Message-Passing Estimation from Quantized Samples

Ulugbek Kamilov, Vivek K Goyal, and Sundeep Rangan

***Abstract*—Estimation of a vector from quantized linear measurements is a common problem for which simple linear techniques are suboptimal—sometimes greatly so. This paper develops generalized approximate message passing (GAMP) algorithms for minimum mean-squared error estimation of a random vector from quantized linear measurements, notably allowing the linear expansion to be overcomplete or undercomplete and the scalar quantization to be regular or non-regular. GAMP is a recently-developed class of algorithms that uses Gaussian approximations in belief propagation and allows arbitrary separable input and output channels. Scalar quantization of measurements is incorporated into the output channel formalism, leading to the first tractable and effective method for high-dimensional estimation problems involving non-regular scalar quantization. Non-regular quantization is empirically demonstrated to greatly improve rate–distortion performance in some problems with oversampling or with undersampling combined with a sparsity-inducing prior. Under the assumption of a Gaussian measurement matrix with i.i.d. entries, the asymptotic error performance of GAMP can be accurately predicted and tracked through the state evolution formalism. We additionally use state evolution to design MSE-optimal scalar quantizers for GAMP signal reconstruction and empirically demonstrate the superior error performance of the resulting quantizers.**

***Index Terms*—analog-to-digital conversion, approximate message passing, belief propagation, compressed sensing, frames, non-regular quantizers, Slepian–Wolf coding, quantization, Wyner–Ziv coding**

I. Introduction

Estimation of a signal from quantized samples is a fundamental problem in signal processing. It arises both from the discretization in digital acquisition devices and the quantization performed for lossy compression. In some settings, much can be gained from treating quantization with care. A key example is analog-to-digital conversion (ADC), where the advantage from oversampling is increased by replacing conventional linear estimation with nonlinear estimation procedures [1]–[9]. Sophisticated approaches are also helpful when using sparsity or compressibility to reconstruct an undersampled signal [10]–[12].

This paper focuses on using a simple message-passing algorithm based on belief propagation (BP). Implementation of BP for estimation of a continuous-valued quantity requires discretization of densities; this is inherently inexact and leads to high computational complexity. To handle quantization effects without any heuristic additive noise model [13] and with low complexity, we use a recently-developed Gaussian-approximated BP algorithm, called *generalized approximate message passing* (GAMP) [14] or *relaxed belief propagation* [15], which extends earlier methods [16], [17] to nonlinear output channels.

A. Contributions

Our first main contribution is to demonstrate that GAMP provides significantly-improved performance over traditional methods for estimating from quantized samples. Gaussian approximations of BP have previously been shown to be effective in several other applications [15]–[20]; for our application to estimation from quantized samples, the extension to general output channels [14], [15] is essential.

Our second main contribution concerns the quantizer design. When quantizer outputs are used as inputs to a nonlinear estimation algorithm, minimizing the mean-squared error (MSE) between quantizer inputs and outputs is generally not equivalent to minimizing the MSE of the final reconstruction [21]. To optimize the quantizer for the GAMP algorithm, we use the fact that the MSE under large random mixing matrices $\mathbf{A}$ can be predicted accurately from a set of simple state evolution (SE) equations [14], [22]. Then, by modeling the quantizer as a part of the measurement channel, we use the SE formalism to optimize the quantizer to asymptotically minimize distortions after the reconstruction by GAMP. Note that our use of random $\mathbf{A}$ is for rigor of the SE formalism; the effectiveness of GAMP does not depend on this.

B. Outline

The remainder of the paper is organized as follows. Section II provides basic background material on quantization, compressed sensing, and belief propagation. Section III introduces the problem of estimating a random vector from quantized linear transform coefficients. It concentrates on geometric insights for both the oversampled and undersampled settings. The main results in this paper apply under a Bayesian formulation introduced in Section IV. Note that this Bayesian

This material is based upon work supported by the National Science Foundation under Grant No. 0729069 and by the DARPA InPho program through the US Army Research Office award W911-NF-10-1-0404. The material in this paper will be presented in part at the IEEE International Symposium on Information Theory, St. Petersburg, Russian, July–August 2011.

U. Kamilov (email: ulugbek.kamilov@epfl.ch) is with École Polytechnique Fédérale de Lausanne. This work was completed while he was with the Research Laboratory of Electronics, Massachusetts Institute of Technology.

V. K. Goyal (email: vgoyal@mit.edu) is with the Department of Electrical Engineering and Computer Science and the Research Laboratory of Electronics, Massachusetts Institute of Technology.

S. Rangan (email: srangan@poly.edu) is with the Polytechnic Institute of New York University.

formulation does not require sparsity of the signal nor specify undersampling or oversampling. The use of generalized approximate message passing to find optimal estimates under this Bayesian formulation is derived in Section V. Section VI describes the use of state evolution to predict the performance of GAMP for our problem. Optimization of quantizers using state evolution is developed in Section VII, and experimental results are presented in Section VIII. Section IX concludes the paper.

C. Notation

Vectors and matrices will be written in boldface type ($\mathbf{A}$, $\mathbf{x}$, $\mathbf{y}$, ...) to distinguish from scalars written in normal weight (m, n, ...). Random and non-random quantities (or random variables and their realizations) are not distinguished typographically since the use of capital letters for random variables would conflict with the convention of using capital letters for matrices (or in the case of quantization, an operator on a vector rather than a scalar). The probability density function (p.d.f.) of random vector $\mathbf{x}$ is denoted $p_{\mathbf{x}}$, and the conditional p.d.f. of $\mathbf{y}$ given $\mathbf{x}$ is denoted $p_{\mathbf{y}|\mathbf{x}}$. When these densities are separable and identical across components, we repeat the previous notations: $p_{\mathbf{x}}$ for the scalar p.d.f. and $p_{\mathbf{y}|\mathbf{x}}$ for the scalar conditional p.d.f. Writing $x \sim \mathcal{N}(a, b)$ indicates that x is a Gaussian random variable with mean a and variance b. The resulting p.d.f. is written as $p_x(t) = \phi(t\,;\, a,\, b)$.

II. Background

This section establishes concepts and notations central to the paper. For a comprehensive tutorial history of quantization, we recommend [23]; for an introduction to compressed sensing, [24]; and for the basics of belief propagation, [25]–[27].

A. Scalar Quantization

A K-level scalar quantizer $q : \mathbb{R} \to \mathbb{R}$ is defined by its *output levels* or *reproduction points* $\mathcal{C} = \{c_i\}_{i=1}^{K}$ and *(partition) cells* $\{q^{-1}(c_i)\}_{i=1}^{K}$. It can be decomposed into a composition of two mappings $q = \beta \circ \alpha$ where $\alpha : \mathbb{R} \to \{1, 2, \ldots, K\}$ is the *(lossy) encoder* and $\beta : \{1, 2, \ldots, K\} \to \mathcal{C}$ is the *decoder*. The boundaries of the cells are called *decision thresholds*. One may allow $K = \infty$ to denote that $\mathcal{C}$ is countably infinite.

A quantizer is called *regular* when each cell is a convex set, i.e., a single interval. Each cell of a regular scalar quantizer thus has a boundary of one point (if the cell is unbounded) or two points (if the cell is bounded). If the input to a quantizer is a continuous random variable, then the probability of the input being a boundary point is zero. Thus it suffices to specify the cells of a K-point regular scalar quantizer by its decision thresholds $\{b_i\}_{i=0}^{K}$, with $b_0 = -\infty$ and $b_K = \infty$. The encoder satisfies

$$\alpha(x) = i \qquad \text{for } x \in (b_{i-1},\, b_i),$$

and the output for boundary points can be safely ignored.

The lossy encoder of a non-regular quantizer can be decomposed into the lossy encoder of a regular quantizer followed by a many-to-one integer-to-integer mapping. Suppose K-level non-regular scalar quantizer q' has decision thresholds $\{b'_i\}_{i=0}^{K'}$, and let α be the lossy encoder of a regular quantizer with these decision thresholds. Since q' is not regular, $K' > K$. Let $\alpha' : \mathbb{R} \to \{1, 2, \ldots, K\}$ denote the lossy encoder of q'. Then $\alpha' = \lambda \circ \alpha$, where

$$\lambda : \{1, 2, \ldots, K'\} \to \{1, 2, \ldots, K\}$$

is called a *binning function*, *labeling function*, or *index assignment*. The binning function is not invertible.

The *distortion* of a quantizer q applied to scalar random variable x is typically measured by the MSE

$$D = \mathbb{E}[(x - q(x))^2].$$

A quantizer is called optimal at fixed rate $R = \log_2 K$ when it minimizes distortion D among all K-level quantizers. To optimize scalar quantizers under MSE distortion, it suffices to consider only regular quantizers; a non-regular quantizer will never perform strictly better.

While regular quantizers are optimal for the standard lossy compression problem, non-regular quantizers are sometimes useful when some information aside from $q(x)$ is available when estimating x. Two key examples are Wyner–Ziv coding [28] and multiple description coding [29]. One method for Wyner–Ziv coding is to apply Slepian–Wolf coding across a block of samples after regular scalar quantization [30]; the Slepian–Wolf coding is binning, but across a block rather than for a single scalar. In multiple description scalar quantization [31], two binning functions are used that together are invertible but individually are not. In these uses of non-regular quantizers, side information aids in recovering x with resolution commensurate with K' while the rate is only commensurate with K, with $K' > K$.

Optimization of a quantizer can rarely be done exactly or analytically. One standard way of optimizing q is via the *Lloyd algorithm*, which iteratively updates the decision boundaries and output levels by applying necessary conditions for quantizer optimality.

A quantizer $Q : \mathbb{R}^m \to \mathbb{R}^m$ is called a scalar quantizer when it is the Cartesian product of m scalar quantizers $q_i : \mathbb{R} \to \mathbb{R}$. In this paper, Q always represents a scalar quantizer with component quantizers $\{q_i\}_{i=1}^{m}$.

B. Compressed Sensing

Conventionally, one does not attempt to estimate an n-dimensional signal $\mathbf{x}$ from fewer than n scalar quantities; it would not seem to work from a simple counting of degrees of freedom. Compressed sensing (CS) [32]–[34] encapsulates a variety of techniques for estimating $\mathbf{x}$ from $m < n$ scalar linear measurements, possibly including some noise, by exploiting knowledge that $\mathbf{x}$ is sparse or approximately sparse in some given transform domain. Measurements are of the form

$$\mathbf{z} = \mathbf{A}\mathbf{x}, \tag{1}$$

where $\mathbf{A} \in \mathbb{R}^{m \times n}$ is the *measurement matrix*, or

$$\mathbf{y} = \mathbf{z} + \mathbf{d} = \mathbf{A}\mathbf{x} + \mathbf{d}, \tag{2}$$

where $\mathbf{d} \in \mathbb{R}^m$ is additive noise. Many theoretical guarantees for compressed sensing are given with high probability of success over a random selection of $\mathbf{A}$. Note that it is always assumed that $\mathbf{A}$ is available when estimating $\mathbf{x}$ from $\mathbf{z}$ or $\mathbf{y}$.

In this paper, we simplify notation and expressions by assuming that $\mathbf{x}$ itself is sparse or approximately sparse without requiring the use of a transform domain. Also, since our focus is on estimation in the presence of degradation of measurements caused by quantization, we do not consider further the noiseless measurement model (1).

The most commonly-studied estimator for the measurement model (2) is the *lasso* estimator [35]

$$\widehat{\mathbf{x}} = \underset{\mathbf{x}\in\mathbb{R}^n}{\arg\min} \left(\tfrac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \gamma\|\mathbf{x}\|_1 \right),$$

where algorithm parameter $\gamma > 0$ trades off data fidelity against sparsity of the solution. This may be interpreted as a Lagrangian form of the estimator

$$\widehat{\mathbf{x}} = \underset{\mathbf{x}\,:\,\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 \leq \epsilon}{\arg\min} \|\mathbf{x}\|_1,$$

which could be justified heuristically by $\|\mathbf{d}\|_2^2 \leq \epsilon$.

Most of the CS literature has considered signal recovery with no noise or with $\|\mathbf{d}\|_2^2 \leq \epsilon$. However, in many practical applications, measurements have to be discretized to a finite number of bits. The effect of such quantization on the performance of CS reconstruction has been studied in [36], [37]. In [38], high-resolution functional scalar quantization theory was used to design quantizers for lasso estimation. Better yet is to change the reconstruction algorithm: In [10]–[12], the authors demonstrate that when $\mathbf{d}$ represents quantization error,

$$\mathbf{d} = Q(\mathbf{A}\mathbf{x}) - \mathbf{A}\mathbf{x},$$

significant improvements can be obtained by replacing the constraint $\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 \leq \epsilon$ by one that uses the partition cells of the quantizers that compose Q.

While convex optimization formulations are prominent in CS, estimation with generic convex program solvers often has excessively high computational cost. Thus, there is significant interest in greedy and iterative methods. The use of belief propagation for CS estimation was first proposed in [39]; however, as explained in Section II-C, belief propagation has high complexity for the estimation of continuous-valued quantities. Lower-complexity approximations to belief propagation were first proposed for CS estimation in [20]. To handle the effects of quantization precisely, in this paper we use the generalization of the technique of [17], [20] developed by Rangan [14].

C. Belief Propagation

Consider the problem of estimating a random vector $\mathbf{x} \in \mathbb{R}^n$ from noisy measurements $\mathbf{y} \in \mathbb{R}^m$, where the noise is described by a measurement channel $p_{\mathbf{y}|\mathbf{z}}$ that acts separably and identically on each entry of the vector $\mathbf{z}$ obtained via (1). Moreover suppose that elements in the vector $\mathbf{x}$ are distributed i.i.d. according to $p_{\mathbf{x}}$. We can construct the following conditional probability distribution over random vector $\mathbf{x}$ given the measurements $\mathbf{y}$:

$$p_{\mathbf{x}|\mathbf{y}}(\mathbf{x} \mid \mathbf{y}) = \frac{1}{Z} \prod_{j=1}^{n} p_{\mathbf{x}}(\mathbf{x}_j) \prod_{i=1}^{m} p_{\mathbf{y}|\mathbf{z}}\left(\mathbf{y}_i \mid \mathbf{z}_i\right), \tag{3}$$

where Z is the normalization constant and $\mathbf{z}_i = (\mathbf{A}\mathbf{x})_i$. In principle, it is possible to estimate each $\mathbf{x}_j$ by marginalizing this distribution.

Belief propagation replaces the computationally intractable direct marginalization of $p_{\mathbf{x}|\mathbf{y}}$ with an iterative algorithm. To apply BP, construct a bipartite factor graph $G = (V, F, E)$ from (3) and pass the following messages along the edges E of the graph:

$$\mu_{i\leftarrow j}^{t+1}(\mathbf{x}_j) \quad \propto \quad p_{\mathbf{x}}(\mathbf{x}_j) \prod_{\ell \neq i} \mu_{\ell\to j}^{t}(\mathbf{x}_j), \tag{4a}$$

$$\mu_{i\to j}^{t}(\mathbf{x}_j) \quad \propto \quad \int p_{\mathbf{y}|\mathbf{z}}(\mathbf{y}_i \mid \mathbf{z}_i) \prod_{k \neq j} \mu_{i\leftarrow k}^{t}(\mathbf{x}_j)\, d\mathbf{x}_{\setminus j}, \tag{4b}$$

where $\propto$ means that the distribution is to be normalized so that it has unit integral and integration is over all the elements of $\mathbf{x}$ except $\mathbf{x}_j$. We refer to messages $\{\mu_{i\leftarrow j}\}_{(i,j)\in E}$ as variable updates and to messages $\{\mu_{i\to j}\}_{(i,j)\in E}$ as measurement updates. BP is initialized by setting $\mu_{i\leftarrow j}^{0}(\mathbf{x}_j) = p_{\mathbf{x}}(\mathbf{x}_j)$.

Earlier works on BP reconstruction have shown that it is asymptotically MSE optimal under certain verifiable conditions. These conditions involve simple single-dimensional recursive equations called *state evolution* (SE), which predicts that BP is optimal when the corresponding SE admits a unique fixed point [16], [22]. However, direct implementation of BP is impractical due to the dense structure of $\mathbf{A}$, which implies that the algorithm must compute the marginal of a high-dimensional distribution at each measurement node; i.e., the integration in (4b) is over many variables. Furthermore, integration must be approximated through some discrete quadrature rule.

BP can be simplified through various Gaussian approximations, including the *relaxed BP* method [15], [16] and *approximate message passing (AMP)* [14], [20]. Recent theoretical work and extensive numerical experiments have demonstrated that, in the case of certain large random measurement matrices, the error performance of both relaxed BP and AMP can also be accurately predicted by SE.

III. Quantized Linear Expansions

This paper focuses on the general quantized measurement abstraction of

$$\mathbf{y} = Q(\mathbf{A}\mathbf{x}), \tag{5}$$

where $\mathbf{x} \in \mathbb{R}^n$ is a signal of interest, $\mathbf{A} \in \mathbb{R}^{m\times n}$ is a linear *mixing matrix*, and $Q : \mathbb{R}^m \to \mathbb{R}^m$ is a scalar quantizer. We will be primarily interested in (per-component) MSE $n^{-1}\mathbb{E}[\|\mathbf{x}-\widehat{\mathbf{x}}\|^2]$ for various estimators $\widehat{\mathbf{x}}$ that depend on $\mathbf{y}$, $\mathbf{A}$, and Q. The cases of $m \geq n$ and $m < n$ are both of interest. We sometimes use $\mathbf{z} = \mathbf{A}\mathbf{x}$ to simplify expressions.

A. Overcomplete Expansions

Let $\mathbf{A} \in \mathbb{R}^{m\times n}$ have rank n. Then $\{\mathbf{a}_i\}_{i=1}^m$ is a *frame* in $\mathbb{R}^n$, where $\mathbf{a}_i^T$ is row i of $\mathbf{A}$. Rank n can occur only with $m \geq n$, so $\mathbf{Ax}$ is called an *overcomplete expansion* of $\mathbf{x}$ and $Q(\mathbf{Ax})$ as in (5) is called a *quantized overcomplete expansion*. In some cases of interest, the frame may be *uniform*, meaning $\|\mathbf{a}_i\| = 1$ for each i, or *tight*, meaning $\mathbf{A}^T\mathbf{A} = c\mathbf{I}_n$ for some scalar c.

Commonly-used *linear reconstruction* forms estimate

$$\widehat{\mathbf{x}} = \mathbf{A}^\dagger \mathbf{y} = \mathbf{A}^\dagger Q(\mathbf{Ax}), \tag{6}$$

where $\mathbf{A}^\dagger = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$ is the pseudoinverse of $\mathbf{A}$. Under several reasonable models, linear reconstruction has MSE inversely proportional to m. For example, suppose the frame is uniform and tight and $\mathbf{x}$ is an unknown deterministic quantity. By modeling scalar quantization $\mathbf{y}_i = q_i(\mathbf{z}_i)$ with an additive noise as

$$\mathbf{y}_i = \mathbf{z}_i + \mathbf{d}_i \tag{7a}$$

where

$$\mathbb{E}[\mathbf{d}_i] = 0, \tag{7b}$$

$$\mathbb{E}[\mathbf{d}_i\mathbf{d}_j] = \sigma_{\mathbf{d}}^2\delta_{ij}, \tag{7c}$$

one can compute the MSE to be $n\sigma_{\mathbf{d}}^2/m$ [40].

Even when the model (7) is accurate [41], the linear reconstruction (6) may be far from optimal. More sophisticated algorithms have focused on enforcing *consistency* of an estimate with the quantized samples. A nonlinear estimate may exploit the boundedness of the sets

$$\mathcal{S}_i(\mathbf{y}_i) = \{\mathbf{x} \in \mathbb{R}^n \,|\, q_i(\mathbf{z}_i) = \mathbf{y}_i\}, \qquad i = 1, 2, \ldots, m,$$

which we call *single-sample consistent sets*. Assuming for now that scalar quantizer q_i is regular and its cells are bounded, the boundary of $\mathcal{S}_i(\mathbf{y}_i)$ is two parallel hyperplanes. The full set of hyperplanes obtained for one index i by varying $\mathbf{y}_i$ over the output levels of q_i is called a hyperplane wave partition [42], as illustrated for a uniform quantizer in Figure 1(a). The set enclosed by two neighboring hyperplanes in a hyperplane wave partition is called a *slab*; one slab is shaded in Figure 1(a). Intersecting $\mathcal{S}_i(\mathbf{y}_i)$ for n distinct indexes specifies an n-dimensional parallelotope as illustrated in Figure 1(b). Using more than n of these single-sample consistent sets restricts $\mathbf{x}$ to a finer partition, as illustrated in Figure 1(c) for $m = 3$.

The intersection

$$\mathcal{S}(\mathbf{y}) = \bigcap_{i=1}^m \mathcal{S}_i(\mathbf{y}_i)$$

is called the *consistent set*. Since each $\mathcal{S}_i(\mathbf{y}_i)$ is convex, one may reach $\mathcal{S}(\mathbf{y})$ asymptotically through a sequence of projections onto $\mathcal{S}_i(\mathbf{y}_i)$ using each infinitely often [1], [2].

In a variety of settings, nonlinear estimates achieve MSE inversely proportional to m^2, which is the best possible dependence on m [42]. The first result of this sort was in [1]. When $\mathbf{A}$ is an oversampled discrete Fourier transform matrix and Q is a uniform quantizer, $\mathbf{z} = \mathbf{Ax}$ represents uniformly quantized samples above Nyquist rate of a periodic bandlimited signal. For this case, it was proven in [1] that any $\widehat{\mathbf{x}} \in \mathcal{S}(\mathbf{y})$ has $O(m^{-2})$ MSE, under a mild assumption on $\|\mathbf{x}\|$. This was extended empirically to arbitrary uniform frames in [3], where it was also shown that consistent estimates can be computed through a linear program. The techniques of alternating projections and linear programming suffer from high computational complexity; yet, since they generally find a corner of the consistent set (rather than the centroid), the MSE performance is suboptimal.

Full consistency is not necessary for optimal MSE dependence on m. It was shown in [4] that $O(m^{-2})$ MSE is guaranteed for a simple algorithm that uses each $\mathcal{S}_i(\mathbf{y}_i)$ only once, recursively, under mild conditions on randomized selection of $\{\mathbf{a}_i\}_{i=1}^m$. These results were strengthened and extended to deterministic frames in [9].

Quantized overcomplete expansions arise naturally in acquisition subsystems such as ADCs, where m/n represents oversampling factor relative to Nyquist rate. In such systems, high oversampling factor may be motivated by a trade-off between MSE and power consumption or manufacturing cost: within certain bounds, faster sampling is cheaper than a higher number of quantization bits per sample [43]. However, high oversampling does not give a good trade-off between MSE and raw number of bits produced by the acquisition system: combining the proportionality of bit rate R to number of samples m with the best-case $\Theta(m^{-2})$ MSE, we obtain $\Theta(R^{-2})$ MSE; this is poor compared to the exponential decrease of MSE with R obtained with scalar quantization of Nyquist-rate samples.

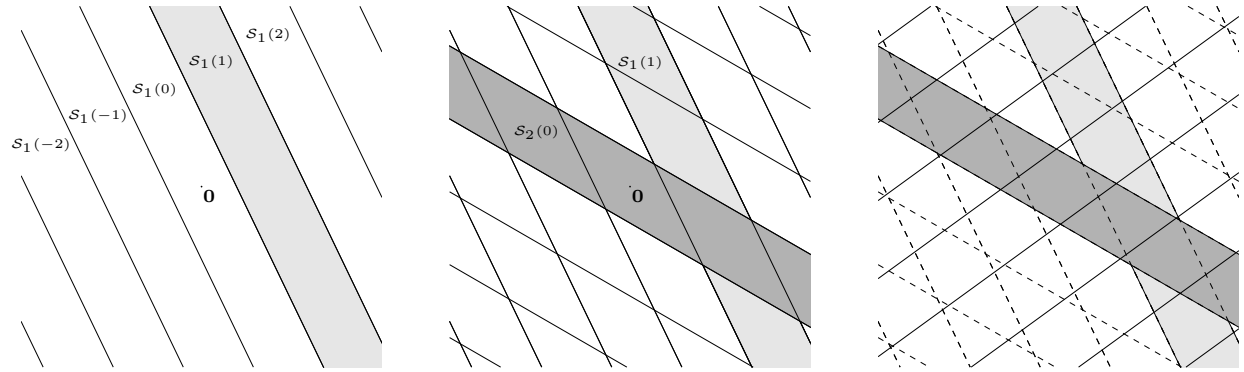
Ordinarily, the bit-rate inefficiency of the raw output is made irrelevant by recoding, at or near Nyquist rate, soon after acquisition or within the ADC. An alternative explored in this paper is to combat this bit-rate inefficiency through the use of non-regular quantization.

B. Non-Regular Quantization

The bit-rate inefficiency of the raw output with regular quantization is easily understood with reference to Figure 1(c). After $\mathbf{y}_1$ and $\mathbf{y}_2$ are fixed, $\mathbf{x}$ is known to lie in the intersection of the shaded strips. Only four values of $\mathbf{y}_3$ are possible (i.e., the solid hyperplane wave breaks $\mathcal{S}_1(1)\cap\mathcal{S}_2(0)$ into four cells), and bits are wasted if this is not exploited in the representation of $\mathbf{y}_3$.

Recall the discussion of generating a non-regular quantizer by using a binning function λ in Section II-A. Binning does not change the boundaries of the single-sample consistent sets, but it makes these sets unions of slabs that may not even be connected. Thus, while binning reduces the quantization rate, in the absence of side information that specifies which slab contains $\mathbf{x}$ (at least with moderately high probability), it increases distortion significantly. The increase in distortion is due to *ambiguity* among slabs. Taking $m > n$ quantized samples together may provide adequate information to disambiguate among slabs, thus removing the distortion penalty.

The key concepts in the use of non-regular quantization are illustrated in Figure 2. Suppose one quantized sample $\mathbf{y}_1$ specifies a single-sample consistent set $\mathcal{S}_1(\mathbf{y}_1)$ composed of two slabs, such as the shaded region in Figure 2(a). A second quantized sample $\mathbf{y}_2$ will not disambiguate between



(a) (b) (c)

Fig. 1: Visualizing the information present in a quantized overcomplete expansion of $\mathbf{x} \in \mathbb{R}^2$ when each q_i is a regular quantizer. (a) A single hyperplane wave partition with one single-sample consistent set shaded. (b) Partition boundaries from two hyperplane waves; $\mathbf{x}$ is specified to the intersection of two single-sample consistent sets, which is a bounded convex cell. (c) Partition from part (b) in dashed lines with a third hyperplane wave added in solid lines.

the two slabs. In the example shown in Figure 2(b), $\mathcal{S}_2(\mathbf{y}_2)$ is composed of two slabs, and $\mathcal{S}_1(\mathbf{y}_1) \cap \mathcal{S}_2(\mathbf{y}_2)$ is the union of four connected sets. A third quantized sample $\mathbf{y}_3$ may now completely disambiguate; the particular example of $\mathcal{S}_3(\mathbf{y}_3)$ shown in Figure 2(c) makes $\mathcal{S} = \mathcal{S}_1(\mathbf{y}_1) \cap \mathcal{S}_2(\mathbf{y}_2) \cap \mathcal{S}_3(\mathbf{y}_3)$ a single convex set.

When the quantized samples together completely disambiguate the slabs as in the example, the rate reduction from binning comes with no increase in distortion. The price to pay comes in complexity of estimation.

The use of binned quantization of linear expansions was introduced in [44], where the only reconstruction method proposed is intractable in high dimensions because it is combinatorial over the binning functions. Specifically, using the notation from Section II-A, let the quantizer forming $\mathbf{y}_i$ be defined by $(\alpha_i, \beta_i, \lambda_i)$. Then $\lambda_i^{-1}(\beta_i^{-1}(\mathbf{y}_i))$ will be a set of possible values of $\alpha_i(\mathbf{z}_i)$ specified by $\mathbf{y}_i$. One can try every combination, i.e., element of

$$\lambda_1^{-1}(\beta_1^{-1}(\mathbf{y}_1)) \times \lambda_2^{-1}(\beta_2^{-1}(\mathbf{y}_2)) \times \cdots \times \lambda_m^{-1}(\beta_m^{-1}(\mathbf{y}_m)), \quad (8)$$

to seek a consistent estimate. If the binning is effective, most combinations yield an empty consistent set; if the slabs are disambiguated, exactly one combination yields a non-empty set, which is then the consistent set $\mathcal{S}$. This technique has complexity exponential in m (assuming non-trivial binning). The recent manuscript [45] provides bounds on reconstruction error for consistent estimation with binned quantization; it does not address reconstruction.

This paper provides a tractable and effective method for reconstruction from a quantized linear expansion with non-regular quantizers. To the best of our knowledge, this is the first such method.

C. Undercomplete Expansions

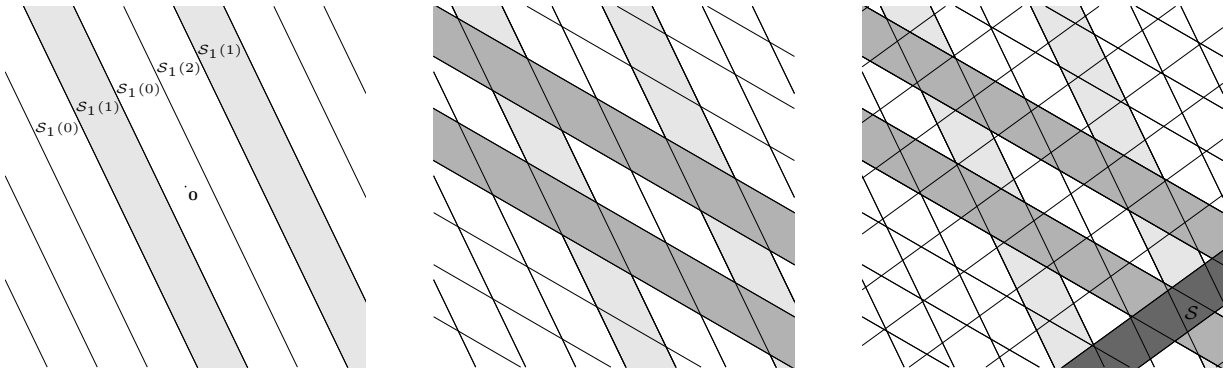
Maintaining the quantized measurement model (5), let us turn to the case of $m < n$. We now call $Q(\mathbf{A}\mathbf{x})$ a *quantized undercomplete expansion* of $\mathbf{x}$.

Since the rank of $\mathbf{A}$ is less than n, $\mathbf{A}$ is a many-to-one mapping. Thus, even without quantization, one cannot recover $\mathbf{x}$ from $\mathbf{A}\mathbf{x}$. Rather, $\mathbf{A}\mathbf{x}$ specifies a proper subspace of $\mathbb{R}^n$ containing $\mathbf{x}$; when $\mathbf{A}$ is in general position, the subspace is of dimension $n - m$. Quantization increases the ambiguity in the value of $\mathbf{x}$, yielding consist sets similar to those depicted in Figures 1(a) and 2(a). However, as described in Section II-B, knowledge that $\mathbf{x}$ is sparse or approximately sparse could be exploited to enable accurate estimation of $\mathbf{x}$ from $Q(\mathbf{A}\mathbf{x})$.

For ease of explanation, consider only the case where $\mathbf{x}$ is known to be k-sparse with $k < m$. Let $\mathcal{J} \subset \{1, 2, \ldots, n\}$ be the support (sparsity pattern) of $\mathbf{x}$, with $|\mathcal{J}| = k$. The product $\mathbf{A}\mathbf{x}$ is equal to $\mathbf{A}_{\mathcal{J}}\mathbf{x}_{\mathcal{J}}$, where $\mathbf{x}_{\mathcal{J}}$ denotes the restriction of the domain of $\mathbf{x}$ to $\mathcal{J}$ and $\mathbf{A}_{\mathcal{J}}$ is the $m \times k$ submatrix of $\mathbf{A}$ containing the $\mathcal{J}$-indexed columns. Assuming $\mathbf{A}_{\mathcal{J}}$ has rank k (i.e., full rank), $Q(\mathbf{A}\mathbf{x}) = Q(\mathbf{A}_{\mathcal{J}}\mathbf{x}_{\mathcal{J}})$ is a quantized *overcomplete* expansion of $\mathbf{x}_{\mathcal{J}}$. All discussion of estimation of $\mathbf{x}_{\mathcal{J}}$ from the previous subsections thus applies, assuming $\mathcal{J}$ is known.

The key remaining issue is that $Q(\mathbf{A}\mathbf{x})$ may or may not provide enough information to infer $\mathcal{J}$. In an overcomplete representation, most vectors of quantizer outputs cannot occur; this redundancy was used to enable binning in Figure 2, and it can be used to show that certain subsets $\mathcal{J}$ are inconsistent with the sparse signal model. In principle, one may enumerate the sets $\mathcal{J}$ of size k and apply a consistent reconstruction method for each $\mathcal{J}$. If only one candidate $\mathcal{J}$ yields a non-empty consistent set, then $\mathcal{J}$ is determined. This is intractable except for small problem sizes because there are $\binom{n}{k}$ candidates for $\mathcal{J}$.

The key concepts are illustrated in Figure 3. To have an interpretable diagram with $k < m < n$, we let $(k, m, n) = (1, 2, 3)$ and draw the space of unquantized measurements $\mathbf{z} \in \mathbb{R}^2$. (This contrasts with Figures 1 and 2 where the space of $\mathbf{x} \in \mathbb{R}^2$ is drawn.) The vector $\mathbf{x}$ has one of $\binom{n}{k} = \binom{3}{1} = 3$ possible supports $\mathcal{J}$. Thus, $\mathbf{z}$ lies in one of 3 subspaces of dimension 1, which are depicted by the angled solid lines. Scalar quantization of $\mathbf{z}$ corresponds to separable partitioning



(a) (b) (c)

Fig. 2: Visualizing the information present in a quantized overcomplete expansion of $\mathbf{x} \in \mathbb{R}^2$ when using non-regular (binned) quantizers. (a) A single hyperplane wave partition with one single-sample consistent set shaded. Note that binning makes the shaded set not connected. (b) Partition boundaries from two hyperplane waves; $\mathbf{x}$ is specified to the intersection of two single-sample consistent sets, which is now the union of four convex cells. (c) A third sample now specifies $\mathbf{x}$ to within a consistent set $\mathcal{S}$ that is convex.

of $\mathbb{R}^2$ with cell boundaries aligned with coordinate axes, as shown with lighter solid lines.

Only one quantized measurement $\mathbf{y}_1$ is not adequate to specify $\mathcal{J}$, as shown in Figure 3(a) by the fact that a single shaded cell intersects all the subspaces.[1] Two quantized measurements together will usually specify $\mathcal{J}$, as shown in Figure 3(b) by the fact that only one subspace intersects the specified square cell; for fixed scalar quantizers, ambiguity becomes less likely as k decreases, n increases, m increases, or $\|x\|$ increases. Figure 3(c) shows a case where non-regular (binned) quantization still allows unambiguous determination of $\mathcal{J}$.

The naïve reconstruction method implied by Figure 3(c) is to search combinatorially over both $\mathcal{J}$ and the combinations in (8); this is extremely complex. While the use of binning for quantized undercomplete expansions of sparse signals has appeared in the literature, first in [44] and later in [45], to the best of our knowledge this paper is the first to provide a tractable and effective reconstruction method.

IV. Estimation from Quantized Samples: Bayesian Formulation

We now specify more explicitly the class of problems for which we derive new estimation algorithms. Generalizing (5), let

$$\mathbf{y} = Q(\mathbf{z} + \mathbf{w}) \qquad \text{where} \qquad \mathbf{z} = \mathbf{A}\mathbf{x}, \tag{9}$$

as depicted in Figure 4. The input vector $\mathbf{x} \in \mathbb{R}^n$ is random with i.i.d. entries with prior p.d.f. $p_{\mathbf{x}}$. The linear mixing matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ is random with i.i.d. entries $a_{ij} \sim \mathcal{N}(0, 1/m)$. The (pre-quantization) additive noise $\mathbf{w} \in \mathbb{R}^m$ is random with i.i.d. entries $w_i \sim \mathcal{N}(0, \sigma^2)$. The quantizer Q is a scalar quantizer, and each of its component quantizers q_i is identical and has K output levels.

The estimator $\widehat{\mathbf{x}}$ is a function of $\mathbf{A}$, $\mathbf{y}$, Q, and σ^2. We wish to minimize the MSE $n^{-1}\mathbb{E}[\|\mathbf{x} - \widehat{\mathbf{x}}\|^2]$.

[1] Intersections with two subspaces are shown within the range of the diagram.

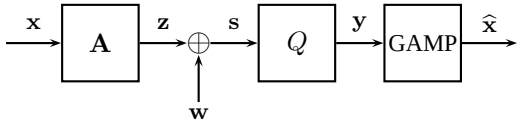


Fig. 4: Quantized linear measurement model for which GAMP estimator is derived. Vector $\mathbf{x} \in \mathbb{R}^n$ with an i.i.d. prior is estimated from scalar quantized measurements $\mathbf{y} \in \mathbb{R}^m$. The quantizer input $\mathbf{s}$ is the sum of $\mathbf{z} = \mathbf{A}\mathbf{x} \in \mathbb{R}^m$ and an i.i.d. Gaussian noise vector $\mathbf{w}$. Including noise variance σ^2 in the model clarifies certain derivations; setting the noise variance to zero recovers acquisition model (5).

Our primary interest is in the case of $\sigma^2 = 0$, but allowing a nontrivial distribution for $\mathbf{w}$ is not only more general but also makes the derivations more clear.

V. Generalized Approximate Message Passing for a Quantizer Output Channel

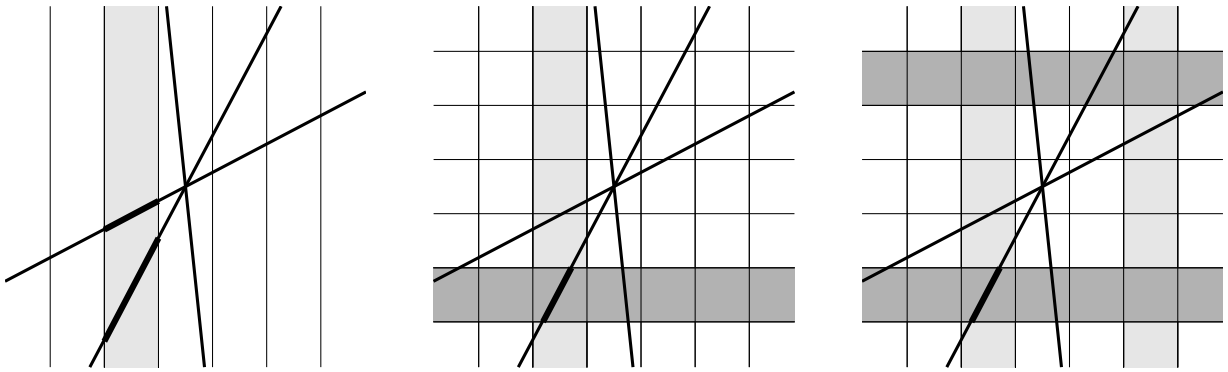
The acquisition model (9) is suitable for GAMP estimation under the conditions in [14] after one simple observation: the mapping from $\mathbf{z}$ to $\mathbf{y}$ is a separable probabilistic mapping with identical marginals. Specifically, quantized measurement $\mathbf{y}_i$ indicates $\mathbf{s}_i \in q_i^{-1}(\mathbf{y}_i)$, so each component *output channel* can be characterized as

$$p_{\mathbf{y}|\mathbf{z}}(y \mid z) = \int_{q_i^{-1}(y)} \phi\left(t\,;\, z,\, \sigma^2\right)\, dt,$$

where ϕ is the Gaussian function

$$\phi\left(t\,;\, a,\, b\right) = \frac{1}{\sqrt{2\pi b}} \exp\left(-\frac{(t-a)^2}{2b}\right).$$

GAMP can be derived by approximating the updates in (4) by two scalar parameters each and introducing some first-order approximations, as discussed in [14]. Then given the estimation functions F_{in}, $\mathcal{E}_{\text{in}}$, D_1, and D_2 described below, for each iteration $t = 0,\, 1,\, 2,\, \ldots$, the GAMP algorithm produces



(a) (b) (c)

Fig. 3: Visualizing the information present in a quantized undercomplete expansion $Q(\mathbf{Ax})$ of a 1-sparse signal $\mathbf{x} \in \mathbb{R}^3$ when $\mathbf{Ax} \in \mathbb{R}^2$. The depicted 2-dimensional plane represents the vector of measurements $\mathbf{z} = \mathbf{Ax}$. Since $\mathbf{x}$ is 1-sparse, the measurement lies in a union of 1-dimensional subspaces (the angled solid lines); since $\mathbf{x}$ is 3 dimensional, there are three such subspaces. (a) Scalar quantization of $\mathbf{z}_1$ divides the plane of possible values for $\mathbf{z}$ into vertical strips. One particular value of $\mathbf{y}_1 = q_1(\mathbf{z}_1)$ does not specify which entry of $\mathbf{x}$ is nonzero since the shaded strip intersects all the angled solid lines. For each possible support, the value of the nonzero entry is specified to an interval. (b) Scalar quantization of both components of $\mathbf{z}$ specifies $\mathbf{z}$ to a rectangular cell. In most cases, including the one highlighted, the quantized values specify which entry of $\mathbf{x}$ is nonzero because only one angled solid line intersects the cell. The value of the nonzero entry is specified to an interval. (c) In many cases, including the one highlighted, the quantizers can be non-regular (binned) and yet still uniquely specify which entry of $\mathbf{x}$ is nonzero.

estimates $\widehat{\mathbf{x}}^t$ of the true signal $\mathbf{x}$ according to the following rules:

$$\widehat{\mathbf{x}}^{t+1} \equiv F_{\text{in}}\left(\widehat{\mathbf{x}}^t + \frac{\mathbf{A}^T\mathbf{u}^t}{\left(\mathbf{A}^T\right)^2\boldsymbol{\tau}^t}, \frac{1}{\left(\mathbf{A}^T\right)^2\boldsymbol{\tau}^t}\right), \tag{10a}$$

$$\widehat{\boldsymbol{\tau}}^{t+1} \equiv \mathcal{E}_{\text{in}}\left(\widehat{\mathbf{x}}^t + \frac{\mathbf{A}^T\mathbf{u}^t}{\left(\mathbf{A}^T\right)^2\boldsymbol{\tau}^t}, \frac{1}{\left(\mathbf{A}^T\right)^2\boldsymbol{\tau}^t}\right), \tag{10b}$$

$$\mathbf{u}^t \equiv D_1\left(\mathbf{y}, \mathbf{A}\widehat{\mathbf{x}}^t - \mathbf{u}^{t-1}\mathbf{A}^2\widehat{\boldsymbol{\tau}}^t, \mathbf{A}^2\widehat{\boldsymbol{\tau}}^t + \sigma^2\mathbf{I}_n\right), \tag{10c}$$

$$\boldsymbol{\tau}^t \equiv D_2\left(\mathbf{y}, \mathbf{A}\widehat{\mathbf{x}}^t - \mathbf{u}^{t-1}\mathbf{A}^2\widehat{\boldsymbol{\tau}}^t, \mathbf{A}^2\widehat{\boldsymbol{\tau}}^t + \sigma^2\mathbf{I}_n\right). \tag{10d}$$

Note that in (10) the notation $\mathbf{A}^2$ denotes the element-wise product of a matrix with itself, i.e. $(\mathbf{A}^2)_{ij} = (\mathbf{A}_{ij})^2$. The estimation functions F_{in}, $\mathcal{E}_{\text{in}}$, D_1, and D_2 described below are applied to their inputs component-by-component.

We refer to messages $\{\hat{x}_j, \hat{\tau}_j\}_{j\in V}$ as variable updates and to messages $\{u_i, \tau_i\}_{i\in F}$ as measurement updates. The algorithm is initialized by setting $\hat{x}_j^0 = \hat{x}_{\text{init}}$, $\hat{\tau}_j^0 = \hat{\tau}_{\text{init}}$, and $u_i^{-1} = 0$, where $\hat{x}_{\text{init}}$ and $\hat{\tau}_{\text{init}}$ are the mean and variance of the prior $p_{\mathbf{x}}$. The nonlinear functions F_{in} and $\mathcal{E}_{\text{in}}$ are the conditional mean and variance

$$F_{\text{in}}(q,\nu) \equiv \mathbb{E}[x \mid q],$$
$$\mathcal{E}_{\text{in}}(q,\nu) \equiv \text{var}(x \mid q),$$

where $q = x + v$ with $x \sim p_{\mathbf{x}}$ and $v \sim \mathcal{N}(0,\nu)$. Note that these functions can easily be evaluated numerically for any given values of q and σ^2. Similarly, the functions D_1 and D_2 can be computed via

$$D_1(y,\hat{z},\nu) \equiv \frac{1}{\nu}\left(F_{\text{out}}(y,\hat{z},\nu) - \hat{z}\right), \tag{11a}$$

$$D_2(y,\hat{z},\nu) \equiv \frac{1}{\nu}\left(1 - \frac{\mathcal{E}_{\text{out}}(y,\hat{z},\nu)}{\nu}\right), \tag{11b}$$

where the functions F_{out} and $\mathcal{E}_{\text{out}}$ are the conditional mean and variance

$$F_{\text{out}}(y,\hat{z},\nu) \equiv \mathbb{E}\left[z \mid z \in q_i^{-1}(y)\right], \tag{12a}$$

$$\mathcal{E}_{\text{out}}(y,\hat{z},\nu) \equiv \text{var}\left(z \mid z \in q_i^{-1}(y)\right), \tag{12b}$$

of the random variable $z \sim \mathcal{N}(\hat{z},\nu)$. These functions admit closed-form expressions in terms of $\text{erf}(z) = \frac{2}{\sqrt{\pi}}\int_0^z e^{-t^2}\,dt$.

VI. State Evolution for GAMP

The equations (10) are easy to implement, however they provide us no insight into the performance of the algorithm. The goal of SE equations is to describe the asymptotic behavior of GAMP under large measurement matrices.

The SE for our setting in Figure 4 is given by the recursion

$$\bar{\tau}_{t+1} = \bar{\mathcal{E}}_{\text{in}}\left(\frac{1}{\bar{D}_2\left(\beta\bar{\tau}_t, \sigma^2\right)}\right), \tag{13}$$

where $t \geq 0$ is the iteration number, $\beta = n/m$ is a fixed number denoting the measurement ratio, and σ^2 is the variance of the additive white Gaussian noise (AWGN), which is also fixed. We initialize the recursion by setting $\bar{\tau}_0 = \hat{\tau}_{\text{init}}$, where τ_{init} is the variance of $\mathbf{x}_j$ according to the prior $p_{\mathbf{x}}$. We define the function $\bar{\mathcal{E}}_{\text{in}}$ as

$$\bar{\mathcal{E}}_{\text{in}}(\nu) = \mathbb{E}\left[\mathcal{E}_{\text{in}}(q,\nu)\right], \tag{14}$$

where the expectation is taken over the scalar random variable $q = x + v$, with $x \sim p_{\mathbf{x}}$ and $v \sim \mathcal{N}(0,\nu)$. Similarly, the function $\bar{D}_2$ is defined as

$$\bar{D}_2\left(\nu,\sigma^2\right) = \mathbb{E}\left[D_2\left(y,\hat{z},\nu+\sigma^2\right)\right], \tag{15}$$

where D_2 is given by (11b) and the expectation is taken over $p_{\mathbf{y}|\mathbf{z}}$ and $(z, \hat{z}) \sim \mathcal{N}(0, P_z(\nu))$, with the covariance matrix

$$P_z(\nu) = \begin{pmatrix} \beta\hat{\tau}_{\text{init}} & \beta\hat{\tau}_{\text{init}} - \nu \\ \beta\hat{\tau}_{\text{init}} - \nu & \beta\hat{\tau}_{\text{init}} - \nu \end{pmatrix}. \tag{16}$$

One of the main results of [14] was to demonstrate the convergence of the error performance of the GAMP algorithm to the SE equations. In the asymptotic analysis, $m, n \to \infty$ with $n/m \to \beta$.

Another important result regarding SE recursion in (13) is that it admits at least one fixed point. It has been shown that as $t \to \infty$ the recursion decreases monotonically to its largest fixed point and if the SE admits a unique fixed point, then GAMP is asymptotically mean-square optimal [14].

VII. Quantizer Optimization

Ordinarily, quantizer designs depend on the distribution of the quantizer input, with an implicit aim of minimizing the MSE between the quantizer input and output. Often, only uniform quantizers are considered, in which case the "design" is to choose the loading factor of the quantizer. When quantized data is used as an input to a nonlinear function, overall system performance may be improved by adjusting the quantizer designs appropriately [21]. In the present setting, conventional quantizer design minimizes $m^{-1}\mathbb{E}[\|\mathbf{z} - Q(\mathbf{z})\|^2]$, but minimizing $n^{-1}\mathbb{E}[\|\mathbf{x} - \widehat{\mathbf{x}}\|^2]$ is desired instead.

The SE description of GAMP performance facilitates the desired optimization. By modeling the quantizer as part of the channel and working out the resulting equations for GAMP and SE, we can make use of the convergence results to recast our optimization problem to

$$Q^{\text{SE}} = \arg\min_{Q} \left\{ \lim_{t\to\infty} \bar{\tau}_t \right\}, \tag{17}$$

where minimization is done over all K-level regular scalar quantizers. In practice, about 10 to 20 iterations are sufficient to reach the fixed point of $\bar{\tau}_t$. Then by applying convergence results from [14], we know that the asymptotic error performance of the optimal quantizer for GAMP will be identical to that of Q^{SE}. It is important to note that the SE recursion behaves well under quantizer optimization. This is due to the fact that SE is independent of actual output levels and small changes in the quantizer boundaries result in only minor change in the recursion (see (12b)). Although closed-form expressions for the derivatives of $\bar{\tau}_t$ for large t's are difficult to obtain, we can approximate them by using finite difference methods. Finally, the recursion itself is fast to evaluate, which makes the scheme in (17) practically realizable under standard optimization methods like sequential quadratic programming (SQP).

VIII. Experimental Results

A. Overcomplete Expansions

We first consider overcomplete expansion of $\mathbf{x}$ as discussed in Section III-A. We generate the signal $\mathbf{x}$ with i.i.d. elements from the standard Gaussian distribution $\mathbf{x}_j \sim \mathcal{N}(0, 1)$. We form the measurement matrix $\mathbf{A}$ from i.i.d. zero-mean Gaussian random variables. To concentrate on the degradation due to quantization we assume noiseless measurement model (5); i.e., $\sigma^2 = 0$ in (9).

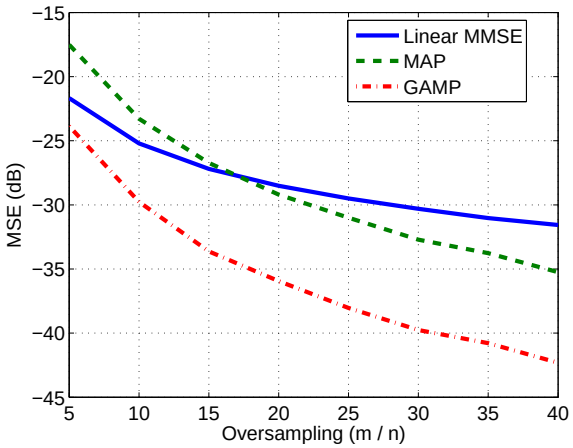


Fig. 5: Performance comparison for oversampled observation of a jointly Gaussian signal vector (no sparsity). GAMP outperforms linear MMSE and MAP estimators.

Figure 5 presents squared-error performance of three estimation algorithm while varying the oversampling ratio m/n and holding $n = 100$. To generate the plot we considered estimation from measurements discretized by a 16-level regular uniform quantizer. For each value of m/n, 100 random realizations of the problem were generated; the curves show the median squared error over these 100 Monte Carlo trials. For comparison to GAMP, we also plot the performance for two other common reconstruction methods: linear MMSE and maximum a posteriori probability (MAP). The MAP estimator was implemented using quadratic programming (QP) and obtains the minimum Euclidean-norm estimate consistent with the quantized samples.

We see that GAMP offers significantly better performance with more than 5 dB improvement for many values of m/n. MAP provides poor performance compared to GAMP because it finds a corner of the consistent set, which is suboptimal as compared to the centroid of the consistent set.

B. Compressive Sensing with Quantized Measurements

We now would like to estimate a sparse signal $\mathbf{x}$ from $m < n$ random measurements. We assume that the signal $\mathbf{x}$ is generated with i.i.d. elements from the Gauss–Bernoulli distribution

$$\mathbf{x}_j \sim \begin{cases} \mathcal{N}(0, 1/\rho), & \text{with probability } \rho; \\ 0, & \text{with probability } 1-\rho, \end{cases} \tag{18}$$

where ρ is the sparsity ratio that represents the average fraction of nonzero components of $\mathbf{x}$. In the following experiments we assume $\rho = 0.1$. Similarly to overcomplete case, we form the measurement matrix $\mathbf{A}$ from i.i.d. Gaussian random variables, i.e., $A_{ij} \sim \mathcal{N}(0, 1/m)$; and we assume no additive noise ($\sigma^2 = 0$ in (9)).

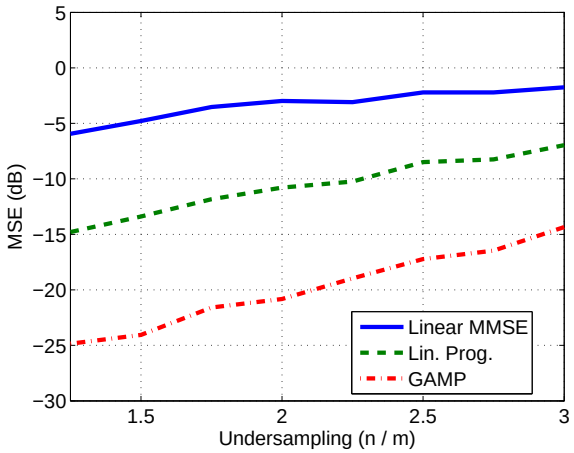


Fig. 6: Performance comparison of GAMP with other sparse estimation methods.

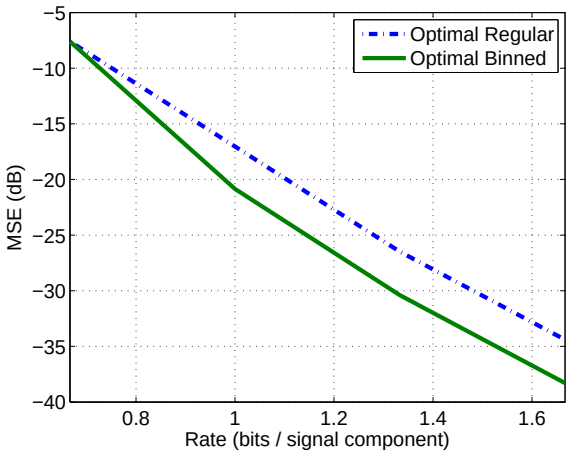


Fig. 7: Performance comparison of GAMP with regular and binned quantizers under sparse Gauss-Bernoulli prior.

Figure 6 compares squared-error performance of GAMP with two other sparse estimation methods. We obtain the curves by varying the undersampling ratio $\beta = n/m$ and holding $m = 100$. We perform estimation from measurements obtained from a 16-level regular uniform quantizer. The figure plots the median of squared error values from 100 Monte Carlo trials for each value of β. The top curve (worst performance) is for linear MMSE estimation; and middle curve labeled is for the basis pursuit estimator

$$\widehat{\mathbf{x}} = \underset{\mathbf{x}\,:\,\mathbf{y}=Q(\mathbf{A}\mathbf{x})}{\arg\min} \|\mathbf{x}\|_1,$$

which can be cast and solved as a linear program (LP). Again GAMP offers substantial improvement over the other methods.

C. Non-Regular Quantization

We now repeat the case of the undersampled sparse signal by again using $\mathbf{x}$ with an i.i.d. Gauss–Bernoulli distribution with $\rho = 0.1$. To study the effect of non-regular quantization, we introduce a binning function

$$\lambda : \{1, 2, \ldots, N\} \to \{1, 2, \ldots, N/2\}$$

that reduces the rate of the quantizer by 1 bit per sample (halving the number of output levels) by performing reduction modulo 2. Hence input indexes i and $i + N/2$ get mapped to the same output index i.

Figure 7 plots the MSE performance of GAMP under regular and non-regular quantization. Both quantizers were optimized using (17). We obtain the plot by varying the number of bits per component of $\mathbf{x}$ and holding $\beta = 3$. We see that, in comparison to regular quantizers, binned quantizers with GAMP estimation achieve much lower distortions for the same rates. This indicates that binning can be an effective strategy to favorably shift rate–distortion performance of the estimation.

IX. Conclusions

We present generalized approximate message passing as an effective and efficient algorithm for estimation from quantized linear measurements. The problem formulation is Bayesian, with an i.i.d. prior over the components of the signal of interest $\mathbf{x}$; the prior may or may not induce sparsity of $\mathbf{x}$. Also, the number of measurements may be more or less than the dimension of $\mathbf{x}$, and the quantizers applied to the linear measurements may be regular or not. Experiments show significant performance improvement over traditional reconstruction schemes, some of which have higher computational complexity. Matlab code for experiments with GAMP is available online [46].

References

[1] N. T. Thao and M. Vetterli, "Reduction of the MSE in R-times oversampled A/D conversion from $O(1/R)$ to $O(1/R^2)$," *IEEE Trans. Signal Process.*, vol. 42, no. 1, pp. 200–203, Jan. 1994.

[2] ——, "Deterministic analysis of oversampled A/D conversion and decoding improvement based on consistent estimates," *IEEE Trans. Signal Process.*, vol. 42, no. 3, pp. 519–531, Mar. 1994.

[3] V. K. Goyal, M. Vetterli, and N. T. Thao, "Quantized overcomplete expansions in $\mathbb{R}^N$: Analysis, synthesis, and algorithms," *IEEE Trans. Inform. Theory*, vol. 44, no. 1, pp. 16–31, Jan. 1998.

[4] S. Rangan and V. K. Goyal, "Recursive consistent estimation with bounded noise," *IEEE Trans. Inform. Theory*, vol. 47, no. 1, pp. 457–464, Jan. 2001.

[5] Z. Cvetković, "Resilience properties of redundant expansions under additive noise and quantization," *IEEE Trans. Inform. Theory*, vol. 49, no. 3, pp. 644–656, Mar. 2003.

[6] J. J. Benedetto, A. M. Powell, and Ö. Yilmaz, "Sigma–Delta ($\Sigma\Delta$) quantization and finite frames," *IEEE Trans. Inform. Theory*, vol. 52, no. 5, pp. 1990–2005, May 2006.

[7] B. G. Bodmann and V. I. Paulsen, "Frame paths and error bounds for sigma-delta quantization," *Appl. Comput. Harm. Anal.*, vol. 22, no. 2, pp. 176–197, Mar. 2007.

[8] B. G. Bodmann and S. P. Lipshitz, "Randomly dithered quantization and sigma–delta noise shaping for finite frames," *Appl. Comput. Harm. Anal.*, vol. 25, no. 3, pp. 367–380, Nov. 2008.

[9] A. M. Powell, "Mean squared error bounds for the Rangan–Goyal soft thresholding algorithm," *Appl. Comput. Harm. Anal.*, vol. 29, no. 3, pp. 251–271, Nov. 2010.

[10] A. Zymnis, S. Boyd, and E. Candès, "Compressed sensing with quantized measurements," *IEEE Signal Process. Lett.*, vol. 17, no. 2, pp. 149–152, Feb. 2010.

[11] L. Jacques, D. K. Hammond, and J. M. Fadili, "Dequantizing compressed sensing: When oversampling and non-Gaussian constraints combine," *IEEE Trans. Inform. Theory*, vol. 57, no. 1, pp. 559–571, Jan. 2011.

[12] J. N. Laska, P. T. Boufounos, M. A. Davenport, and R. G. Baraniuk, "Democracy in action: Quantization, saturation, and compressive sensing," *Appl. Comput. Harm. Anal.*, vol. 31, 2011.
[13] A. Gersho, "Principles of quantization," *IEEE Trans. Circuits Syst.*, vol. CAS-25, no. 7, pp. 427–436, Jul. 1978.
[14] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," arXiv:1010.5141v1 [cs.IT]., Oct. 2010.
[15] ——, "Estimation with random linear mixing, belief propagation and compressed sensing," in *Proc. Conf. on Inform. Sci. & Sys.*, Princeton, NJ, Mar. 2010, pp. 1–6.
[16] D. Guo and C.-C. Wang, "Asymptotic mean-square optimality of belief propagation for sparse linear systems," in *Proc. IEEE Inform. Theory Workshop*, Chengdu, China, Oct. 2006, pp. 194–198.
[17] M. Bayati and A. Montanari, "The dynamics of message passing on dense graphs, with applications to compressed sensing," *IEEE Trans. Inform. Theory*, vol. 57, no. 2, pp. 764–785, Feb. 2011.
[18] J. Boutros and G. Caire, "Iterative multiuser joint decoding: Unified framework and asymptotic analysis," *IEEE Trans. Inform. Theory*, vol. 48, no. 7, pp. 1772–1793, Jul. 2002.
[19] T. Tanaka and M. Okada, "Approximate belief propagation, density evolution, and neurodynamics for CDMA multiuser detection," *IEEE Trans. Inform. Theory*, vol. 51, no. 2, pp. 700–706, Feb. 2005.
[20] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proc. Nat. Acad. Sci.*, vol. 106, no. 45, pp. 18 914–18 919, Nov. 2009.
[21] V. Misra, V. K. Goyal, and L. R. Varshney, "Distributed scalar quantization for computing: High-resolution analysis and extensions," *IEEE Trans. Inform. Theory*, vol. 57, no. 8, Aug. 2011, to appear.
[22] D. Guo and C.-C. Wang, "Random sparse linear systems observed via arbitrary channels: A decoupling principle," in *Proc. IEEE Int. Symp. Inform. Theory*, Nice, France, Jun. 2007, pp. 946–950.
[23] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2325–2383, Oct. 1998.
[24] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.
[25] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, CA: Morgan Kaufmann Publ., 1988.
[26] T. Richardson and R. Urbanke, "The capacity of low-density parity check codes under message-passing decoding," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 599–618, Feb. 2001.
[27] J. S. Yedidia, W. T. Freeman, and Y. Weiss, "Understanding belief propagation and its generalizations," in *Exploring Artificial Intelligence in the New Millennium*. San Francisco, CA: Morgan Kaufmann Publishers, 2003, pp. 239–269.
[28] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inform. Theory*, vol. IT-22, no. 1, pp. 1–10, Jan. 1976.
[29] V. K. Goyal, "Multiple description coding: Compression meets the network," *IEEE Signal Process. Mag.*, vol. 18, no. 5, pp. 74–93, Sep. 2001.
[30] Z. Liu, S. Cheng, A. D. Liveris, and Z. Xiong, "Slepian–Wolf coded nested lattice quantization for Wyner–Ziv coding: High-rate performance analysis and code design," *IEEE Trans. Inform. Theory*, vol. 52, no. 10, pp. 4358–4379, Oct. 2006.
[31] V. A. Vaishampayan, "Design of multiple description scalar quantizers," *IEEE Trans. Inform. Theory*, vol. 39, no. 3, pp. 821–834, May 1993.
[32] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.
[33] E. J. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Inform. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.
[34] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
[35] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Royal Stat. Soc., Ser. B*, vol. 58, no. 1, pp. 267–288, 1996.
[36] E. J. Candès and J. Romberg, "Encoding the ℓ_p ball from limited measurements," in *Proc. IEEE Data Compression Conf.*, Snowbird, UT, Mar. 2006, pp. 33–42.
[37] V. K. Goyal, A. K. Fletcher, and S. Rangan, "Compressive sampling and lossy compression," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 48–56, Mar. 2008.
[38] J. Z. Sun and V. K. Goyal, "Optimal quantization of random measurements in compressed sensing," in *Proc. IEEE Int. Symp. Inform. Theory*, Seoul, Korea, Jun.–Jul. 2009, pp. 6–10.
[39] D. Baron, S. Sarvotham, and R. G. Baraniuk, "Bayesian compressive sensing via belief propagation," *IEEE Trans. Signal Process.*, vol. 58, no. 1, pp. 269–280, Jan. 2010.
[40] V. K. Goyal, J. Kovačević, and J. A. Kelner, "Quantized frame expansions with erasures," *Appl. Comput. Harm. Anal.*, vol. 10, no. 3, pp. 203–233, May 2001.
[41] H. Viswanathan and R. Zamir, "On the whiteness of high-resolution quantization errors," *IEEE Trans. Inform. Theory*, vol. 47, no. 5, pp. 2029–2038, Jul. 2001.
[42] N. T. Thao and M. Vetterli, "Lower bound on the mean-squared error in oversampled quantization of periodic signals using vector quantization analysis," *IEEE Trans. Inform. Theory*, vol. 42, no. 2, pp. 469–479, Mar. 1996.
[43] R. H. Walden, "Analog-to-digital converter survey and analysis," *IEEE J. Sel. Areas Comm.*, vol. 17, no. 4, pp. 539–550, Apr. 1999.
[44] R. J. Pai, "Nonadaptive lossy encoding of sparse signals," Master's thesis, Massachusetts Inst. of Tech., Cambridge, MA, Aug. 2006.
[45] P. T. Boufounos, "Universal rate-efficient scalar quantization," arXiv:1009.3145v2 [cs.IT]., Oct. 2010.
[46] S. Rangan *et al.*, "Generalized approximate message passing," SourceForge.net project gampmatlab, available on-line at `http://gampmatlab.sourceforge.net/`.