$. Update the estimated signal
$$x_i^{t+1}=\eta\left(\tilde{y}_i^t;\theta_i^t\right),\ \forall i\in[n],$$
where the denoiser $\eta(\cdot)$ and the threshold θ_i^t are defined in (20) and (21) respectively.
3) Update the "residual" signal $\boldsymbol{r}^{t+1}$ by
$$r_a^{t+1}=y_a-\sum_{i=1}^n A_{ai}x_i^{t+1}+\sum_{i=1}^n A_{ai}^2\eta'\left(\tilde{y}_i^t;\theta_i^t\right)r_a^t,\ \forall a\in[m]. \tag{26}$$
4) Compute $\sigma_{\mathrm{out},i}^2$ via Equation (24).
5) [Optional] Adjust the "output" noise variance.
Let $\bar{\sigma}_r^2=\frac{1}{m}\sum_{i=1}^n\|\boldsymbol{A}_i\|_2^2\sigma_{\mathrm{out},i}^2$. Set $\sigma_{\mathrm{out},i}^2=c\sigma_{\mathrm{out},i}^2$, where
$$c=\left(\frac{1}{m}\left\|\boldsymbol{r}^{t+1}\right\|_2^2-\sigma_w^2\right)_+/\bar{\sigma}_r^2.$$
6) Go back to step 1 unless the stop criteria are satisfied.

With above notations, the IAMP algorithm is detailed in Algorithm 1. In the initialisation step, an estimation of the variance of σ_x^2 will be needed. From the model $\boldsymbol{y}=\boldsymbol{A}\boldsymbol{x}+\boldsymbol{w}$, it is approximately true that $\|\boldsymbol{A}\boldsymbol{x}\|_2^2=\|\boldsymbol{y}\|_2^2-m\sigma_w^2$. On the other hand, assume that x_i's are independent and $\sigma_{x_i}^2=\sigma_{x_j}^2=\sigma_x^2$, $\forall i\neq j$. Then $\|\boldsymbol{A}\boldsymbol{x}\|_2^2\approx\sum_i\sigma_{x_i}^2\|\boldsymbol{A}_i\|_2^2=\sigma_x^2\|\boldsymbol{A}\|_F^2$. As a result, one can set $\sigma_x^2=\left(\|\boldsymbol{y}\|_2^2-m\sigma_w^2\right)/\|\boldsymbol{A}\|_F^2$.

The major differences between AMP and IAMP are as follows. Assume that $\sigma_{\mathrm{in},j}^2=\sigma_{\mathrm{in},k}^2$ and $\sigma_{\mathrm{out},j}^2=\sigma_{\mathrm{out},k}^2$, $j\neq k$, and therefore $\sigma_{\mathrm{in},i}^2$ and $\sigma_{\mathrm{out},i}^2$ are replaced by σ_{in}^2 and σ_{out}^2 respectively. In AMP, Equation (25) becomes $\sigma_{\mathrm{in}}^2=\frac{1}{\delta}\sigma_{\mathrm{out}}^2+\sigma_w^2$ with $\delta:=\frac{m}{n}$. The implementation of the denoising function η in (2) depends on this information. The second difference is that the last term in (26) becomes $\frac{1}{m}\sum_{i=1}^n\eta'\left(\tilde{y}_i^t;\theta_i^t\right)r_a^t=\frac{1}{\delta}\left\langle\eta'\left(\tilde{y}_i^t;\theta_i^t\right)\right\rangle r_a^t$.

The 'operational' complexity of IAMP is the same as that of AMP. The most computationally intensive step is the evaluation of (25), of which the complexity is $O\left(n^3\right)$. However, $\boldsymbol{A}_i^T\boldsymbol{A}_j$, $\forall i\neq j$, can be computed off-line. All other steps only involve at most $O\left(n^2\right)$ computations.

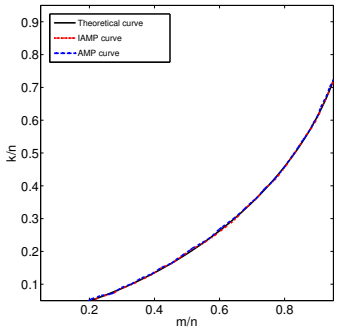


Figure 2: Phase Transition for standard Gaussian matrices.

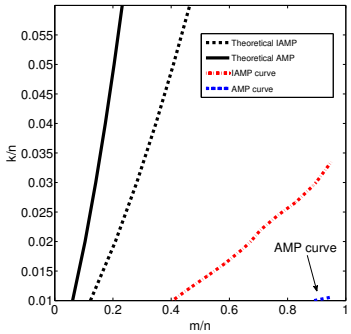


Figure 3: Phase Transition for non-standard Gaussian matrices.

IV. Performance Discussions

In this section, we will first show that if the measurement matrix $\boldsymbol{A}$ is a standard Gaussian random matrix, then IAMP reduces to AMP. Next, we construct a Gaussian random matrix such that the marginal distribution of each entry is still $\mathcal{N}\left(0, \frac{1}{m}\right)$ but the entries are dependent. For this scenario, we show the significant performance improvement of IAMP. Finally, we demonstrated the improvement of IAMP using synthetic data of a real application — radar imaging.

A. The Standard Gaussian Random Matrix

In this subsection, we consider the behavior of IAMP for standard Gaussian random matrices, i.e., the entries are independently generated from $\mathcal{N}\left(0, \frac{1}{m}\right)$. Under this assumption and using the approximation techniques mentioned in [11], the IAMP algorithm can be simplified when the sizes of the system m and n are sufficiently large. In particular, it can be shown that $\boldsymbol{A}_i^T \boldsymbol{A}_j = \frac{1}{m} + o\left(\frac{1}{m}\right)$ and hence Equation (25) becomes $\sigma_{\text{in}}^2 = \frac{1}{\delta}\sigma_{\text{out}}^2 + \sigma_w^2 + o(1)$. Furthermore, each component of the matrix $A_{a,i} = O\left(\frac{1}{\sqrt{m}}\right)$. The last term in (26) becomes $\frac{1}{m}\sum_{i=1}^{n} \eta'\left(\tilde{y}_i^t; \theta_i^t\right) r_a^t = \frac{1}{\delta}\left\langle \eta'\left(\tilde{y}_i^t; \theta_i^t\right)\right\rangle r_a^t$. Hence, IAMP reduces to AMP.

Figure 2 provides the numerical comparison between AMP and IAMP. We consider the noise free case, i.e., $\sigma_w^2 = 0$. We are interested in the phase transition curve, that is, the exact reconstruction happens with dominant probability in the region below the curve while the recovery is not accurate with dominant probability in the region above the curve. (In empirical study, we use 50% probability to draw the phase transition curve.) The theoretic curve is obtained by asymptotic analysis presented in [11]. The empirical results are obtained via 100 independent trials. In the simulations, $n = 1000$ so that asymptotic analysis should be accurate. The simulation results suggest that the theoretical phase transition curve predicts the actual performance and the AMP and IAMP algorithms give the identical numerical performance.

B. Non-standard Gaussian matrices

The more interesting results are obtained when the measurements are not the standard Gaussian matrix. Let $\boldsymbol{B} \in \mathbb{R}^{m\times n}$ be a standard Gaussian random matrix. Let $\boldsymbol{D} \in \mathbb{R}^{m\times m}$ be a diagonal matrix whose first $m/2$ diagonal entries are $\sqrt{\rho/2}$ and the rest $m/2$ diagonal entries are given by $\sqrt{(4-\rho)/2}$, where $\rho \in [0,4]$ is a given design constant. Let $\boldsymbol{H}$ be a normalised Hadamard matrix such that $\boldsymbol{H}^T\boldsymbol{H} = \boldsymbol{H}\boldsymbol{H}^T = \boldsymbol{I}$. Define the measurement matrix as $\boldsymbol{A} = \boldsymbol{H}\boldsymbol{D}\boldsymbol{B}$.

This definition is motivated by equation (9). It is clear that $\boldsymbol{A}$ is a Gaussian random matrix. The marginal distribution of an entry $A_{a,i}$ is given by $A_{a,i} \sim \mathcal{N}\left(0, \frac{1}{m}\right)$. Furthermore, it can be shown that the cross-correlation between two columns has variance given by

$$\begin{aligned}
&\mathrm{E}\left[\left(\boldsymbol{A}_i^T\boldsymbol{A}_j\right)^2\right] = \mathrm{E}\left[\left(\boldsymbol{B}_i^T\boldsymbol{D}^T\boldsymbol{H}^T\boldsymbol{H}\boldsymbol{D}\boldsymbol{B}_j\right)^2\right] \\
&= \mathrm{E}\left[\left(\boldsymbol{B}_i^T\boldsymbol{D}^T\boldsymbol{D}\boldsymbol{B}_j\right)^2\right] = \mathrm{E}\left[\left(\sum_k d_k^2 B_{k,i}B_{k,j}\right)^2\right] \\
&= \frac{1}{m^2}\sum_k d_k^4 = \frac{\rho^2 + (4-\rho)^2}{8m} =: \frac{1}{m}\sigma_c^2(\rho),
\end{aligned}$$

which is not $1/m$ (the value for standard Gaussian random matrix) unless $\rho = 2$.

The resulted IAMP behaves quite different from AMP. Firstly, Equation (25) can be approximated by $\sigma_{\text{in}}^2 = \frac{1}{\delta}\sigma_c^2(\rho)\,\sigma_{\text{out}}^2 + \sigma_w^2$. Secondly, if we mimic the state evolution analysis in [11] for IAMP for the asymptotic case, the theoretic phase transition is given by the implicit equation $M^{\#}(\alpha^*, \epsilon) = \frac{1}{\delta}/\sigma_c^2(\rho)$. Note that $\rho_c^2(\rho) = 2$ when $\rho = 4$, and this means for a given ϵ, twice many measurements are theoretically expected for exact reconstruction.

Figure 3 compares AMP and IAMP. IAMP gives much better performance than AMP, and theoretic prediction of IAMP is also better than that of AMP. Unfortunately, in this case, neither of the theoretical predictions is accurate.

C. Radar Imaging

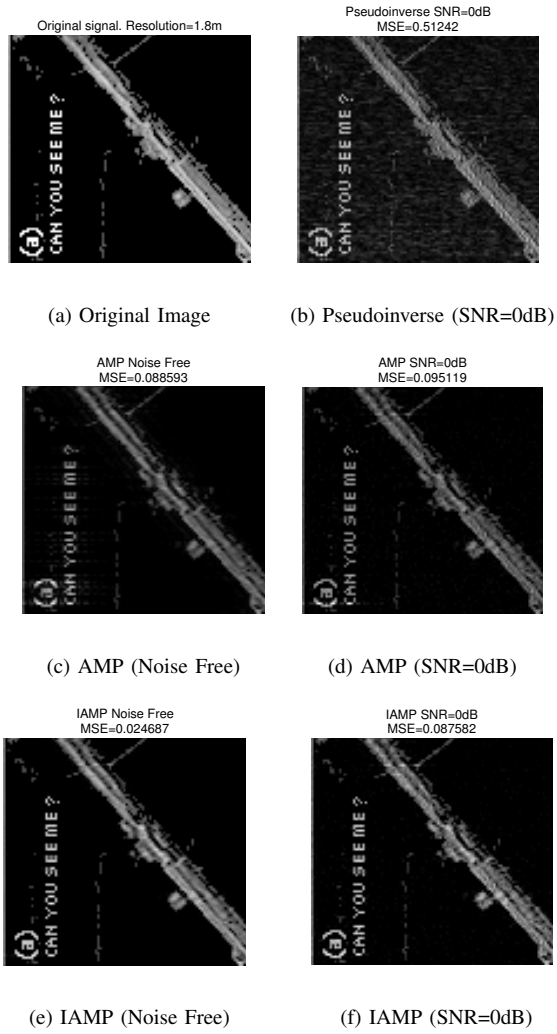
For simplicity, we consider the 1-D radar imaging. (The 2-D image in Figure 4 is obtained by scan the picture line by line. The size of the image is 207×194) A linear frequency modulated signal is transmitted and reflected by the existing targets in the scene. The received signal is then the superposition of the reflected signals. When the number of existing targets is small, this superposition is sparse. Depending on the distances between the radar system and the targets, the

reflected signals are scaled versions of the transmitted signal with different delays. Mathematically, the received signal is given by $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$, where columns of $\boldsymbol{A} \in \mathbb{C}^{m\times n}$ are the transmitted signal with different delays, and $\boldsymbol{x} \in \mathbb{C}^n$ denotes the reflection coefficient vector and is sparse. The matrix $\boldsymbol{A}$ has two interesting structures. First, it is deterministic and Toeplitz. Second, it is tall rather than flat. Here we do not consider the compressed sensing scenario, i.e., no sub-sampling is performed. In practice, the sampling rate can be very high resulting $m > n$.

The simulated results are given in Figure 4. Besides AMP and IAMP, the least squares approach is also included. This is motivated by the fact that least squares approach can perfectly recover the signal $\boldsymbol{x}$ for the noise free case. However, least squares approach cannot incorporate the sparse prior information and therefore does not give a sparse solution for the noisy case. Figure 4b demonstrates this point at SNR=0dB. By contrast, due to accommodating sparse prior information, both AMP and IAMP perform well consistently for both high SNR and low SNR. As of the comparison between AMP and IAMP, it can be observed that IAMP results in less artifacts (see bottom left corner of Figure 4c) and sharper images.[1] In summary, among the tested algorithms, IAMP is the most robust one against the noise.

V. Conclusions

An improved AMP algorithm has been derived for non-ideal measurement matrices. The performance improvement has been demonstrated by using a particularly constructed Gaussian matrix and a matrix from real applications. It turns out that the improvement is obtained by considering the correlation profile of the columns of the measurement matrix.



(a) Original Image (b) Pseudoinverse (SNR=0dB)

(c) AMP (Noise Free) (d) AMP (SNR=0dB)

(e) IAMP (Noise Free) (f) IAMP (SNR=0dB)

Figure 4: Radar imaging.

References

[1] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM journal on scientific computing*, vol. 20, no. 1, pp. 33–61, 1998.

[2] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *Information Theory, IEEE Transactions on*, vol. 52, pp. 489–509, Feb 2006.

[3] T. Jin, M. Yanting, and D. Baron, "Compressive imaging via approximate message passing with image denoising," *Signal Processing, IEEE Transactions on*, vol. 63, pp. 2085–2092, April 2015.

[4] A. Tavakoli and A. Pourmohammad, "Image denoising based on compressed sensing," *International Journal of Computer Theory and Engineering*, vol. 4, no. 2, pp. 266–269, 2012.

[5] A. Maleki and D. Donoho, "Optimally tuned iterative reconstruction algorithms for compressed sensing," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 4, pp. 330–341, April 2010.

[6] D. Wei and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *Information Theory, IEEE Transactions on*, vol. 55, pp. 2230–2249, May 2009.

[7] E. Candes and T. Tao, "Decoding by linear programming," *Information Theory, IEEE Transactions on*, vol. 51, pp. 4203–4215, Dec 2005.

[8] Y. Pati, R. Rezaiifar, and P. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," in *Signals, Systems and Computers, 1993. 1993 Conference Record of The Twenty-Seventh Asilomar Conference on*, pp. 40–44 vol.1, Nov 1993.

[1] It is interesting to observe that the visual performance of AMP improves when SNR decreases. We don't fully understand the reason but suspect that it may be because the biased estimation of σ^2_{in} (25) is neutralised by the large noise variance σ^2_w.

[9] J. Tropp, D. Needell, and R. Vershynin, "Iterative signal recovery from incomplete and inaccurate measurements," *Information Theory and Applications*, 2008.

[10] D. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing: Ii. analysis and validation," in *Information Theory (ITW 2010, Cairo), 2010 IEEE Information Theory Workshop on*, pp. 1–5, Jan 2010.

[11] A. Montanari, "Graphical models concepts in compressed sensing," *Compressed Sensing: Theory and Applications*, pp. 394–438, 2012.

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

[13] M. Bayati and A. Montanari, "The dynamics of message passing on dense graphs, with applications to compressed sensing," in *Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on*, pp. 1528–1532, June 2010.

[14] S. Rangan, "Estimation with random linear mixing, belief propagation and compressed sensing," in *Information Sciences and Systems (CISS), 2010 44th Annual Conference on*, pp. 1–6, March 2010.

[15] S. Rangan, P. Schniter, and A. Fletcher, "On the convergence of approximate message passing with arbitrary matrices," in *Information Theory (ISIT), 2014 IEEE International Symposium on*, pp. 236–240, June 2014.

[16] A. Manoel, F. Krzakala, E. Tramel, and L. Zdeborov, "Sparse estimation with the swept approximated message-passing algorithm," *arXiv preprint arXiv:1406.4311*, 2014.

[17] S. Rangan, A. Fletcher, P. Schniter, and U. Kamilov, "Inference for generalized linear models via alternating directions and bethe free energy minimization," *arXiv preprint arXiv:1501.01797*, 2015.