

# Performance Trade-Offs in Multi-Processor Approximate Message Passing

Junan Zhu,<sup>\*</sup> Ahmad Beirami,<sup>†</sup> and Dror Baron<sup>\*</sup>

<sup>\*</sup>Department of Electrical and Computer Engineering, North Carolina State University, Email: {jzhu9,barondror}@ncsu.edu

<sup>†</sup>Research Laboratory of Electronics, Massachusetts Institute of Technology, Email: beirami@mit.edu

**Abstract**—We consider large-scale linear inverse problems in Bayesian settings. Our general approach follows a recent line of work that applies the approximate message passing (AMP) framework in multi-processor (MP) computational systems by storing and processing a subset of rows of the measurement matrix along with corresponding measurements at each MP node. In each MP-AMP iteration, nodes of the MP system and its fusion center exchange lossily compressed messages pertaining to their estimates of the input. There is a trade-off between the physical costs of the reconstruction process including computation time, communication loads, and the reconstruction quality, and it is impossible to simultaneously minimize all the costs. We pose this minimization as a multi-objective optimization problem (MOP), and study the properties of the best trade-offs (Pareto optimality) in this MOP. We prove that the achievable region of this MOP is convex, and conjecture how the combined cost of computation and communication scales with the desired mean squared error. These properties are verified numerically.

**Index Terms**—approximate message passing, distributed linear systems, multi-objective optimization, Pareto optimality.

## I. INTRODUCTION

Many scientific and engineering problems [1, 2] can be approximated as linear systems of the form

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{z}, \quad (1)$$

where  $\mathbf{x} \in \mathbb{R}^N$  is the unknown input signal,  $\mathbf{A} \in \mathbb{R}^{M \times N}$  is the matrix that characterizes the linear system, and  $\mathbf{z} \in \mathbb{R}^M$  is measurement noise. The goal is to estimate  $\mathbf{x}$  from the noisy measurements  $\mathbf{y}$  given  $\mathbf{A}$  and statistical information about  $\mathbf{z}$ . Alternately, one could view the estimation of  $\mathbf{x}$  as fitting or learning a linear model for the data comprised of  $\mathbf{y}$  and  $\mathbf{A}$ .

When  $M \ll N$ , the setup (1) is known as compressed sensing (CS) [1, 2]; by posing a sparsity or compressibility requirement on the signal, it is indeed possible to accurately recover  $\mathbf{x}$  from the ill-posed linear system [1, 2] when the number of measurements  $M$  is large enough, and the noise level is modest. However, we might need  $M > N$  when the signal is dense or the noise is substantial.

Approximate message passing (AMP) [3–5] is an iterative framework that solves linear inverse problems by successively decoupling [6–8] matrix channel problems into scalar channel denoising problems with additive white Gaussian noise (AWGN). AMP has received considerable attention, because of its fast convergence and the state evolution (SE)

formalism [3, 5], which offers a precise characterization of the AWGN denoising problem in each iteration. In the Bayesian setting, AMP often achieves the minimum mean squared error (MMSE) [9, 10] in the limit of large linear systems.

In real-world applications, a multi-processor (MP) version of CS could be of interest, due to either storage limitations in each individual processor node, or the need for fast computation. This paper considers multi-processor CS (MP-CS) [11–16], in which there are  $P$  distributed nodes (processor nodes) and a fusion center. Each distributed node stores  $\frac{M}{P}$  rows of the matrix  $\mathbf{A}$ , and acquires the corresponding linear measurements of the underlying signal  $\mathbf{x}$ . Without loss of generality, we model the measurement system in distributed node  $p \in \{1, \dots, P\}$  as

$$y_i = \mathbf{A}_i \mathbf{x} + z_i, \quad i \in \left\{ \frac{M(p-1)}{P} + 1, \dots, \frac{Mp}{P} \right\}, \quad (2)$$

where  $\mathbf{A}_i$  is the  $i$ -th row of  $\mathbf{A}$ , and  $y_i$  and  $z_i$  are the  $i$ -th entries of  $\mathbf{y}$  and  $\mathbf{z}$ , respectively. Once every  $y_i$  is collected, we run distributed algorithms among the fusion center and  $P$  distributed nodes to reconstruct the signal  $\mathbf{x}$ . MP versions of AMP (MP-AMP) for MP-CS have been studied in the literature [13, 16]. Usually, MP platforms are designed for distributed settings such as sensor networks [17, 18] or large-scale “big data” computing systems [19]. We reduce the communication costs of MP platforms by applying lossy compression [20–22] to the communication portion of MP-AMP.

In this paper, we consider a rich design space that includes various costs, such as the number of iterations  $T$ , aggregate coding rates  $R_{agg}$  (defined later in (13)), and the mean squared error (MSE) achieved by the reconstruction algorithm. In such a rich design space, reducing any cost is likely to incur an increase in other costs, and it is difficult to simultaneously minimize all the costs. Han et al. [13] reduce the communication costs, and Ma et al. [23] develop an algorithm with reduced computation; both works [13, 23] achieve a reasonable MSE. However, the optimal trade-offs in this rich design space are not studied. We pose the problem of finding the best trade-offs among the individual costs  $T$ ,  $R_{agg}$ , and MSE as a multi-objective optimization problem (MOP), and study the properties of the Pareto optimal tuples  $(T^*, R_{agg}^*, \text{MSE}^*)$  [24] of this MOP. (Note that we do not intend to provide a practical implementation to achieve the optimal trade-offs.) Finally, we conjecture that the combined cost of computation and communication scales as  $O(\log^2(1/(\text{MSE} - \text{MMSE})))$ ;

The work was supported by the National Science Foundation under the Grant CCF-1217749.

these properties are verified numerically using a dynamic programming (DP, cf. Bertsekas [25]) scheme from our prior work [26].

## II. BACKGROUND

### A. Centralized CS using AMP

In the linear system (1), we consider an independent and identically distributed (i.i.d.) Gaussian measurement matrix  $\mathbf{A}$ , i.e.,  $\mathbf{A}_{i,j} \sim \mathcal{N}(0, \frac{1}{M})$ . The signal entries follow an i.i.d. *Bernoulli Gaussian* distribution,

$$x_j \sim \epsilon \mathcal{N}(0, 1) + (1 - \epsilon) \delta(x_j), \quad (3)$$

where  $\delta(\cdot)$  is the Dirac delta function and  $\epsilon$  is called the *sparcity rate* of the signal. The noise entries obey  $z_i \sim \mathcal{N}(0, \sigma_Z^2)$ , where  $\sigma_Z^2$  is the noise variance. Note that the results in this paper can be easily extended to priors other than (3).

Starting from  $\mathbf{x}_0 = \mathbf{0}$ , the AMP framework [3] proceeds iteratively according to

$$\mathbf{x}_{t+1} = \eta_t(\mathbf{A}^T \mathbf{r}_t + \mathbf{x}_t), \quad (4)$$

$$\mathbf{r}_t = \mathbf{y} - \mathbf{A} \mathbf{x}_t + \frac{1}{\kappa} \mathbf{r}_{t-1} \langle d\eta_{t-1}(\mathbf{A}^T \mathbf{r}_{t-1} + \mathbf{x}_{t-1}) \rangle, \quad (5)$$

where  $\eta_t(\cdot)$  is a denoising function,  $d\eta_t(\cdot) = \frac{d\eta_t(\cdot)}{d\{\cdot\}}$  is shorthand for the derivative of  $\eta_t(\cdot)$ , and  $\langle \mathbf{u} \rangle = \frac{1}{N} \sum_{i=1}^N u_i$  for some vector  $\mathbf{u} \in \mathbb{R}^N$ . The subscript  $t$  represents the iteration index,  $\mathcal{T}$  denotes transpose, and  $\kappa = \frac{M}{N}$  is the measurement rate. Owing to the decoupling effect [6–8], in each AMP iteration [4, 5], the vector  $\mathbf{f}_t = \mathbf{A}^T \mathbf{r}_t + \mathbf{x}_t$  in (4) is statistically equivalent to the input signal  $\mathbf{x}$  corrupted by AWGN  $\mathbf{w}_t$  generated by a source  $W \sim \mathcal{N}(0, \sigma_t^2)$ ,

$$\mathbf{f}_t = \mathbf{x} + \mathbf{w}_t. \quad (6)$$

In large systems ( $N \rightarrow \infty, \frac{M}{N} \rightarrow \kappa$ ), a useful property of AMP [4, 5] is that the noise variance  $\sigma_t^2$  evolves following state evolution (SE):  $\sigma_{t+1}^2 = \sigma_Z^2 + \frac{1}{\kappa} \text{MSE}(\eta_t, \sigma_t^2)$ , where the mean squared error  $\text{MSE}(\eta_t, \sigma_t^2) = \mathbb{E}_{X,W} \left[ (\eta_t(X+W) - X)^2 \right]$ ,  $\mathbb{E}_{X,W}(\cdot)$  is expectation with respect to  $X$  and  $W$ , and  $X \sim f_X$  is the source that generates  $\mathbf{x}$ . Note that  $\sigma_1^2 = \sigma_Z^2 + \frac{\mathbb{E}[X^2]}{\kappa}$ , because of the all-zero initial estimate for  $\mathbf{x}$ . Formal statements for SE appear in prior work [4, 5].

This paper considers the Bayesian setting, in which we assume knowledge of the true prior for the signal  $\mathbf{x}$ . Therefore, the MMSE-achieving denoiser is the conditional expectation,  $\eta_t(\cdot) = \mathbb{E}[\mathbf{x} | \mathbf{f}_t]$ , which can be easily obtained. Other denoisers such as soft thresholding [3–5] yield MSE's that are greater than that of the Bayesian denoiser. When the true prior for  $\mathbf{x}$  is unavailable, parameter estimation techniques can be used.

### B. MP-CS using lossy MP-AMP

In the sensing problem formulated in (2), the measurement matrix is stored in a distributed manner in each distributed node. Lossy MP-AMP [16, 26] iteratively solves MP-CS problems using lossily compressed messages:

$$\text{Distributed nodes: } \mathbf{r}_t^p = \mathbf{y}^p - \mathbf{A}^p \mathbf{x}_t + \frac{1}{\kappa} \mathbf{r}_{t-1}^p g_{t-1}, \quad (7)$$

$$\mathbf{f}_t^p = \frac{1}{P} \mathbf{x}_t + (\mathbf{A}^p)^T \mathbf{r}_t^p, \quad (8)$$

$$\text{Fusion center: } \mathbf{f}_{Q,t} = \sum_{p=1}^P Q(\mathbf{f}_t^p), \quad g_t = \langle d\eta_t(\mathbf{f}_{Q,t}) \rangle, \quad (9)$$

$$\mathbf{x}_{t+1} = \eta_t(\mathbf{f}_{Q,t}), \quad (10)$$

where  $Q(\cdot)$  denotes quantization, and an MP-AMP iteration refers to the process from (7) to (10). The reader might notice that the fusion center also needs to transmit the denoised signal vector  $\mathbf{x}_t$  and a scalar  $g_{t-1}$  to the distributed nodes. The transmission of the scalar  $g_{t-1}$  is negligible, and the fusion center may broadcast  $\mathbf{x}_t$  so that naive compression of  $\mathbf{x}_t$ , such as compression with a fixed quantizer, is sufficient. Hence, we will not discuss possible lossy compression of the messages transmitted by the fusion center.

Assume that we quantize  $\mathbf{f}_t^p, \forall p$ , and use  $C$  bits to encode the quantized vector  $Q(\mathbf{f}_t^p) \in \mathbb{R}^N$ . The *coding rate* is  $R = \frac{C}{N}$ . We incur a *distortion* (or quantization error)  $D_t = \frac{1}{N} \sum_{i=1}^N (Q(f_{t,i}^p) - f_{t,i}^p)^2$  at iteration  $t$  in each distributed node,<sup>1</sup> where  $Q(f_{t,i}^p)$  and  $f_{t,i}^p$  are the  $i$ -th entries of the vectors  $Q(\mathbf{f}_t^p)$  and  $\mathbf{f}_t^p$ , respectively. The rate distortion function, denoted by  $R(D)$ , offers the fundamental information theoretic limit on the coding rate  $R$  for communicating a sequence up to distortion  $D$  [20–22, 27]. A pivotal conclusion from RD theory is that coding rates can be greatly reduced even if  $D$  is quite small. The function  $R(D)$  can be computed in various ways [28–30], and can be achieved by an RD-optimal quantization scheme. Other quantization schemes require larger coding rates to achieve the same expected distortion  $D$ .

The goal of this paper is to understand the fundamental trade-offs for MP-CS using MP-AMP. Hence, throughout this paper, we assume that appropriate vector quantization (VQ) schemes [22, 31, 32] that achieve  $R(D)$  are applied within each MP-AMP iteration, although our analysis is readily extended to practical quantizers such as scalar quantizer with entropy coding [21, 22]. Therefore, the signal *at the fusion center* before denoising can be modeled as

$$\mathbf{f}_{Q,t} = \sum_{p=1}^P Q(\mathbf{f}_t^p) = \mathbf{x} + \mathbf{w}_t + \mathbf{n}_t, \quad (11)$$

where  $\mathbf{w}_t$  is the equivalent scalar channel noise (6) and  $\mathbf{n}_t$  is the overall quantization error whose entries follow  $\mathcal{N}(0, PD_t)$ . For large block sizes, we expect the VQ quantization error,  $\mathbf{n}_t$ , to resemble Gaussian noise, which is independent of  $\mathbf{x} + \mathbf{w}_t$ . The SE for the lossy MP-AMP [16, 26] follows

$$\sigma_{t+1}^2 = \sigma_Z^2 + \frac{1}{\kappa} \text{MSE}(\eta_t, \sigma_t^2 + PD_t), \quad (12)$$

where  $\sigma_t^2$  can be estimated by  $\hat{\sigma}_t^2 = \frac{1}{M} \|\mathbf{r}_t\|_2^2$  with  $\|\cdot\|_p$  denoting the  $\ell_p$  norm [4, 5], and  $\sigma_{t+1}^2$  is the variance of  $\mathbf{w}_{t+1}$ . The rigorous justification of (12) by extending Bayati and Montanari [5] is left for future work.

<sup>1</sup>Because we assume that the matrix  $\mathbf{A}$  and noise  $\mathbf{z}$  are both i.i.d., the expected distortions are the same over all  $P$  nodes, and can be denoted by  $D_t$  for simplicity. Other distortion metrics  $d(\cdot, \cdot)$  can also be used [21].

Denote the coding rate used to transmit  $Q(\mathbf{f}_t^p)$  at iteration  $t$  by  $R_t$ . The sequence of  $R_t$ ,  $t = 1, \dots, T$ , where  $T$  is the total number of MP-AMP iterations, is called the *coding rate sequence*, and is denoted by the vector  $\mathbf{R} = [R_1, \dots, R_T]$ . Given the coding rate sequence  $\mathbf{R}$ , the distortion  $D_t$  can be evaluated with  $R(D)$ , and the scalar channel noise variance  $\sigma_t^2$  can be evaluated with (12). Hence, the MSE for  $\mathbf{R}$  can be predicted; we call it SE-predicted MSE. The MSE at the last iteration is called the *final MSE*.

### III. ACHIEVABLE PERFORMANCE REGION

Following the discussion of Sec. II, we can see that the lossy compression of  $\mathbf{f}_t^p, \forall p \in \{1, \dots, P\}$ , can reduce communication costs. On the other hand, the greater the savings in the coding rate sequence  $\mathbf{R}$ , the worse the final MSE is expected to be. If a certain level of final MSE is desired under a small budget of coding rates, more iterations  $T$  will be needed. Define the aggregate coding rate  $R_{agg}$  as the sum of all the coding rates in  $\mathbf{R}$ ,

$$R_{agg} = \|\mathbf{R}\|_1 = \sum_{t=1}^T R_t. \quad (13)$$

As mentioned above, there is a trade-off between  $T$ ,  $R_{agg}$ , and the final MSE, and there is no optimal solution that minimizes them simultaneously. To deal with such trade-offs in a multi-objective optimization (MOP) problem, it is customary to think about the concept of *Pareto optimality* [24].

#### A. Properties of achievable region

Define the computation cost rate,  $C_1$ , as the cost of computation in one MP-AMP iteration, and define the communication cost rate,  $C_2$ , as the cost of transmitting 1 bit for  $Q(\mathbf{f}_t^p)$  (9). We further define the *relative cost* as

$$b = \frac{C_1}{C_2}. \quad (14)$$

For notational convenience, denote by  $\mathcal{E}(T, R_{agg})$  all of the MSE values that would be provided by the pair  $(T, R_{agg})$  for some relative cost  $b$  (14), among which the smallest MSE is denoted by  $\text{MSE}^*(R_{agg}, T)$ . Furthermore, the achievable set  $\mathcal{C}$  is defined as<sup>2</sup>

$$\mathcal{C} := \{(T, R_{agg}, \text{MSE}) \in \mathbb{R}_{\geq 0}^3 : \text{MSE} \in \mathcal{E}(T, R_{agg})\},$$

i.e., there exists an instantiation of the MP-AMP algorithm that could reconstruct the signal with  $T$  iterations and an aggregate coding rate  $R_{agg}$ , and yield a certain MSE.

*Definition 1:* The point  $\mathcal{X}_1 \in \mathcal{C}$  is said to dominate another point  $\mathcal{X}_2 \in \mathcal{C}$ , denoted by  $\mathcal{X}_1 \prec \mathcal{X}_2$ , if and only if  $T_1 \leq T_2$ ,  $R_{agg1} \leq R_{agg2}$ , and  $\text{MSE}_1 \leq \text{MSE}_2$ . A point  $\mathcal{X}^* \in \mathcal{C}$  is said to be Pareto optimal if and only if there does not exist  $\mathcal{X} \in \mathcal{C}$  satisfying  $\mathcal{X} \prec \mathcal{X}^*$ . Furthermore, let  $\mathcal{P}$  denote the set of all Pareto optimal points,

$$\mathcal{P} := \{\mathcal{X} \in \mathcal{C} : \mathcal{X} \text{ is Pareto optimal}\}. \quad (15)$$

<sup>2</sup> $\mathbb{R}_{\geq 0}$  denotes the set of non-negative real numbers.

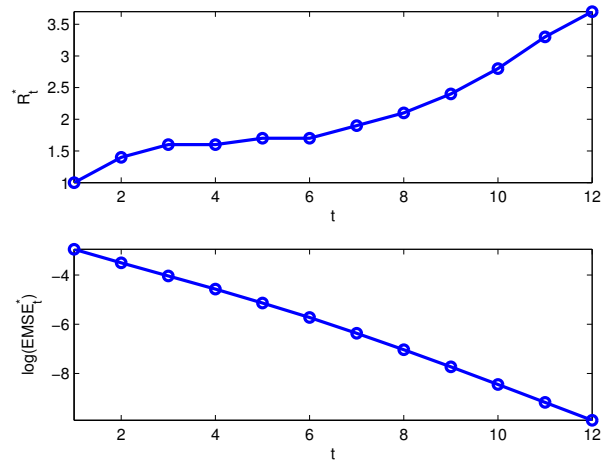


Fig. 1. The optimal coding rate sequence  $\mathbf{R}$  (top panel) and  $\text{EMSE}_t^*$  (bottom panel) are shown as functions of  $t$ . ( $\kappa = 0.4$ ,  $P = 100$ ,  $\sigma_Z^2 = \frac{1}{400}$ , and  $b = 2$ .)

In words, the tuple  $(T, R_{agg}, \text{MSE})$  is Pareto optimal if no other tuple  $(T', R'_{agg}, \text{MSE}')$  exists such that  $T' \leq T$ ,  $R'_{agg} \leq R_{agg}$ , and  $\text{MSE}' \leq \text{MSE}$ . These are the tuples that belong to the boundary of  $\mathcal{C}$ .

We extend the definition of the number of iterations  $T$  to a probabilistic one. We assume that the number of iterations is drawn from a probability distribution  $\pi$  over  $\mathbb{N}$ , such that  $\sum_{i=1}^{\infty} \pi_i = 1$ . Of course, this definition contains a deterministic  $T = j$  as a special case with  $\pi_j = 1$  and  $\pi_i = 0$  for all  $i \neq j$ . Armed with this definition of Pareto optimality and the probabilistic definition of the number of iterations, we have the following lemma.

*Lemma 1:* For a fixed noise variance  $\sigma_Z^2$ , measurement rate  $\kappa$ , and  $P$  distributed nodes in MP-AMP, the achievable set  $\mathcal{C}$  is a convex set.

*Proof:* We need to show that for any  $(T^{(1)}, R_{agg}^{(1)}, \text{MSE}^{(1)})$ ,  $(T^{(2)}, R_{agg}^{(2)}, \text{MSE}^{(2)}) \in \mathcal{C}$  and any  $0 < \lambda < 1$ ,

$$(\lambda T^{(1)} + (1 - \lambda)T^{(2)}, \lambda R_{agg}^{(1)} + (1 - \lambda)R_{agg}^{(2)}, \lambda \text{MSE}^{(1)} + (1 - \lambda)\text{MSE}^{(2)}) \in \mathcal{C}. \quad (16)$$

We use the well-known time-sharing argument (see Cover and Thomas [21]). Assume that  $(T^{(1)}, R_{agg}^{(1)}, \text{MSE}^{(1)})$ ,  $(T^{(2)}, R_{agg}^{(2)}, \text{MSE}^{(2)}) \in \mathcal{C}$  are achieved by probability distributions  $\pi^{(1)}$  and  $\pi^{(2)}$ , respectively. Let us select all the parameters of the first tuple with probability  $\lambda$  and those of the second tuple with probability  $(1 - \lambda)$ . Hence, we have  $\pi = \lambda\pi^{(1)} + (1 - \lambda)\pi^{(2)}$ . Due to the linearity of expectation, we have  $T = \lambda T^{(1)} + (1 - \lambda)T^{(2)}$ , and  $\text{MSE} = \lambda \text{MSE}^{(1)} + (1 - \lambda)\text{MSE}^{(2)}$ . Again, due to the linearity of expectation,  $R_{agg} = \lambda R_{agg}^{(1)} + (1 - \lambda)R_{agg}^{(2)}$ , implying that (16) is satisfied, and the proof is complete. ■

*Definition 2:* Let the function  $R^*(T, \text{MSE}) : \mathbb{R}_{\geq 0}^2 \rightarrow \mathbb{R}_{\geq 0}$  be the Pareto optimal rate function, which is implicitly described as  $R^*(T, \text{MSE}) = R_{agg}^* \Leftrightarrow (T, R_{agg}^*, \text{MSE}) \in \mathcal{P}$ . We further define implicit functions  $T^*(\text{MSE}, R_{agg})$  and

$MSE^*(R_{agg}, T)$  in a similar way.

*Corollary 1:* The functions  $R^*(T, MSE)$ ,  $T^*(MSE, R_{agg})$ , and  $MSE^*(R_{agg}, T)$  are convex in their arguments.

Note that our proof for the convexity of the set  $\mathcal{C}$  might be extended to other distributed iterative learning algorithms that might use lossy compression.

### B. Limiting performance of MP-AMP

These discussions raise the question whether we can provide some asymptotic analysis of the achievable region. We believe that such an analysis is indeed possible in the limit of MSE that approaches the MMSE. Define the excess MSE (EMSE) [33],  $EMSE = MSE - MMSE$ . Consider a case where we aim to reach a very low EMSE. Montanari [4] provided a graphical interpretation of the relation between the MSE performance of AMP in iteration  $t$  and the statistical properties of the denoiser  $\eta_t(\cdot)$  being used. In the limit of small EMSE, the EMSE decreases by a nearly-constant multiplicative factor in each AMP iteration, yielding a geometric decay of the MMSE. In MP-AMP, in addition to the equivalent scalar channel noise  $\mathbf{w}_t$ , we have additive quantization error  $\mathbf{n}_t$  (11). In order for the EMSE in an MP-AMP system to decay geometrically, the quantization error  $D_t$  must decay at least as quickly. To obtain this geometric decay in  $D_t$ , recall that in the high resolution limit, the distortion-rate function typically takes the form  $D(R) = C_3 2^{-2R}$  [22], where  $C_3 > 0$  is some constant. We propose for  $R_t$  to have the form,

$$R_t = C_4 + C_5 t, \quad (17)$$

where  $C_4$  and  $C_5$  are constants. This rate will not yield a distortion that decays exactly geometrically, because the distribution of  $\mathbf{f}_t^p$  will be dependent on  $t$ . That said, in the limit of small EMSE, the distribution barely changes between iterations, and so it is plausible to expect  $D_t \approx C_6 2^{-C_7 t} [1 + o_t(1)]$ , where the decay rate  $C_7$  is a function of the extra coding rate  $C_5$  per iteration (17), and the multiplicative term  $1 + o_t(1)$  converges to 1 in the limit of large  $t$ , because the distribution barely changes between iterations for large  $t$ . Now that we have driven down the quantization error geometrically, we conjecture that the Pareto optimal EMSE,  $EMSE^* = MSE^* - MMSE$ , decays at the same rate,

$$EMSE_t^* \approx C_8 2^{-C_7 t} [1 + o_t(1)]. \quad (18)$$

Combining (17) and (18), and considering the definition of  $R_{agg}$  (13), the total computation and communication cost is  $O(T) + O(R_{agg}) = O(T^2)$ , which is  $O(\log^2(1/EMSE^*))$ . We have the following conjecture.

*Conjecture 1:* The total computation and communication cost scales as  $O(\log^2(1/EMSE^*))$ .

Having provided this conjecture, we back it up numerically by running our unconstrained DP scheme (Sec. IV-A) [26] on a problem with relatively small  $EMSE_T^* = 5 \times 10^{-5}$  in the last iteration  $T$ . Consider reconstructing a Bernoulli Gaussian signal (3) with  $\epsilon = 0.1$ . The signal is measured in an MP platform with  $P = 100$  distributed nodes according to (2). The measurement rate is  $\kappa = \frac{M}{N} = 0.4$ , and the noise variance is  $\sigma_Z^2 = \frac{1}{400}$ . The relative cost is  $b = 2$  (14). Fig. 1 illustrates the

optimal coding rate sequence  $\mathbf{R}$  and  $EMSE_t^*$  as functions of the iteration number  $t$ . It is readily seen that after the first 5–6 iterations the coding rate seems near-linear, which confirms (18); and  $EMSE_t^*$  decays geometrically, as predicted by (18).

## IV. NUMERICAL RESULTS

After proving that the achievable set  $\mathcal{C}$  is convex, we apply the unconstrained DP developed in Zhu and Baron [26] to find the Pareto optimal points for various relative costs (14), and illustrate the convexity of the achievable set.

### A. Unconstrained DP

The unconstrained DP [26] finds a coding rate sequence  $\mathbf{R}$  over the MP-AMP iterations such that the final MSE is less than  $\Delta$ , while achieving the minimum cost  $\Psi$ . The cost  $\Psi$  for a given computation cost rate  $C_1$  and communication cost rate  $C_2$  is a function of the number of remaining iterations ( $T - t$ ) and the current scalar channel noise variance  $\sigma_t^2$  (11). In the basis case,  $T - t = 0$ , the cost is  $C_1 + C_2 R_T$ . After solving the basis case, we iterate back in time by decreasing  $t$ ,

$$\Psi_{T-t}(\sigma_t^2) = \min_{R'} \{C_1 \times \mathbb{1}_{R' \neq 0} + C_2 R' + \Psi_{T-(t+1)}(\sigma_{t+1}^2(R'))\},$$

where  $R'$  is the coding rate used in the current MP-AMP iteration  $t$ ,  $\mathbb{1}_{\mathcal{A}}$  is the indicator function, which is 1 if the condition  $\mathcal{A}$  is met, else 0, and  $\sigma_{t+1}^2(R')$  is the variance of the noise  $\mathbf{w}_{t+1}$  of the scalar channel (11) in the next MP-AMP iteration after transmitting  $\mathbf{f}_t^p$  at rate  $R'$ . A discretized search space of  $\sigma_t^2$  and  $R'$  is utilized [26].

The coding rates  $R'$  that minimize the cost function  $\Psi_{T-t}(\sigma_t^2)$  for different  $t$  and  $\sigma_t^2$  are stored in a table  $\mathcal{R}(t, \sigma_t^2)$ . After the unconstrained DP finishes, we obtain the coding rate sequence  $\mathbf{R}$  from the table  $\mathcal{R}(t, \sigma_t^2)$ .

### B. Pareto optimal points via unconstrained DP

According to Definition 1, the resulting tuple  $(T, \|\mathbf{R}\|_1, \Delta)$  from the unconstrained DP in Sec. IV-A is Pareto optimal. Hence, in this subsection, we run the unconstrained DP to obtain the Pareto optimal points for a certain distributed linear system under various relative costs (14).

Consider the same setting as in Fig. 1, except that we analyze MP platforms [17–19] with a variety of relative costs (14). Running the unconstrained DP scheme developed in Sec. IV-A, we obtain the optimal coding rate sequence  $\mathbf{R}$  that yields the lowest combined cost while helping MP-AMP achieve an MSE that is at most  $\Delta \in \{2, 3, \dots, 6\} \times MMSE$ . In Fig. 2(a), we draw the Pareto optimal surface obtained by running the unconstrained DP; the circles on the surface are the Pareto optimal points we analyzed. Fig. 2(b) plots the aggregate coding rate as a function of different MSE with different optimal numbers of MP-AMP iterations  $T$ . Fig. 2(c) plots the aggregate coding rate as a function of different  $T$  with different optimal MSE. We can see that the surface comprised of the Pareto optimal points is indeed convex.

With stricter requirements on the final MSE (meaning smaller  $\Delta$ ), more iterations  $T$  and greater aggregate coding rates  $R_{agg}$  (13) are needed. Optimal coding rate sequences increase the coding rate to reduce the number of iterations

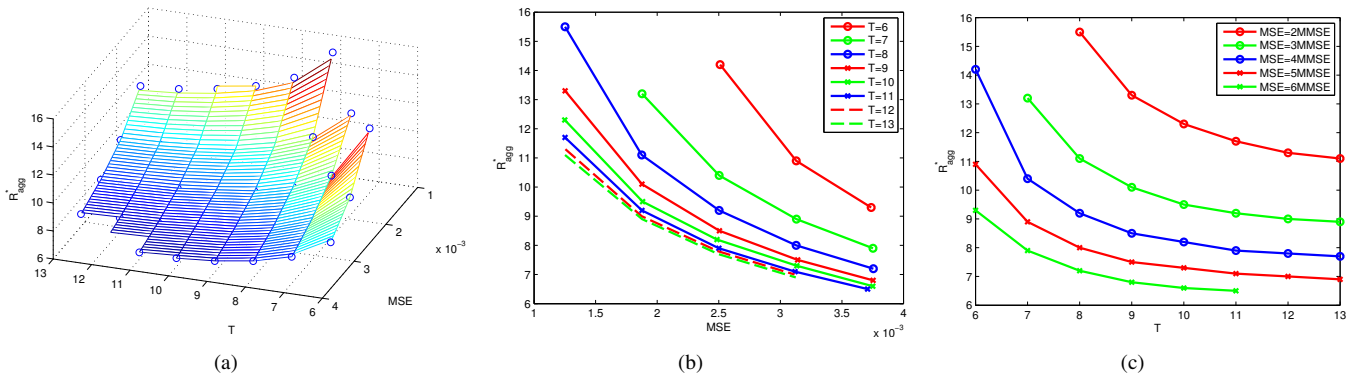


Fig. 2. Pareto optimal results provided by unconstrained DP under a variety of relative costs (14): (a) Pareto optimal surface, (b) Pareto optimal aggregate coding rate  $R_{agg}^*$  (13) versus the achieved MSE for different optimal MP-AMP iterations  $T$ , and (c) Pareto optimal  $R_{agg}^*$  (13) versus the number of iterations  $T$  for different optimal MSE. The signal is Bernoulli Gaussian (3) with  $\epsilon = 0.1$ . ( $\kappa = 0.4$ ,  $P = 100$ , and  $\sigma_Z^2 = \frac{1}{400}$ .)

when communication costs are low [26] (examples are commercial cloud computing systems [19], multi-processor CPUs, and graphic processing units), whereas more iterations allow to reduce the coding rate when communication is costly [26] (for example, in sensor networks [17, 18]).

#### ACKNOWLEDGMENTS

The authors thank Puxiao Han and Ruixin Niu for numerous discussions about MP settings of CS and AMP. We also thank Yanting Ma for useful suggestions.

#### REFERENCES

- [1] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [2] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.
- [3] D. L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing," *Proc. Nat. Academy Sci.*, vol. 106, no. 45, pp. 18914–18919, Nov. 2009.
- [4] A. Montanari, "Graphical models concepts in compressed sensing," *Compressed Sensing: Theory and Applications*, pp. 394–438, 2012.
- [5] M. Bayati and A. Montanari, "The dynamics of message passing on dense graphs, with applications to compressed sensing," *IEEE Trans. Inf. Theory*, vol. 57, no. 2, pp. 764–785, Feb. 2011.
- [6] T. Tanaka, "A statistical-mechanics approach to large-system analysis of CDMA multiuser detectors," *IEEE Trans. Inf. Theory*, vol. 48, no. 11, pp. 2888–2910, Nov. 2002.
- [7] D. Guo and S. Verdú, "Randomly spread CDMA: Asymptotics via statistical physics," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 1983–2010, June 2005.
- [8] D. Guo and C. C. Wang, "Multiuser detection of sparsely spread CDMA," *IEEE J. Select. Areas Commun.*, vol. 26, no. 3, pp. 421–431, Apr. 2008.
- [9] J. Zhu and D. Baron, "Performance regions in compressed sensing from noisy measurements," in *Proc. 2013 Conf. Inference Sci. Syst. (CISS)*, Baltimore, MD, Mar. 2013.
- [10] F. Krzakala, M. Mézard, F. Sausset, Y. Sun, and L. Zdeborová, "Probabilistic reconstruction in compressed sensing: Algorithms, phase diagrams, and threshold achieving matrices," *J. Stat. Mech. - Theory E*, vol. 2012, no. 08, pp. P08009, Aug. 2012.
- [11] J. Mota, J. Xavier, and P. Aguiar, "Distributed basis pursuit," *IEEE Trans. Signal Process.*, vol. 60, no. 4, pp. 1942–1956, Apr. 2012.
- [12] S. Patterson, Y. C. Eldar, and I. Keidar, "Distributed compressed sensing for static and time-varying networks," *IEEE Trans. Signal Proc.*, vol. 62, no. 19, pp. 4931–4946, Oct. 2014.
- [13] P. Han, R. Niu, M. Ren, and Y. C. Eldar, "Distributed approximate message passing for sparse signal recovery," in *Proc. IEEE Global Conf. Signal Inf. Process.*, Atlanta, GA, Dec. 2014, pp. 497–501.
- [14] C. Ravazzi, S. M. Fosson, and E. Magli, "Distributed iterative thresholding for  $\ell_0/\ell_1$ -regularized linear inverse problems," *IEEE Trans. Inf. Theory*, vol. 61, no. 4, pp. 2081–2100, Apr. 2015.
- [15] P. Han, R. Niu, and Y. C. Eldar, "Communication-efficient distributed IHT," in *Proc. Signal Processing with Adaptive Sparse Structured Representations Workshop (SPARS)*, Cambridge, United Kingdom, July 2015.
- [16] P. Han, J. Zhu, R. Niu, and D. Baron, "Multi-processor approximate message passing using lossy compression," in *IEEE Int. Conf. Acoustics, Speech, Signal Process. (ICASSP)*, Shanghai, China, Mar. 2016.
- [17] G. J. Pottie and W. J. Kaiser, "Wireless integrated network sensors," *Commun. ACM*, vol. 43, no. 5, pp. 51–58, May 2000.
- [18] D. Estrin, D. Culler, K. Pister, and G. Sukhatme, "Connecting the physical world with pervasive networks," *IEEE Pervasive Comput.*, vol. 1, no. 1, pp. 59–69, Jan. 2002.
- [19] "Amazon EC2," <https://aws.amazon.com/ec2/>.
- [20] T. Berger, *Rate distortion theory; a mathematical basis for data compression*, Prentice-Hall Englewood Cliffs, NJ, 1971.
- [21] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, New York, NY, USA: Wiley-Interscience, 2006.
- [22] A. Gersho and R. M. Gray, *Vector quantization and signal compression*, Kluwer, 1993.
- [23] Y. Ma, D. Baron, and D. Needell, "Two-part reconstruction with noisy-sudocodes," *IEEE Trans. Signal Process.*, vol. 62, no. 23, pp. 6323–6334, Dec. 2014.
- [24] I. Das and J. E. Dennis, "Normal-boundary intersection: A new method for generating the Pareto surface in nonlinear multicriteria optimization problems," *SIAM J. Optimization*, vol. 8, no. 3, pp. 631–657, Aug. 1998.
- [25] D. P. Bertsekas, *Dynamic programming and optimal control*, vol. 1, Athena Scientific Belmont, MA, 1995.
- [26] J. Zhu and D. Baron, "Multi-processor approximate message passing with lossy compression," *Arxiv preprint arXiv:1601.03790*, Jan. 2016.
- [27] C. Weidmann and M. Vetterli, "Rate distortion behavior of sparse sources," *IEEE Trans. Inf. Theory*, vol. 58, no. 8, pp. 4969–4992, Aug. 2012.
- [28] S. Arimoto, "An algorithm for calculating the capacity of an arbitrary discrete memoryless channel," *IEEE Trans. Inf. Theory*, vol. 18, pp. 14–20, Jan. 1972.
- [29] R. E. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Trans. Inf. Theory*, vol. 18, no. 4, pp. 460–473, July 1972.
- [30] K. Rose, "A mapping approach to rate-distortion computation and analysis," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 1939–1952, Nov. 1994.
- [31] Y. Linde, A. Buzo, and R. M. Gray, "An algorithm for vector quantizer design," *IEEE Trans. Comm.*, vol. 28, no. 1, pp. 84–95, Jan. 1980.
- [32] R. M. Gray, "Vector quantization," *IEEE ASSP Magazine*, vol. 1, no. 2, pp. 4–29, Apr. 1984.
- [33] Y. Ma, D. Baron, and A. Beirami, "Mismatched estimation in large linear systems," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, July 2015, pp. 760–764.