Compressed Sensing

Supplemental Material for Graphical Models and Inference
Henry D. Pfister

November 26th, 2014

1 Introduction

Compressed sensing (CS) is an area of signal processing and statistics that emerged in the late 1990's and then exploded in the mid-2000's [1, 2, 3, 4, 5]. The main idea is that many signals can be reconstructed and/or processed using many fewer measurements than predicted by a naive application of Nyquist's theorem. CS originated with the observation that many signal processing systems first sample a large amount of data, then perform a linear (e.g., wavelet) transform, and finally throw all the small coefficients away. If the locations of the large-magnitude transform coefficients were known in advance, then one could just sample those values directly and reduce the complexity enormously. CS is a research area that emerged when people realized that a random transform could be used (with a small penalty) to achieve the same result without prior knowledge about the locations of important coefficients. CS can be broken into two phases: sampling and reconstruction. In the sampling phase, the signal of interest (SOI) is sampled by computing its dot product with a set of sampling kernels. The reconstruction phase estimates the SOI from the observed samples. In many cases, the number of samples required for a good estimate is much smaller than other methods (e.g., Nyquist sampling at twice the maximum frequency).

The recent excitement about CS is largely motivated by a series of practical signal reconstruction algorithms: basis pursuit [1, 4, 5], LASSO [6, 7], and approximate message-passing (AMP) [8, 9]. The original basis-pursuit algorithm was actually developed to solve the (essentially equivalent) *best-basis selection* problem, whose goal is to find a good sparse approximation for a vector in terms of an *over-complete dictionary* of vectors. Best-basis selection is a classic problem in signal processing and statistics and basis-pursuit defines a linear program (LP) which efficiently approximates the answer. The LASSO uses a linearly constrained quadratic program to solve the same problem when there is measurement noise. Finally, AMP provides a very efficient message-passing solution for the LASSO problem. Other natural applications of these algorithms include compression, denoising, and super-resolution estimation.

To motivate CS, assume that some linear transform of the SOI is either (i) exactly sparse or (ii) approximately sparse. For example, a 1D piecewise constant signal with a small number of jumps is exactly sparse after first-order differencing because the number of non-zero elements in the difference sequence is equal to the number of jumps. This result is easily generalized to piecewise polynomial functions by considering higher order differences. Likewise, many 2D images associated real applications are approximately sparse in some wavelet transform domain. In fact, most 2D signals associated with physical phenomena (e.g., topographic maps) "nice" enough to be approximately sparse in some natural transform domain. In general, compressed sensing can be applied to signals which are approximately sparse in some known transform domain.

One can compare sampling and reconstruction algorithms for CS in terms of their efficiency and complexity. The efficiency of a CS system increases when the number of samples required for "good" reconstruction decreases. A system has near-optimum efficiency if the number of samples required is only slightly larger than the theoretical minimum. For exactly sparse signals, it is now known that the theoretical minimum of one measurement per non-zero entry in the vector is asymptotically achievable [10, 11]. The complexity of a CS system is the computational complexity of signal reconstruction algorithm relative to the problem size. Implementation of the basis-pursuit algorithm based on a direct application of LP has roughly cubic complexity in the problem size [1], but this can reduced significantly. In particular, the message-passing reconstruction techniques are linear in the signal dimension [12, 13, 8].

1.1 Basic Problem Statement

In CS, the reconstruction problem is to take m dot-product samples of a length-n discrete-time *signal-of-interest* (SOI) and then to use those samples to reconstruct the SOI. For simplicity, we assume that the SOI is (approximately) sparse. Let $\underline{x} \in \mathbb{R}^n$ the signal vector, $\Phi \in \mathbb{R}^{m\times n}$ be the $m \times n$ measurement matrix, and $\underline{y} = \Phi\underline{x}$ be the m linear observations of $\underline{x}$. If $m < n$, then unique reconstruction of $\underline{x}$ from $\underline{y}$ is not possible without making some assumptions. For example, if one assumes the vector $\underline{x}$ is i.i.d. zero-mean Laplacian (i.e., entries drawn$\sim f_X(x) \propto e^{-\lambda|x|}$), then a *maximum a posteriori* (MAP) estimate $\hat{\underline{x}}$ is recovered by the basis-pursuit problem

$$\hat{\underline{x}} = \arg \min_{\underline{x}:\Phi\underline{x}=\underline{y}} \|\underline{x}\|_1 ,$$

where $\|\underline{x}\|_p^p \triangleq \sum_{i=0}^{n-1} |x_i|^p$ defines the p-norm[1] for $p \in (0, \infty)$.

Exercise 1.1. Formulate basis-pursuit problem as a linear program in $2n$ variables using the decomposition $\underline{x} = \underline{x}^+ - \underline{x}^-$, where $\underline{x}^+ = (x_1^+, \ldots, x_n^+) \in \mathbb{R}^n_{\geq 0}$ and $\underline{x}^- = (x_1^-, \ldots, x_n^-) \in \mathbb{R}^n_{\geq 0}$.

If instead, we know that the signal has independent entries that are non-zero with probability $\theta < 1/2$, then all MAP solutions are recovered by

$$\hat{\underline{x}} = \arg \min_{\underline{x}:\Phi\underline{x}=\underline{y}} \|\underline{x}\|_H ,$$

where $\|\underline{x}\|_H = \lim_{p\to 0} \|\underline{x}\|_p^p$ is the Hamming weight of $\underline{x}$. Solving this problem, however, is NP Hard. In both cases, the purpose of the prior distribution for $\boldsymbol{x}$ is simply to provide a partial ordering of the set $\left\{\underline{x} \in \mathbb{R}^n \mid \Phi\underline{x} = \underline{y}\right\}$ so that the "best" element can be chosen.

In the case of measurement noise, a common assumption is that $\underline{y} = \Phi\underline{x} + \underline{z}$, where $\underline{z} \in \mathbb{R}^n$ is additive noise. If one assumes that z is i.i.d. standard Gaussian noise and that the signal is i.i.d. zero-mean Laplacian with parameter λ, then a MAP estimate $\hat{\underline{x}}$ is recovered by the LASSO problem

$$\hat{\underline{x}} = \arg \min_{x\in\mathbb{R}^n} \frac{1}{2} \left\|\underline{y} - \Phi\underline{x}\right\|_2^2 + \lambda \|\underline{x}\|_1 . \tag{1}$$

The case where $\underline{x}$ is only sparse in some transform domain $\underline{y} = \Psi\underline{x}$ can be handled by replacing $\|\underline{x}\|_1$ by $\|\Psi\underline{x}\|_1$ in basis-pursuit and LASSO problems. It is well-known in convex optimization that a point, $\underline{z}$, is a minimizer of a convex function $h : \mathbb{R}^n \to \mathbb{R}$ iff $h(\underline{x}) \geq h(\underline{z})$ for all $\underline{x} \in \mathbb{R}^n$. This is equivalent to saying that $\underline{0} \in \partial h(\underline{x})$ (i.e., $\underline{0}$ is a subgradient of h at $\underline{x}$). Applying this to (1) gives the necessary and sufficient condition

$$\left[\Phi^\top(\underline{y} - \Phi\underline{x})\right]_i \in \begin{cases} \{\lambda\} & \text{if } x_i > 0 \\ \{-\lambda\} & \text{if } x_i < 0 \\ [-\lambda, \lambda] & \text{if } x_i = 0. \end{cases} \tag{2}$$

Exercise 1.2. Formulate the LASSO problem as a quadratic program in $2n$ variables using the decomposition $\underline{x} = \underline{x}^+ - \underline{x}^-$, where $\underline{x}^+ = (x_1^+, \ldots, x_n^+) \in \mathbb{R}^n_{\geq 0}$ and $\underline{x}^- = (x_1^-, \ldots, x_n^-) \in \mathbb{R}^n_{\geq 0}$.

The most impressive results of CS are based on approximately sparse signals defined by $\|\underline{x}\|_p \leq \infty$ for some $p \in (0, 2)$. These signals occur in practice as the wavelet transforms of natural image classes [4]. The reason these signals are interesting is that they can be approximated closely in the 2-norm based only their top N elements of largest magnitude. For example, if $\underline{x}^{(N)}$ is a modified version of $\underline{x}$ where all but the largest N elements are set to zero, then

$$\left\|\underline{x} - \underline{x}^{(N)}\right\|_2 \leq C_p \|\underline{x}\|_p N^{1/2-1/p}, \tag{3}$$

where C_p is a constant. From a geometric point of view, the reason for this is that points in the n-dimensional p-norm ball $\|\underline{x}\|_p \leq R$ cannot lie very far away from a sparse subspace. An impressive result of [4] is that the basis-pursuit solution $\hat{\underline{x}}$ with m measurements satisfies

$$\|\underline{x} - \hat{\underline{x}}\|_2 \leq C'_p \|\underline{x}\|_p \left(\frac{m}{\log n}\right)^{1/2-1/p} .$$

[1] For $p \in (0, 1)$, this is only a quasi-norm because it does not satisfy the triangle inequality.

Since (3) required knowledge of the position of the largest N coefficients, we see that LP gives performance is within a $\log n$ factor of sampling only the top N elements.

Whether or not this $\log n$ factor can be eliminated was an important open question in CS for some time. Message-passing reconstruction, for the noiseless case, provided an early hint that the $\log n$ can be removed [13, 14]. Results from information theory then showed that this is possible using optimal decoding [10]. In 2011, those results were extended to the noisy case [11] and to low-complexity reconstruction [15, 16] based on spatially-coupled measurement matrices [17].

2 Bayesian Reconstruction

In the early compressive sensing literature, the typical signal model is that the signal $\underline{x} \in \mathbb{R}^n$ is a random vector drawn uniformly from the n-dimensional unit $\|\cdot\|_q$ ball. For $0 < q < 2$, this enforces an approximate sparsity condition. Recent advances have been made using a stochastic model for the signal. For example, one simple assumption is that each component of the signal vector is drawn i.i.d. according to some distribution $f_X(\cdot)$. In this case, optimal reconstruction refers either to the MAP or or minimum mean-square error (MMSE) estimate of $\underline{x}$ given the observation $\underline{y}$. This also allows us to view the CS reconstruction problem in a Bayesian framework where a very important role is played by the prior distribution assumed for the signal.

Consider a signal where, conditional on an i.i.d. Bernoulli(p) random variable, each entry is either exactly zero or chosen i.i.d. from a Gaussian distribution with mean μ_1 and variance σ_1^2. Such a signal is called Bernoulli-Gaussian and can be modeled by choosing its pdf to be

$$f_X(x) = (1-p)\delta(x) + \frac{p}{\sigma_1}\phi\left(\frac{x-\mu_1}{\sigma_1}\right),$$

where $\phi(x)$ is the pdf of a standard Gaussian. For approximately sparse signals, one can use instead a two-Gaussian model

$$f_X(x) = \frac{1-p}{\sigma_0}\phi\left(\frac{x}{\sigma_0}\right) + \frac{p}{\sigma_1}\phi\left(\frac{x-\mu_1}{\sigma_1}\right),$$

where choosing $\sigma_0 \ll \sigma_1$ results in roughly $(1-p)n$ entries that are close to zero.

For a random observation $\underline{Y} = \Phi\underline{X} + \underline{Z}$ of the signal $\underline{X}$, it is generally very difficult to find either the MAP estimate

$$\underline{\hat{x}}^{MAP} = \arg\max_{\underline{x}\in\mathbb{R}^n} \Pr\left(\underline{X} = \underline{x} \,|\, \underline{Y} = \underline{y}\right)$$

or the MMSE estimate

$$\underline{\hat{x}}^{MMSE} = \arg\min_{\underline{x}\in\mathbb{R}^n} \mathbb{E}\left[(\underline{X}-\underline{x})^2|\underline{Y}=\underline{y}\right] = \mathbb{E}\left[\underline{X}|\underline{Y}\right].$$

This type of inference problem can be tackled by defining an appropriate factor graph and using belief propagation. If the signal alphabet is small, then the complexity is quite manageable. But, for the real numbers, this leads to messages that are functions (e.g., pdfs). In practice, one must either quantize the alphabet or approximate the message functions. The former approach is discussed in [18, 19] and has some issues with complexity and accuracy. The latter approach was taken in [20] where relaxed belief-propagation (RBP) was defined based on a Gaussian approximation.

Historically, some of the approximations used for the CS problem were originally introduced for the code-division multiple-access problem [21, 22, 23]. These approaches led to RBP for CS in [20] and a more efficient version called approximate message-passing (AMP) in [8, 9]. These algorithms gives rise to a natural conjectured density evolution analysis called state evolution. For large problem sizes with Gaussian measurement matrices, this analysis can be made precise [24] and one finds that the AMP algorithm solves the LASSO problem in this case. The state-of-the-art in factor-graph based CS currently consists of combining specially-designed pseudo-random measurement matrices with the AMP algorithm for reconstruction [17, 15, 16]. One big advantage of Bayesian reconstruction is the factor-graph framework which leads naturally to low-complexity message-passing solutions. In particular, this framework allows one to tie together factor graphs for many different systems and easily compute approximate solutions to the joint MAP estimation problem [25, 26].

3 Factor Graph Formulation

In this section, it is assumed that $\underline{X}$ is drawn independently $X_i \sim f_{X_i}(x_i)$. Then, the measurements are computed with $\underline{Y} = \Phi \underline{X} + \underline{Z}$ where $\underline{Z}$ is drawn independently $Z_a \sim f_{Z_a}(z_a)$. In this case, the joint pdf is given by

$$f_{\underline{X},\underline{Y}}\left(\underline{x},\underline{y}\right) = \left(\prod_{i=1}^{n} f_{X_i}(x_i)\right)\left(\prod_{a=1}^{m} f_{Z_a}\left(y_a - \sum_{i=1}^{n} \Phi_{a,i} x_i\right)\right).$$

This naturally defines a factor graph with variable nodes $V = [n]$ and factor nodes $F = [m]$. The set of edges is defined to be $E \triangleq \{(i,a) \in V \times F \,|\, \Phi_{a,i} \neq 0\}$. Let $V(a) \triangleq \{i \in V \mid (i,a) \in E\}$ denote the set of variable nodes (VNs) adjacent to a and $F(i) \triangleq \{a \in F \mid (i,a) \in E\}$ denote the set of factor nodes (FNs) adjacent to i.

In contrast to previously studied problems, we now have $\mathcal{X} = \mathbb{R}$ and the sums in the standard belief-propagation equations must be replaced by integrals over $\mathbb{R}$. The messages also become functions (e.g., pdfs) and we find that

$$\hat{\mu}^{(\ell)}_{a\to i}(x_i) = \frac{1}{\hat{C}^{(\ell)}_{a\to i}} \int_{\underline{x}\setminus x_i} f_{Z_a}\left(y_a - \sum_{k\neq i} \Phi_{a,k} x_k - \Phi_{a,i} x_i\right) \prod_{j\neq i} \mu^{(\ell)}_{j\to a}(x_j) \mathrm{d}x_j$$

$$\mu^{(\ell+1)}_{i\to a}(x_i) = \frac{1}{C^{(\ell+1)}_{i\to a}} f_X(x_i) \prod_{b\neq a} \hat{\mu}^{(\ell)}_{b\to i}(x_i),$$

where $\hat{C}^{(\ell)}_{a\to i}$ and $C^{(\ell+1)}_{i\to a}$ are the normalization factors that guarantee $\int \hat{\mu}^{(\ell)}_{a\to i}(x_i)\mathrm{d}x_i = \int \mu^{(\ell+1)}_{i\to a}(x_i)\mathrm{d}x_i = 1$.

3.1 Relaxed Belief Propagation

In this section, we derive the RBP algorithm which was introduced for CS [20]. This can be seen as an approximation of belief propagation whose error vanishes asymptotically as $n \to \infty$. This algorithm is closely related to the AMP algorithm and was introduced for CS as "relaxed belief propagation" in [20].

3.1.1 FN-toVN Message Update

Let us first consider the FN-to-VN message update

$$\begin{aligned}\hat{\mu}^{(\ell)}_{a\to i}(x_i) &\propto \int_{\underline{x}\setminus x_i} f_{Z_a}\left(y_a - \sum_{k\neq i} \Phi_{a,k} x_k - \Phi_{a,i} x_i\right) \prod_{j\neq i} \mu^{(\ell)}_{j\to a}(x_j)\mathrm{d}x_j \\ &= \int \mathrm{d}z_a \int_{\underline{x}\setminus x_i} \delta\left(z_a - y_a + \sum_{k\neq i} \Phi_{a,k} x_k + \Phi_{a,i} x_i\right) \exp\left(-\frac{z_a^2}{2\sigma_a^2}\right) \prod_{j\neq i} \mu^{(\ell)}_{j\to a}(x_j)\mathrm{d}x_j. \end{aligned} \tag{4}$$

The following lemma will simplify our derivation.

Lemma 3.1. *If $f_{\underline{W}}(\underline{w})$ is the pdf of a random vector $\underline{W} \in \mathbb{R}^p$ and $g : \mathbb{R}^p \to \mathbb{R}$, then $V = \frac{1}{\beta} g(\underline{W})$ is equivalent to*

$$f_V(v) \propto \int \delta\left(\beta v - g(\underline{w})\right) f_{\underline{W}}(\underline{w})\mathrm{d}\underline{w}. \tag{5}$$

Proof. If $V = \frac{1}{\beta} g(\underline{W})$, then

$$\Pr(V \leq v) = \Pr\left(g(\underline{W}) \leq \beta v\right) = \int u\left(\beta v - g(w)\right) f_{\underline{W}}(\underline{w})\mathrm{d}\underline{w},$$

where u is the unit step function. Thus,

$$f_V(v) = \frac{\mathrm{d}}{\mathrm{d}v} \int u\left(\beta v - g(w)\right) f_{\underline{W}}(\underline{w})\mathrm{d}\underline{w} = \beta \int \delta\left(\beta v - g(w)\right) f_{\underline{W}}(\underline{w})\mathrm{d}\underline{w}.$$

Likewise, if (5) holds, then $f_V(v) = \beta \int \delta\left(\beta v - g(w)\right) f_{\underline{W}}(\underline{w}) \mathrm{d}\underline{w}$ because

$$\int \beta \int \delta\left(\beta v - g(\underline{w})\right) f_{\underline{W}}(\underline{w}) \mathrm{d}\underline{w} \mathrm{d}v = \int \int \delta\left(v' - g(\underline{w})\right) \mathrm{d}v' f_{\underline{W}}(\underline{w}) \mathrm{d}\underline{w} = \int f_{\underline{W}}(\underline{w}) \mathrm{d}\underline{w} = 1.$$

Working backwards from $f_V(v)$ through the equalities of the first part proves that $V = \frac{1}{\beta} g(\underline{W})$. □

To apply this lemma to (4), we first define independent random variables $X_{i\to a}^{(\ell)} \sim \mu_{i\to a}^{(\ell)}(x)$ and $\hat{X}_{a\to i}^{(\ell)} \sim \hat{\mu}_{i\to a}^{(\ell)}(x)$ for all $(i,a) \in [n] \times [m]$. Then, we let $\underline{W} = \left(Z_a, \{X_{k\to a}\}_{k\neq i}\right)$ be a n-dimensional random vector and observe that its joint pdf is

$$f_{\underline{W}}\left(\{x_k\}_{k\neq i}, z_a\right) = f_{Z_a}(z_a) \prod_{k\neq i} f_{X_{k\to a}}(x_k) \propto \exp\left(-\frac{z_a^2}{2\sigma_a^2}\right) \prod_{k\neq i} \mu_{k\to a}^{(\ell)}(x_k).$$

Using $g(\{x_k\}_{k\neq i}, z_a) = y_a - z_a - \sum_{k\neq i} \Phi_{a,k} x_k$, $\beta = \Phi_{a,i}$, and $V = \hat{X}_{a\to i}$ in Lemma 3.1 shows that (4) is equivalent to

$$\hat{X}_{a\to i} \stackrel{d}{=} \frac{1}{\Phi_{a,i}} \left(y_a - Z_a - \sum_{k\neq i} \Phi_{a,k} X_{k\to a}\right). \tag{6}$$

By construction, the random variables Z_a and $X_{k\to a}$ for $k \neq i$ are all independent. If they are also Gaussian, then (6) implies that $\hat{X}_{a\to i}$ is also Gaussian[2]. If $\Phi_{a,i}$ scales like $\Theta(\frac{1}{\sqrt{n}})$, then the same conclusion holds asymptotically even if the $X_{k\to a}$ are not Gaussian! Focusing on the sum in (6), we see that $S_n = \sum_{k\neq i} \Phi_{a,k} X_{k\to a}$ is the scaled sum of $n-1$ independent random variables. Under mild technical conditions, the central limit theorem (CLT) implies that $S_n - \mathbb{E}[S_n]$ converges in distribution, as $n \to \infty$, to a Gaussian random variable with variance $\sum_{k\neq i} \Phi_{a,k}^2 \mathrm{Var}(X_{k\to a})$. In the limit, the distribution of $\hat{X}_{a\to i}$ depends only on the mean and variance of $X_{k\to a}$ for $k \neq i$, which are denoted

$$\begin{aligned} m_{k\to a}^{(\ell)} &= \int \mu_{k\to a}^{(\ell)}(x_k) x_k \mathrm{d}x_k = \mathbb{E}\left[X_{k\to a}^{(\ell)}\right] \\ v_{k\to a}^{(\ell)} &= \int \mu_{k\to a}^{(\ell)}(x_k) \left(x_k - m_{k\to a}^{(\ell)}\right)^2 \mathrm{d}x_k = \mathrm{Var}\left(X_{k\to a}^{(\ell)}\right). \end{aligned}$$

If we define

$$\begin{aligned} \hat{m}_{a\to i}^{(\ell)} &= \int \hat{\mu}_{a\to i}^{(\ell)}(x_i) x_i \mathrm{d}x_i = \mathbb{E}\left[\hat{X}_{a\to i}^{(\ell)}\right] \\ \hat{v}_{a\to i}^{(\ell)} &= \int \hat{\mu}_{a\to i}^{(\ell)}(x_i) \left(x_i - \hat{\mu}_{a\to i}^{(\ell)}\right)^2 \mathrm{d}x_i = \mathrm{Var}\left(\hat{X}_{a\to i}^{(\ell)}\right), \end{aligned}$$

then (6) implies that

$$\begin{aligned} \hat{m}_{a\to i}^{(\ell)} &= \frac{1}{\Phi_{a,i}} \left(y_a - \sum_{k\neq i} \Phi_{a,k} m_{k\to a}^{(\ell)}\right) \\ \hat{v}_{a\to i}^{(\ell)} &= \frac{1}{\Phi_{a,i}^2} \left(\sigma_a^2 + \sum_{k\neq i} \Phi_{a,k}^2 v_{k\to a}^{(\ell)}\right). \end{aligned} \tag{7}$$

Remark 3.2. The error in this Gaussian approximation can be bounded by tracking the absolute third moment of the $\mu_{k\to a}^{(\ell)}(x_k)$ messages and applying the Berry-Esseen central limit theorem.

[2]This happens when the signal values are independent Gaussians $X_i \sim \mathcal{N}(m_i, v_i)$ and the observation noise values are independent zero-mean Gaussians $Z_a \sim \mathcal{N}(0, \sigma_a^2)$. In this case, the algorithm derived here is known as Gaussian belief propagation.

3.2 VN-to-FN Update

The following simple lemma will simply our analysis of the VN-to-FN update.

Lemma 3.3. *A function* $\nu(x)$ *is equal to the Gaussian function* $\frac{1}{\sqrt{2\pi v}} e^{-(x-m)^2/(2v)}$ *if and only if* $\int \nu(x) \mathrm{d}x = 1$ *and* $\nu(x) \propto \exp\left(-\frac{1}{2v}x^2 + \frac{m}{v}x\right)$.

Proof. The "only if" is immediate. The "if" follows from completing the square $-\frac{1}{2v}x^2 + \frac{m}{v}x = -\frac{1}{2v}\left[(x-m)^2 + m^2\right]$ and computing the Gaussian integral. □

Using the fact that $\hat{\mu}_{b\to i}^{(\ell)}(x_b)$ is asymptotically Gaussian, we use Lemma 3.3 to write

$$\hat{\mu}_{b\to i}^{(\ell)}(x_b) \mathring{\propto} \exp\left(-\frac{1}{2\hat{v}_{b\to i}^{(\ell)}} x_i^2 + \frac{\hat{m}_{b\to i}^{(\ell)}}{\hat{v}_{b\to i}^{(\ell)}} x_i\right),$$

where the symbol $\mathring{\propto}$ is used to denote approximately proportional (i.e., the two sides are approximately equal after normalization). Next, we can analyze the VN-to-FN messages $\mu_{i\to a}^{(\ell+1)}(x_i)$ via

$$\begin{aligned}
\mu_{i\to a}^{(\ell+1)}(x_i) &\propto f_{X_i}(x_i) \prod_{b\neq a} \hat{\mu}_{b\to i}^{(\ell)}(x_b) \\
&\mathring{\propto} f_{X_i}(x_i) \prod_{b\neq a} \exp\left(-\frac{1}{2\hat{v}_{b\to i}^{(\ell)}} x_i^2 + \frac{\hat{m}_{b\to i}^{(\ell)}}{\hat{v}_{b\to i}^{(\ell)}} x_i\right) \\
&= f_{X_i}(x_i) \exp\left(-x_i^2 \sum_{b\neq a} \frac{1}{2\hat{v}_{b\to i}^{(\ell)}} + x_i \sum_{b\neq a} \frac{\hat{m}_{b\to i}^{(\ell)}}{\hat{v}_{b\to i}^{(\ell)}}\right) \\
&= f_{X_i}(x_i) \exp\left(-x_i^2 \frac{1}{2\tilde{v}_{i\to a}^{(\ell)}} + x_i \frac{\tilde{m}_{i\to a}^{(\ell)}}{\tilde{v}_{i\to a}^{(\ell)}}\right),
\end{aligned} \tag{8}$$

where matching coefficients in the last two steps shows that

$$\begin{aligned}
\tilde{v}_{i\to a}^{(\ell)} &= \frac{1}{\sum_{b\neq a} 1/\hat{v}_{b\to i}^{(\ell)}} \\
\tilde{m}_{i\to a}^{(\ell)} &= \tilde{v}_{i\to a}^{(\ell)} \sum_{b\neq a} \frac{\hat{m}_{b\to i}^{(\ell)}}{\hat{v}_{b\to i}^{(\ell)}}.
\end{aligned} \tag{9}$$

Applying Lemma 3.3 to (8) shows that

$$\mu_{i\to a}^{(\ell+1)}(x_i) \mathring{\propto} f_{X_i}(x_i) \exp\left(-\frac{1}{2\tilde{v}_{i\to a}^{(\ell)}} x_i^2 + \frac{\tilde{m}_{i\to a}^{(\ell)}}{\tilde{v}_{i\to a}^{(\ell)}} x_i\right). \tag{10}$$

If $X_i \sim \mathcal{N}(m_i, v_i)$ is Gaussian, then it is easy to verify that

$$\begin{aligned}
v_{i\to a}^{(\ell+1)} &= \int \mu_{i\to a}^{(\ell)}(x_i) \left(x_i - m_{i\to a}^{(\ell+1)}\right)^2 \mathrm{d}x_i = \frac{1}{(1/v_i) + \sum_{b\neq a} 1/\hat{v}_{a\to i}^{(\ell)}}. \\
m_{i\to a}^{(\ell+1)} &= \int \mu_{i\to a}^{(\ell+1)}(x_k) x_k \mathrm{d}x_k = v_{i\to a}^{(\ell+1)} \left(\frac{m_i}{v_i} + \sum_{b\neq a} \frac{\hat{m}_{b\to i}^{(\ell)}}{\hat{v}_{a\to i}^{(\ell)}}\right).
\end{aligned}$$

This update is leads to Gaussian belief propagation. Otherwise, $\mu_{i\to a}^{(\ell+1)}(x_i)$ is not Gaussian and it is not immediately clear how to proceed. Fortunately, this analysis of the FN-to-VN messages depends only on their mean and variance! Moreover, for a fixed $f_{X_i}(x_i)$, (10) mean and variance of $\mu_{i\to a}^{(\ell+1)}(x_i)$ only depend on $\tilde{m}_{a\to i}^{(\ell)}$ and $\tilde{v}_{a\to i}^{(\ell)}$. Thus, one can write the updates

$$\begin{aligned}
m_{i\to a}^{(\ell+1)} &= \int \mu_{i\to a}^{(\ell+1)}(x_k) x_k \mathrm{d}x_k = \eta_i\left(\tilde{m}_{a\to i}^{(\ell)}, \tilde{v}_{a\to i}^{(\ell)}\right) \\
v_{i\to a}^{(\ell+1)} &= \int \mu_{i\to a}^{(\ell)}(x_i) \left(x_i - m_{i\to a}^{(\ell)}\right)^2 \mathrm{d}x_i = \theta_i\left(\tilde{m}_{a\to i}^{(\ell)}, \tilde{v}_{a\to i}^{(\ell)}\right)
\end{aligned}$$

in terms of the functions

$$\eta_i(m,v) = \frac{\int f_{X_i}(x)\exp\left(-\frac{1}{2v}(x-m)^2\right) x\,\mathrm{d}x}{\int f_{X_i}(x)\exp\left(-\frac{1}{2v}(x-m)^2\right)\mathrm{d}x}$$

$$\theta_i(m,v) = \frac{\int f_{X_i}(x)\exp\left(-\frac{1}{2v}(x-m)^2\right) x^2\,\mathrm{d}x}{\int f_{X_i}(x)\exp\left(-\frac{1}{2v}(x-m)^2\right)\mathrm{d}x} - (\eta_i(m,v))^2.$$

Remark 3.4. Closed form expressions for η_i and θ_i can be computed for many priors including the Laplacian prior $f_{X_i}(x) = e^{-\beta|x|}$ and the Bernoulli-Gaussian prior $f_{X_i}(x) = (1-p)\delta(x) + \frac{p}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)$. For example, see Section 6. Also, for any prior, direct computation using the chain rule shows that

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}m}\eta_i(m,v) &= \frac{\int f_{X_i}(x)\exp\left(-\frac{1}{2v}x^2 + \frac{m}{v}x\right)\frac{1}{v}x^2\,\mathrm{d}x}{\int f_{X_i}(x)\exp\left(-\frac{1}{2v}x^2 + \frac{m}{v}x\right)\mathrm{d}x} \\
&\quad - \frac{\int f_{X_i}(x)\exp\left(-\frac{1}{2v}x^2 + \frac{m}{v}x\right)x\,\mathrm{d}x \int f_{X_i}(x)\exp\left(-\frac{1}{2v}x^2 + \frac{m}{v}x\right)\frac{1}{v}x\,\mathrm{d}x}{\left(\int f_{X_i}(x)\exp\left(-\frac{1}{2v}x^2 + \frac{m}{v}x\right)\mathrm{d}x\right)^2} \\
&= \frac{1}{v}\left[\frac{\int f_{X_i}(x)\exp\left(-\frac{1}{2v}x^2 + \frac{m}{v}x\right)x^2\,\mathrm{d}x}{\int f_{X_i}(x)\exp\left(-\frac{1}{2v}x^2 + \frac{m}{v}x\right)\mathrm{d}x} - \left(\frac{\int f_{X_i}(x)\exp\left(-\frac{1}{2v}x^2 + \frac{m}{v}x\right)x\,\mathrm{d}x}{\int f_{X_i}(x)\exp\left(-\frac{1}{2v}x^2 + \frac{m}{v}x\right)\mathrm{d}x}\right)^2\right] \\
&= \frac{1}{v}\theta_i(m,v). \qquad (11)
\end{aligned}$$

The Relaxed Belief Propagation Algorithm:

$$\begin{aligned}
m_{i\to a}^{(0)} &= \int f_{X_i}(x_i)x_i\,\mathrm{d}x \\
v_{i\to a}^{(0)} &= \int f_{X_i}(x_i)\left(x_i - m_{i\to a}^{(0)}\right)^2\mathrm{d}x \\
\hat{m}_{a\to i}^{(\ell)} &= \frac{1}{\Phi_{a,i}}\left(y_a - \sum_{k\neq i}\Phi_{a,k}m_{k\to a}^{(\ell)}\right) \\
\hat{v}_{a\to i}^{(\ell)} &= \frac{1}{\Phi_{a,i}^2}\left(\sigma_a^2 + \sum_{k\neq i}\Phi_{a,k}^2 v_{k\to a}^{(\ell)}\right) \\
\tilde{v}_{i\to a}^{(\ell)} &= \frac{1}{\sum_{b\neq a}1/\hat{v}_{b\to i}^{(\ell)}} \\
\tilde{m}_{i\to a}^{(\ell)} &= \tilde{v}_{i\to a}^{(\ell)}\sum_{b\neq a}\frac{\hat{m}_{b\to i}^{(\ell)}}{\hat{v}_{b\to i}^{(\ell)}} \\
m_{i\to a}^{(\ell+1)} &= \eta_i\left(\tilde{m}_{a\to i}^{(\ell)}, \tilde{v}_{a\to i}^{(\ell)}\right) \\
v_{i\to a}^{(\ell+1)} &= \theta_i\left(\tilde{m}_{a\to i}^{(\ell)}, \tilde{v}_{a\to i}^{(\ell)}\right)
\end{aligned}$$

4 Reducing the Number of Messages

In the last section, we saw that a Gaussian approximation can be used to simplify belief propagation from a recursion on functions to recursion on real numbers. Still, the algorithm requires tracking and updating mn messages. Moreover, a brief inspection shows that many of these messages only differ slightly from each other. In this, section we will describe how to reduce the the number of messages to $m+n$. The idea comes from statistical physics and is based on the work of Thouless, Anderson, and

Palmer (TAP). In signal processing and communications, it was first applied to CDMA in [22]. These notes are based on the derivation in [27]. We consider $n \to \infty$ along a sequence where $m = \alpha n$ is integer and assume that the matrix entries are given by $\Phi_{a,i}^{(n)} = \frac{1}{\sqrt{n}} \tilde{\Phi}_{a,i}^{(n)}$, where $\tilde{\Phi}_{a,i}^{(n)}$ is an $m \times n$ array of i.i.d. random variables with mean zero, variance one, and exponential (or faster) tails. Extreme-value theory allows one to bound, with high probability, the maximum entry in $\Phi_{a,i}^{(n)}$ to show that $|\Phi_{a,i}^{(n)}| = O(\frac{\ln n}{\sqrt{n}})$ and $|\Phi_{a,i}^{(n)}|^2 = O(\frac{\ln n}{n})$ hold uniformly in (a,i). Two important special cases are where Φ is Gaussian (i.e., $\tilde{\Phi}_{a,i}^{(n)}$ is standard Gaussians) and Φ is Rademacher (i.e., $\tilde{\Phi}_{a,i}^{(n)}$ is equiprobable on $\{+1, -1\}$). The superscript n is hereafter dropped from $\Phi_{a,i}^{(n)}$ and $\tilde{\Phi}_{a,i}^{(n)}$. We also assume that $\sigma_a^2 = \sigma^2$ for all a and $f_{X_i}(x) = f_X(x)$ for all i.

Consider the estimate of $U_a \triangleq Y_a - Z_a = \sum_i \Phi_{a,i} X_i$ generated by belief propagation from the $\mu_{k \to a}^{(\ell)}(x_k)$ messages after ℓ iterations. This estimate can be written as

$$\nu_a^{(\ell)}(u_a) \propto \int_{\underline{x}} \delta\left(u_a - \sum_{k=1}^{n} \Phi_{a,k} x_k\right) \prod_{k=1}^{n} \mathrm{d}x_k \mu_{k \to a}^{(\ell)}(x_k).$$

Applying Lemma 3.1 shows that the distribution $\nu_a^{(\ell)}(u_a)$ can be associated with the random variable

$$\hat{U}_a \overset{d}{=} \sum_{i=1}^{n} \Phi_{a,i} X_{i \to a}$$

and the CLT trick implies that

$$\nu_a^{(\ell)}(u_a) \mathring{\propto} \exp\left(-u_a^2 \frac{1}{2V_a^{(\ell)}} - u_a \frac{\omega_a^{(\ell)}}{V_a^{(\ell)}}\right),$$

where

$$\begin{aligned} \omega_a^{(\ell)} &\triangleq \sum_{i=1}^{n} \Phi_{a,i} m_{i \to a}^{(\ell)} \\ V_a^{(\ell)} &\triangleq \sum_{i=1}^{n} \Phi_{a,i}^2 v_{i \to a}^{(\ell)}. \end{aligned} \tag{12}$$

These two quantities can be interpreted as the node-based quantities of interest for the factor nodes. We also consider the node-based quantities of interest for the variable nodes. Starting with (9), we define

$$\begin{aligned} \tilde{v}_i^{(\ell)} &\triangleq \frac{1}{\sum_{a=1}^{m} 1/\hat{v}_{a \to i}^{(\ell)}} \\ \tilde{m}_i^{(\ell)} &\triangleq \tilde{v}_i^{(\ell)} \sum_{a=1}^{m} \frac{\hat{m}_{a \to i}^{(\ell)}}{\hat{v}_{a \to i}^{(\ell)}} \end{aligned} \tag{13}$$

and

$$\begin{aligned} m_i^{(\ell)} &\triangleq \frac{1}{m} \sum_{a=1}^{m} m_{i \to a}^{(\ell)} \\ v_i^{(\ell)} &\triangleq \frac{1}{m} \sum_{a=1}^{m} v_{i \to a}^{(\ell)}. \end{aligned} \tag{14}$$

It is worth noting that some quantities used in the algorithm become deterministic as $n \to \infty$ while others depend on the problem instance. For example, consider the quantity

$$V_a^{(\ell)} = \sum_{i=1}^{n} \frac{1}{n} \tilde{\Phi}_{a,i}^2 v_{i \to a}^{(\ell)}.$$

If $\tilde{\Phi}_{a,i} \in \{+1, -1\}$ for all (a, i), then it follows that $V_a^{(\ell)} = \frac{1}{n}\sum_{i=1}^{n} v_{i\to a}^{(\ell)}$. More generally, if we assume that $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \tilde{\Phi}_{a,i}^2 v_{i\to a}^{(\ell)}$ satisfies a central limit theorem, then the Var $\left(V_a^{(\ell)}\right) = O(\frac{1}{n})$ and it follows that

$$V_a^{(\ell)} = \frac{1}{n}\sum_{i=1}^{n} v_{i\to a}^{(\ell)} + O\left(\frac{\ln n}{\sqrt{n}}\right).$$

In contrast, the quantity

$$\omega_a^{(\ell)} = \sum_{i=1}^{n} \frac{1}{\sqrt{n}} \tilde{\Phi}_{a,i} m_{i\to a}^{(\ell)}$$

remains random under the same assumptions (e.g., it has positive variance as $n \to \infty$).

To continue the derivation, it is important that we understand the rough order of each variable as $n \to \infty$. First, we note that the variances of the signal estimates, $v_{i\to a}^{(\ell)}$, should remain uniformly $O(1)$ as $n \to \infty$. This is automatically true if $f_X(x)$ has bounded support. Regardless, if things are working well, $v_{i\to a}^{(\ell)}$ should be non-increasing in ℓ and, thus, be upper bounded by $v_{i\to a}^{(0)}$. This also implies that $V_a^{(\ell)}$ is uniformly $O(1)$ as $n \to \infty$. From the formula for $\hat{v}_{a\to i}^{(\ell)}$, one may also conclude that, for a fixed (a, i), $\hat{v}_{a\to i}^{(\ell)} = \Theta(n)$ as $n \to \infty$. For a similar reason to $v_{i\to a}^{(\ell)}$, one can argue that $\tilde{v}_{i\to a}^{(\ell)}$ should remain $O(1)$ as $n \to \infty$. We will also assume that $\tilde{m}_i^{(\ell-1)} = O(\ln n)$ uniformly in i as $n \to \infty$ because extreme-value theory implies that both that the true signal value and the deviation of the estimate from the true signal value should satisfy this bound.

Lemma 4.1. *Using a first-order Taylor series expansion, we observe that*

$$m_{i\to a}^{(\ell)} \approx \eta\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right) + \eta^{(1,0)}\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right)\mathrm{d}m + \eta^{(0,1)}\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right)\mathrm{d}v \tag{15}$$

$$v_{i\to a}^{(\ell)} \approx \theta\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right) + \theta^{(1,0)}\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right)\mathrm{d}m + \theta^{(0,1)}\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right)\mathrm{d}v \tag{16}$$

where $\mathrm{d}m$ *and* $\mathrm{d}v$ *are given by*

$$\mathrm{d}m = \frac{\Phi_{a,i}^2 \tilde{v}_{i\to a}^{(\ell-1)} \tilde{m}_i^{(\ell-1)}}{\left(\sigma_a^2 + \sum_{k\neq i} \Phi_{a,k}^2 v_{k\to a}^{(\ell-1)}\right)} - \frac{\left(\hat{m}_{a\to i}^{(\ell-1)}/\hat{v}_{a\to i}^{(\ell-1)}\right)}{\left(\sum_{b=1}^{m} 1/\hat{v}_{b\to i}^{(\ell-1)}\right)}\left(1 + O\left(\frac{\ln n}{n}\right)\right)$$

$$\mathrm{d}v = \frac{\Phi_{a,i}^2 \tilde{v}_i^{(\ell-1)} \tilde{v}_{i\to a}^{(\ell-1)}}{\left(\sigma_a^2 + \sum_{k\neq i} \Phi_{a,k}^2 v_{k\to a}^{(\ell-1)}\right)} = O\left(\frac{\ln n}{n}\right).$$

Proof. To see this, we write

$$\begin{aligned} m_{i\to a}^{(\ell)} &= \eta\left(\frac{\sum_{b=1}^{m} \hat{m}_{b\to i}^{(\ell-1)}/\hat{v}_{b\to i}^{(\ell-1)} - \hat{m}_{a\to i}^{(\ell-1)}/\hat{v}_{a\to i}^{(\ell-1)}}{\sum_{b=1}^{m} 1/\hat{v}_{b\to i}^{(\ell-1)} - 1/\hat{v}_{a\to i}^{(\ell-1)}}, \frac{1}{\sum_{b=1}^{m} 1/\hat{v}_{b\to i}^{(\ell-1)} - 1/\hat{v}_{a\to i}^{(\ell-1)}}\right) \\ &\approx \eta\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right) + \eta^{(1,0)}\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right)\mathrm{d}m + \eta^{(0,1)}\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right)\mathrm{d}v, \end{aligned}$$

where dm and dv are differences implied by the first order Taylor series expansion. This implies that

$$
\begin{aligned}
\mathrm{d}m &= \frac{\sum_{b=1}^{m} \hat{m}_{b\to i}^{(\ell-1)}/\hat{v}_{b\to i}^{(\ell-1)} - \hat{m}_{a\to i}/\hat{v}_{a\to i}^{(\ell-1)}}{\sum_{b=1}^{m} 1/\hat{v}_{b\to i} - 1/\hat{v}_{a\to i}} - \frac{\sum_{b=1}^{m} \hat{m}_{b\to i}^{(\ell-1)}/\hat{v}_{b\to i}^{(\ell-1)}}{\sum_{b=1}^{m} 1/\hat{v}_{b\to i}} \\
&= \frac{\left(\sum_{b=1}^{m} \hat{m}_{b\to i}^{(\ell-1)}/\hat{v}_{b\to i}^{(\ell-1)} - \hat{m}_{a\to i}^{(\ell-1)}/\hat{v}_{a\to i}^{(\ell-1)}\right)\left(\sum_{b=1}^{m} 1/\hat{v}_{b\to i}^{(\ell-1)}\right) - \left(\sum_{b=1}^{m} 1/\hat{v}_{b\to i}^{(\ell-1)} - 1/\hat{v}_{a\to i}^{(\ell-1)}\right)\left(\sum_{b=1}^{m} \hat{m}_{b\to i}^{(\ell-1)}/\hat{v}_{b\to i}^{(\ell-1)}\right)}{\left(\sum_{b=1}^{m} 1/\hat{v}_{b\to i}\right)\left(\sum_{b=1}^{m} 1/\hat{v}_{b\to i} - 1/\hat{v}_{a\to i}\right)} \\
&= \frac{-\left(\hat{m}_{a\to i}^{(\ell-1)}/\hat{v}_{a\to i}^{(\ell-1)}\right)\left(\sum_b 1/\hat{v}_{b\to i}^{(\ell-1)}\right) + \left(1/\hat{v}_{a\to i}^{(\ell-1)}\right)\left(\sum_b \hat{m}_{b\to i}^{(\ell-1)}/\hat{v}_{b\to i}^{(\ell-1)}\right)}{\left(\sum_{b=1}^{m} 1/\hat{v}_{b\to i}^{(\ell-1)}\right)\left(\sum_{b=1}^{m} 1/\hat{v}_{b\to i}^{(\ell-1)} - 1/\hat{v}_{a\to i}^{(\ell-1)}\right)} \\
&= \frac{\left(1/\hat{v}_{a\to i}^{(\ell-1)}\right)\left(\sum_b \hat{m}_{b\to i}^{(\ell-1)}/\hat{v}_{b\to i}^{(\ell-1)}\right)}{\left(\sum_{b=1}^{m} 1/\hat{v}_{b\to i}^{(\ell-1)}\right)\left(\sum_{b=1}^{m} 1/\hat{v}_{b\to i}^{(\ell-1)} - 1/\hat{v}_{a\to i}^{(\ell-1)}\right)} - \frac{\left(\hat{m}_{a\to i}^{(\ell-1)}/\hat{v}_{a\to i}^{(\ell-1)}\right)}{\left(\sum_{b=1}^{m} 1/\hat{v}_{b\to i}^{(\ell-1)}\right)}\left(\frac{\left(\sum_{b=1}^{m} 1/\hat{v}_{b\to i}^{(\ell-1)}\right)}{\left(\sum_{b=1}^{m} 1/\hat{v}_{b\to i}^{(\ell-1)} - 1/\hat{v}_{a\to i}^{(\ell-1)}\right)}\right) \\
&= \frac{\Phi_{a,i}^2 \tilde{v}_{i\to a}^{(\ell-1)} \tilde{m}_i^{(\ell-1)}}{\left(\sigma_a^2 + \sum_{k\neq i} \Phi_{a,k}^2 v_{k\to a}^{(\ell-1)}\right)} - \frac{\left(\hat{m}_{a\to i}^{(\ell-1)}/\hat{v}_{a\to i}^{(\ell-1)}\right)}{\left(\sum_{b=1}^{m} 1/\hat{v}_{b\to i}^{(\ell-1)}\right)}\left(1 - \frac{\Phi_{a,i}^2 \tilde{v}_i^{(\ell-1)}}{\left(\sigma_a^2 + \sum_{k\neq i} \Phi_{a,k}^2 v_{k\to a}^{(\ell-1)}\right)}\right)^{-1} \\
&= \frac{\Phi_{a,i}^2 \tilde{v}_{i\to a}^{(\ell-1)} \tilde{m}_i^{(\ell-1)}}{\left(\sigma_a^2 + \sum_{k\neq i} \Phi_{a,k}^2 v_{k\to a}^{(\ell-1)}\right)} - \frac{\left(\hat{m}_{a\to i}^{(\ell-1)}/\hat{v}_{a\to i}^{(\ell-1)}\right)}{\left(\sum_{b=1}^{m} 1/\hat{v}_{b\to i}^{(\ell-1)}\right)}\left(1 + O\left(\frac{\ln n}{n}\right)\right),
\end{aligned}
$$

where the approximations hold because $\tilde{v}_i^{(\ell-1)}$ and $\sum_{k\neq i} \Phi_{a,k}^2 v_{k\to a}^{(\ell-1)}$ are uniformly bounded as $n \to \infty$ and $\Phi_{a,i}^2 = \Theta(\frac{\ln n}{n})$ uniformly in (a, i). Likewise, we can write

$$
\begin{aligned}
\mathrm{d}v &= \frac{1}{\sum_{b=1}^{m} 1/\hat{v}_{b\to i}^{(\ell-1)} - 1/\hat{v}_{a\to i}^{(\ell-1)}} - \frac{1}{\sum_{b=1}^{m} 1/\hat{v}_{b\to i}^{(\ell-1)}} \\
&= \frac{1/\hat{v}_{a\to i}^{(\ell-1)}}{\left(\sum_{b=1}^{m} 1/\hat{v}_{b\to i}^{(\ell-1)}\right)\left(\sum_{b=1}^{m} 1/\hat{v}_{b\to i}^{(\ell-1)} - 1/\hat{v}_{a\to i}^{(\ell-1)}\right)} \\
&= \frac{\Phi_{a,i}^2 \tilde{v}_i^{(\ell-1)} \tilde{v}_{i\to a}^{(\ell-1)}}{\left(\sigma_a^2 + \sum_{k\neq i} \Phi_{a,k}^2 v_{k\to a}^{(\ell-1)}\right)} = O\left(\frac{\ln n}{n}\right),
\end{aligned}
$$

where the approximation holds because $\tilde{v}_i^{(\ell-1)}$ and $\sum_{k\neq i} \Phi_{a,k}^2 v_{k\to a}^{(\ell-1)}$ are uniformly bounded as $n \to \infty$ and $\Phi_{a,i}^2 = \Theta(\frac{\ln n}{n})$ uniformly in (a, i). The proof for $v_{i\to a}^{(\ell)}$ is the same except that η is replaced by θ. □

Lemma 4.2. *The estimates*

$$
\begin{aligned}
m_i^{(\ell)} &= \eta\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right) + O\left(\frac{\ln n}{n}\right) \\
v_i^{(\ell)} &= \theta\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right) + O\left(\frac{\ln n}{n}\right)
\end{aligned}
$$

hold uniformly in i as $n \to \infty$.

Proof. Starting with the definition of $m_i^{(\ell)}$, we use Lemma 4.1 to write

$$\begin{aligned}
m_i^{(\ell)} &= \frac{1}{m}\sum_{a=1}^{m} m_{i\to a}^{(\ell)} \\
&= \frac{1}{m}\sum_{a=1}^{m} \eta\left(\tilde{m}_{i\to a}^{(\ell-1)}, \tilde{v}_{i\to a}^{(\ell-1)}\right) \\
&\approx \frac{1}{m}\sum_{a=1}^{m}\left[\eta\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right) + \eta^{(1,0)}\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right)\mathrm{d}m + \eta^{(0,1)}\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right)\mathrm{d}v\right] \\
&= \frac{1}{m}\sum_{a=1}^{m}\left[\eta\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right) + \eta^{(1,0)}\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right)\mathrm{d}m + \eta^{(0,1)}\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right)O\left(\frac{\ln n}{n}\right)\right] \\
&\approx \eta\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right) + \frac{1}{m}\sum_{a=1}^{m}\eta^{(1,0)}\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right)\left[\frac{\Phi_{a,i}^2 \tilde{v}_{i\to a}^{(\ell-1)}\tilde{m}_i^{(\ell-1)}}{\left(\sigma_a^2 + \sum_{k\neq i}\Phi_{a,k}^2 v_{k\to a}^{(\ell-1)}\right)} - \frac{\left(\hat{m}_{a\to i}^{(\ell-1)}/\hat{v}_{a\to i}^{(\ell-1)}\right)}{\left(\sum_{b=1}^{m} 1/\hat{v}_{b\to i}^{(\ell-1)}\right)}\right] \\
&= \eta\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right) + \frac{1}{m}\sum_{a=1}^{m}\frac{\theta\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right)}{\tilde{v}_i^{(\ell)}}\left[\frac{\Phi_{a,i}^2 \tilde{v}_{i\to a}^{(\ell-1)}\tilde{m}_i^{(\ell-1)}}{\left(\sigma_a^2 + \sum_{k\neq i}\Phi_{a,k}^2 v_{k\to a}^{(\ell-1)}\right)} - \frac{\left(\hat{m}_{a\to i}^{(\ell-1)}/\hat{v}_{a\to i}^{(\ell-1)}\right)}{\left(\sum_{b=1}^{m} 1/\hat{v}_{b\to i}^{(\ell-1)}\right)}\right] \\
&= \eta\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right) + \frac{1}{m}\frac{\theta\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right)}{\tilde{v}_i^{(\ell)}}\left[O\left(\frac{\ln n}{n}\right) - \sum_{a=1}^{m}\frac{\left(\hat{m}_{a\to i}^{(\ell-1)}/\hat{v}_{a\to i}^{(\ell-1)}\right)}{\left(\sum_{b=1}^{m} 1/\hat{v}_{b\to i}^{(\ell-1)}\right)}\right] \\
&\approx \eta\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right) - \frac{1}{m}\theta\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right)\sum_{a=1}^{m}\frac{\hat{m}_{a\to i}^{(\ell-1)}}{\hat{v}_{a\to i}^{(\ell-1)}} \\
&= \eta\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right) - \frac{1}{m}\theta\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right)\frac{\tilde{m}_i^{(\ell-1)}}{\tilde{v}_i^{(\ell-1)}} \\
&= \eta\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right) + O\left(\frac{\ln n}{n}\right),
\end{aligned}$$

where the final approximation holds because $\tilde{m}_i^{(\ell-1)} = O(\ln n)$ uniformly and $\tilde{v}_i^{(\ell-1)}$ and $\frac{1}{\tilde{v}_i^{(\ell-1)}}\theta\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right) =$

$O(1)$ uniformly. Likewise, we can write

$$
\begin{aligned}
v_i^{(\ell)} &= \frac{1}{m}\sum_{a=1}^{m} v_{i\to a}^{(\ell)} \\
&= \frac{1}{m}\sum_{a=1}^{m} \theta\left(\tilde{m}_{i\to a}^{(\ell-1)}, \tilde{v}_{i\to a}^{(\ell-1)}\right) \\
&\approx \frac{1}{m}\sum_{a=1}^{m}\left[\theta\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right) + \theta^{(1,0)}\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right)\mathrm{d}m + \theta^{(0,1)}\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right)\mathrm{d}v\right] \\
&= \frac{1}{m}\sum_{a=1}^{m}\left[\theta\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right) + \theta^{(1,0)}\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right)\mathrm{d}m + \theta^{(0,1)}\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right)O\left(\frac{\ln n}{n}\right)\right] \\
&\approx \theta\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right) + \frac{1}{m}\sum_{a=1}^{m}\theta^{(1,0)}\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right)\left[\frac{\Phi_{a,i}^2 \tilde{v}_{i\to a}^{(\ell-1)}\tilde{m}_i^{(\ell-1)}}{\left(\sigma_a^2 + \sum_{k\neq i}\Phi_{a,k}^2 v_{k\to a}^{(\ell-1)}\right)} - \frac{\left(\hat{m}_{a\to i}^{(\ell-1)}/\hat{v}_{a\to i}^{(\ell-1)}\right)}{\left(\sum_{b=1}^{m} 1/\hat{v}_{b\to i}^{(\ell-1)}\right)}\right] \\
&= \theta\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right) + \frac{1}{m}\sum_{a=1}^{m}\theta^{(1,0)}\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right)\left[\frac{\Phi_{a,i}^2 \tilde{v}_{i\to a}^{(\ell-1)}\tilde{m}_i^{(\ell-1)}}{\left(\sigma_a^2 + \sum_{k\neq i}\Phi_{a,k}^2 v_{k\to a}^{(\ell-1)}\right)} - \frac{\left(\hat{m}_{a\to i}^{(\ell-1)}/\hat{v}_{a\to i}^{(\ell-1)}\right)}{\left(\sum_{b=1}^{m} 1/\hat{v}_{b\to i}^{(\ell-1)}\right)}\right] \\
&= \theta\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right) + \frac{1}{m}\theta^{(1,0)}\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right)\left[O\left(\frac{\ln n}{n}\right) - \sum_{a=1}^{m}\frac{\left(\hat{m}_{a\to i}^{(\ell-1)}/\hat{v}_{a\to i}^{(\ell-1)}\right)}{\left(\sum_{b=1}^{m} 1/\hat{v}_{b\to i}^{(\ell-1)}\right)}\right] \\
&\approx \theta\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right) - \frac{1}{m}\theta\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right)\tilde{v}_i^{(\ell-1)}\sum_{a=1}^{m}\frac{\hat{m}_{a\to i}^{(\ell-1)}}{\hat{v}_{a\to i}^{(\ell-1)}} \\
&= \theta\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right) - \frac{1}{m}\theta\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right)\tilde{m}_i^{(\ell)} \\
&= \theta\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right) + O\left(\frac{\ln n}{n}\right),
\end{aligned}
$$

where the final approximation holds because $\tilde{m}_i^{(\ell-1)} = O(\ln n)$ uniformly and $\theta\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right) = O(1)$ uniformly. □

Now, we will write simplified recursions for $\omega_a^{(\ell)}$ and $V_a^{(\ell)}$ that are asymptotically exact. The goal is to keep all terms that do not vanish asymptotically. Naively, one might guess that $\omega_a^{(\ell)} \approx \sum_{j\neq i}\Phi_{a,i} m_{j\to a}^{(\ell)}$ is good enough to achieve this accuracy. But, this is not the case. Although the error is $O(\frac{1}{\sqrt{n}})$ for each a, the combination of $m = \alpha n$ values during the next stage leads to a constant size correction in the recursion for $\omega_a^{(\ell)}$. As long as $m_{k\to a}^{(\ell)}$ and $v_{k\to a}^{(\ell)}$ messages are bounded, (7) implies that $\hat{m}_{a\to i}^{(\ell)} = O(\sqrt{n})$ and $\hat{v}_{b\to i}^{(\ell)} = \Theta(n)$.

Starting from 12, we use Lemma 4.1 to write

$$
\begin{aligned}
\omega_a^{(\ell)} &= \sum_{i=1}^{n}\Phi_{a,i} m_{i\to a}^{(\ell)} \\
&= \sum_{i=1}^{n}\Phi_{a,i}\left(\eta\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right) + \eta^{(1,0)}\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right)\mathrm{d}m + \eta^{(0,1)}\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right)\mathrm{d}v\right) \\
&= \sum_{i=1}^{n}\Phi_{a,i}\eta\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right) + \Delta_m + \Delta_v,
\end{aligned}
$$

where Δ_m and Δ_v are the Taylor series corrections. The correction to $\omega_a^{(\ell)}$ due to $\mathrm{d}v$ is given by

$$\begin{aligned}
\Delta_v &\triangleq \sum_{i=1}^{n} \Phi_{a,i} \eta^{(0,1)} \left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right) \mathrm{d}v \\
&= \sum_{i=1}^{n} \Phi_{a,i} \eta^{(0,1)} \left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right) \frac{\Phi_{a,i}^2 \tilde{v}_i^{(\ell-1)} \tilde{v}_{i\to a}^{(\ell-1)}}{\left(\sigma_a^2 + \sum_{k\neq i} \Phi_{a,k}^2 v_{k\to a}^{(\ell-1)}\right)} \\
&\leq \left\|\eta^{(0,1)}\right\| \sum_{i=1}^{n} \Phi_{a,i}^3 \frac{\tilde{v}_i^{(\ell-1)} \tilde{v}_{i\to a}^{(\ell-1)}}{\left(\sigma_a^2 + \sum_{k\neq i} \Phi_{a,k}^2 v_{k\to a}^{(\ell-1)}\right)} \\
&= O\left(\frac{\ln n}{n}\right),
\end{aligned}$$

where the approximation holds because extreme-value theory implies that $\max_a \left|\sum_{i=1}^n \Phi_{a,i}^3\right| = O\left(\frac{\ln n}{n}\right)$. The correction to $\omega_a^{(\ell)}$ is given by

$$\begin{aligned}
\Delta_m &\triangleq \sum_{i=1}^{n} \Phi_{a,i} \eta^{(1,0)} \left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right) \mathrm{d}m \\
&= \sum_{i=1}^{n} \Phi_{a,i} \frac{\theta\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right)}{\tilde{v}_i^{(\ell-1)}} \left[\frac{\Phi_{a,i}^2 \tilde{v}_{i\to a}^{(\ell-1)} \tilde{m}_i^{(\ell-1)}}{\left(\sigma^2 + \sum_{k\neq i} \Phi_{a,k}^2 v_{k\to a}^{(\ell-1)}\right)} - \frac{\left(\hat{m}_{a\to i}^{(\ell-1)} / \hat{v}_{a\to i}^{(\ell-1)}\right)}{\left(\sum_{b=1}^{m} 1/\hat{v}_{b\to i}^{(\ell-1)}\right)}\right] \\
&= O(1) \sum_{i=1}^{n} \Phi_{a,i}^3 \tilde{m}_i^{(\ell-1)} - \sum_{i=1}^{n} \Phi_{a,i} \frac{\theta\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right)}{\tilde{v}_i^{(\ell-1)}} \tilde{v}_i^{(\ell-1)} \left(\hat{m}_{a\to i}^{(\ell-1)} / \hat{v}_{a\to i}^{(\ell-1)}\right) \\
&= O\left(\frac{(\ln n)^2}{n}\right) - \sum_{i=1}^{n} \Phi_{a,i} v_i^{(\ell)} \frac{\frac{1}{\Phi_{a,i}}\left(y_a - \sum_{k\neq i} \Phi_{a,k} m_{k\to a}^{(\ell-1)}\right)}{\frac{1}{\Phi_{a,i}^2}\left(\sigma^2 + \sum_{k\neq i} \Phi_{a,k}^2 v_{k\to a}^{(\ell-1)}\right)} \\
&\approx -\sum_{i=1}^{n} \Phi_{a,i}^2 v_i^{(\ell)} \frac{y_a - \omega_a^{(\ell-1)} + \Phi_{a,i} m_{i\to a}^{(\ell-1)}}{\sigma^2 + V_a^{(\ell-1)} - \Phi_{a,i}^2 v_{i\to a}^{(\ell-1)}} \\
&= -\sum_{i=1}^{n} \Phi_{a,i}^2 v_i^{(\ell)} \frac{y_a - \omega_a^{(\ell-1)} + \Phi_{a,i} m_{i\to a}^{(\ell-1)}}{\sigma^2 + V_a^{(\ell-1)}} \left(1 + \Theta\left(\frac{\ln n}{n}\right)\right) \\
&\approx -\sum_{i=1}^{n} \Phi_{a,i}^2 v_i^{(\ell)} \frac{y_a - \omega_a^{(\ell-1)} + \Phi_{a,i} m_{i\to a}^{(\ell-1)}}{\sigma^2 + V_a^{(\ell-1)}} \\
&\approx -\frac{y_a - \omega_a^{(\ell-1)}}{\sigma^2 + V_a^{(\ell-1)}} \sum_{i=1}^{n} \Phi_{a,i}^2 v_i^{(\ell)} - O(1) \sum_{i=1}^{n} \Phi_{a,i}^3 m_{i\to a}^{(\ell-1)} \\
&= -\frac{y_a - \omega_a^{(\ell-1)}}{\sigma^2 + V_a^{(\ell-1)}} V_a^{(\ell-1)} + O\left(\frac{(\ln n)^2}{n}\right),
\end{aligned}$$

where $\sum_{i=1}^n \Phi_{a,i}^3 \tilde{m}_i^{(\ell-1)} \approx \sum_{i=1}^n \Phi_{a,i}^3 m_{i\to a}^{(\ell-1)} = O(\frac{(\ln n)^2}{n})$ because the means are $O(\ln n)$ and extreme-value theory implies that $\max_a \left|\sum_{i=1}^n \Phi_{a,i}^3\right| = O\left(\frac{\ln n}{n}\right)$. This results in

$$\begin{aligned}
\omega_a^{(\ell)} &= \sum_{i=1}^{n} \Phi_{a,i} m_{i\to a}^{(\ell)} \\
&= \sum_{i=1}^{n} \Phi_{a,i} \eta\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right) - \frac{y_a - \omega_a^{(\ell-1)}}{\sigma_a^2 + V_a^{(\ell-1)}} V_a^{(\ell-1)} + O\left(\frac{(\ln n)^2}{n}\right).
\end{aligned}$$

Next, we observe that

$$
\begin{aligned}
V_a^{(\ell)} &= \sum_{i=1}^{n} \Phi_{a,i}^2 v_{i\to a}^{(\ell)} \\
&= \sum_{i=1}^{n} \Phi_{a,i}^2 \theta\left(\tilde{m}_{a\to i}^{(\ell-1)}, \tilde{v}_{a\to i}^{(\ell-1)}\right) \\
&= \sum_{i=1}^{n} \Phi_{a,i}^2 \theta\left(\frac{\sum_{b=1}^{m} \hat{m}_{b\to i}^{(\ell-1)}/\hat{v}_{b\to i}^{(\ell-1)} - \hat{m}_{a\to i}^{(\ell-1)}/\hat{v}_{a\to i}^{(\ell-1)}}{\sum_{b=1}^{m} 1/\hat{v}_{b\to i}^{(\ell-1)} - 1/\hat{v}_{a\to i}^{(\ell-1)}}, \frac{1}{\sum_{b=1}^{m} 1/\hat{v}_{b\to i}^{(\ell-1)} - 1/\hat{v}_{a\to i}^{(\ell-1)}}\right) \\
&\approx \sum_{i=1}^{n} \Phi_{a,i}^2 \left(\theta\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right) + \theta^{(1,0)}\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right) \mathrm{d}m + \theta^{(0,1)}\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right) \mathrm{d}v\right) \\
&= \sum_{i=1}^{n} \Phi_{a,i}^2 \left(\theta\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right) + \theta^{(1,0)}\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right) \mathrm{d}m + \theta^{(0,1)}\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right) O\left(\frac{\ln n}{n}\right)\right) \\
&\approx \sum_{i=1}^{n} \Phi_{a,i}^2 v_i^{(\ell)} + \sum_{i=1}^{n} \Phi_{a,i}^2 \theta^{(1,0)}\left(\tilde{m}_i^{(\ell-1)}, \tilde{v}_i^{(\ell-1)}\right) \left(\frac{\Phi_{a,i}^2 \tilde{v}_{i\to a}^{(\ell-1)} \tilde{m}_i^{(\ell-1)}}{\left(\sigma_a^2 + \sum_{k\neq i} \Phi_{a,k}^2 v_{k\to a}^{(\ell-1)}\right)} - \frac{\left(\hat{m}_{a\to i}^{(\ell-1)}/\hat{v}_{a\to i}^{(\ell-1)}\right)}{\left(\sum_{b=1}^{m} 1/\hat{v}_{b\to i}^{(\ell-1)}\right)}\right) + O\left(\frac{(\ln n)^2}{n}\right) \\
&\approx \sum_{i=1}^{n} \Phi_{a,i}^2 v_i^{(\ell)} + \left\|\theta^{(1,0)}\right\| \sum_{i=1}^{n} \Phi_{a,i}^2 \frac{\Phi_{a,i}^2 \tilde{v}_{i\to a}^{(\ell-1)} \tilde{m}_i^{(\ell-1)}}{\left(\sigma_a^2 + \sum_{k\neq i} \Phi_{a,k}^2 v_{k\to a}^{(\ell-1)}\right)} - \left\|\theta^{(1,0)}\right\| \sum_{i=1}^{n} \Phi_{a,i}^3 v_i^{(\ell-1)} \frac{y_a - \omega_a^{(\ell-1)} + \Phi_{a,i} m_{i\to a}^{(\ell-1)}}{\sigma_a^2 + V_a^{(\ell-1)} - \Phi_{a,i}^2 v_{i\to a}^{(\ell-1)}} \\
&= \sum_{i=1}^{n} \Phi_{a,i}^2 v_i^{(\ell)} + O\left(\frac{(\ln n)^2}{n}\right) + O\left(\frac{(\ln n)^2}{\sqrt{n}}\right),
\end{aligned}
$$

where $\sum_{i=1}^{n} \Phi_{a,i}^4 \tilde{m}_i^{(\ell-1)} = O\left(\frac{\ln n}{n}\right)$ because the $\tilde{m}_i^{(\ell-1)} = O(\ln n)$ uniformly and extreme-value theory implies that $\max_a \left|\sum_{i=1}^{n} \Phi_{a,i}^3\right| = O\left(\frac{\ln n}{n}\right)$. This formula also implies that $V_a^{(\ell)}$ is essentially independent of a. If we assume that a CLT holds for $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{\Phi}_{a,i}^2 v_i^{(\ell)}$ (i.e., without any real proof), then we can write

$$
\sum_{i=1}^{n} \Phi_{a,i}^2 v_i^{(\ell)} \approx \frac{1}{n} \sum_{i=1}^{n} v_i^{(\ell)} + O\left(\frac{\ln n}{\sqrt{n}}\right).
$$

Since the RHS is essentially independent of a, so is the LHS. Thus, we define

$$
\overline{V}^{(\ell)} \triangleq \frac{1}{n} \sum_{i=1}^{n} v_i^{(\ell)}
$$

and make the approximation $V_a^{(\ell)} \approx \overline{V}^{(\ell)}$. The same conclusion can be reached deterministically (i.e., without resorting unproven assumptions) if the measurement matrix is Rademacher. Based on this result, we will replace all occurrences of $V_a^{(\ell)}$ with $\overline{V}^{(\ell)}$ in the final algorithm. Next, we use (7) to write

$$\begin{aligned}
\tilde{v}_i^{(\ell)} &= \frac{1}{\sum_{a=1}^m 1/\hat{v}_{a\to i}^{(\ell)}} \\
&= \left[\sum_{a=1}^m \frac{\Phi_{a,i}^2}{\sigma^2 + V_a^{(\ell)} - \Phi_{a,i}^2 v_{i\to a}^{(\ell)}}\right]^{-1} \\
&\approx \left[\left(\sum_{a=1}^m \frac{\Phi_{a,i}^2}{\sigma^2 + V_a^{(\ell)}} \cdot \frac{1}{1 - \Phi_{a,i}^2 v_{i\to a}^{(\ell)}/(\sigma_a^2 + V_a^{(\ell)})}\right)\right]^{-1} \\
&\overset{(a)}{=} \left[\sum_{a=1}^m \frac{\Phi_{a,i}^2}{\sigma^2 + V_a^{(\ell)}} \left(1 + O\left(\frac{\ln n}{n}\right)\right)\right]^{-1} \\
&\overset{(b)}{=} \left[\sum_{a=1}^m \frac{\Phi_{a,i}^2}{\sigma^2 + \overline{V}^{(\ell)} + O\left(\frac{\ln n}{\sqrt{n}}\right)} \left(1 + O\left(\frac{\ln n}{n}\right)\right)\right]^{-1} \\
&= \left[\sum_{a=1}^m \frac{\Phi_{a,i}^2}{\sigma^2 + \overline{V}^{(\ell)}} \left(1 + O\left(\frac{\ln n}{\sqrt{n}}\right)\right)\left(1 + O\left(\frac{\ln n}{n}\right)\right)\right]^{-1} \\
&\overset{(c)}{=} \left[\frac{\alpha}{\sigma^2 + \overline{V}^{(\ell)}} \left(1 + O\left(\frac{\ln n}{n}\right)\right)\left(1 + O\left(\frac{\ln n}{\sqrt{n}}\right)\right)\left(1 + O\left(\frac{\ln n}{n}\right)\right)\right]^{-1} \\
&= \frac{\sigma^2 + \overline{V}^{(\ell)}}{\alpha} \left(1 + O\left(\frac{\ln n}{\sqrt{n}}\right)\right),
\end{aligned}$$

where (a) holds because $\Phi_{a,i}^2 = \Theta(\frac{\ln n}{n})$ uniformly in (a,i), (b) relies on the assumption that $V_a^{(\ell-1)}$ concentrates around a deterministic quantity that doesn't depend on a, and (c) follows from $\sum_{a=1}^m \Phi_{a,i}^2 = \alpha\left(1 + O\left(\frac{\ln n}{n}\right)\right)$ uniformly in i.

Using an approach similar to the one for $\tilde{v}_i^{(\ell)}$we now analyze $\tilde{m}_i^{(\ell)}$. In particular, we write

$$
\begin{aligned}
\frac{\tilde{m}_i^{(\ell)}}{\tilde{v}_i^{(\ell)}} &\triangleq \sum_{a=1}^{m} \frac{\hat{m}_{a\to i}^{(\ell)}}{\hat{v}_{a\to i}^{(\ell)}} \\
&= \sum_{a=1}^{m} \frac{\Phi_{a,i}(y_a - \omega_a^{(\ell)} + \Phi_{a,i} m_{i\to a}^{(\ell)})}{\sigma^2 + V_a^{(\ell)} - \Phi_{a,i}^2 v_{i\to a}^{(\ell)}} \\
&= \sum_{a=1}^{m} \frac{\Phi_{a,i}(y_a - \omega_a^{(\ell)} + \Phi_{a,i} m_{i\to a}^{(\ell)})}{\sigma^2 + V_a^{(\ell)}} \left(1 + \frac{\Phi_{a,i}^2 v_{i\to a}^{(\ell)}}{\sigma_a^2 + V_a^{(\ell)}}\right)^{-1} \\
&= \sum_{a=1}^{m} \frac{\Phi_{a,i}(y_a - \omega_a^{(\ell)} + \Phi_{a,i} m_{i\to a}^{(\ell)})}{\sigma^2 + V_a^{(\ell)}} \left(1 + O\left(\frac{\ln n}{n}\right)\right)^{-1} \\
&\approx \sum_{a=1}^{m} \frac{\Phi_{a,i}(y_a - \omega_a^{(\ell)})}{\sigma^2 + V_a^{(\ell)}} + \sum_{a=1}^{m} \frac{\Phi_{a,i}^2 m_{i\to a}^{(\ell)}}{\sigma^2 + V_a^{(\ell)}} \\
&\overset{(a)}{=} \sum_{a=1}^{m} \frac{\Phi_{a,i}^2 m_{i\to a}^{(\ell)}}{\sigma^2 + \overline{V}^{(\ell)} + O\left(\frac{\ln n}{\sqrt{n}}\right)} + \sum_{a=1}^{m} \frac{\Phi_{a,i}(y_a - \omega_a^{(\ell)})}{\sigma^2 + \overline{V}^{(\ell)} + O\left(\frac{\ln n}{\sqrt{n}}\right)} \\
&= \sum_{a=1}^{m} \frac{\Phi_{a,i}^2 m_{i\to a}^{(\ell)}}{\sigma^2 + \overline{V}^{(\ell)}} \left(1 + O\left(\frac{\ln n}{\sqrt{n}}\right)\right) + \sum_{a=1}^{m} \frac{\Phi_{a,i}(y_a - \omega_a^{(\ell)})}{\sigma^2 + \overline{V}^{(\ell)}} \left(1 + O\left(\frac{\ln n}{\sqrt{n}}\right)\right) \\
&= \left(\frac{1}{\sigma^2 + \overline{V}^{(\ell)}} \sum_{a=1}^{m} \Phi_{a,i}^2 m_{i\to a}^{(\ell)} + \sum_{a=1}^{m} \frac{\Phi_{a,i}(y_a - \omega_a^{(\ell)})}{\sigma^2 + \overline{V}^{(\ell)}}\right) \left(1 + O\left(\frac{\ln n}{\sqrt{n}}\right)\right) \\
&\overset{(b)}{=} \left(\frac{1}{\sigma^2 + \overline{V}^{(\ell)}} \left(\frac{\alpha}{m} \sum_{a=1}^{m} m_{i\to a}^{(\ell)}\right) \left(1 + O\left(\frac{\ln n}{\sqrt{n}}\right)\right) + \sum_{a=1}^{m} \frac{\Phi_{a,i}(y_a - \omega_a^{(\ell)})}{\sigma^2 + \overline{V}^{(\ell)}}\right) \left(1 + O\left(\frac{\ln n}{\sqrt{n}}\right)\right) \\
&= \left(\frac{\alpha}{\sigma^2 + \overline{V}^{(\ell)}} \tilde{m}_i^{(\ell)} + \sum_{a=1}^{m} \frac{\Phi_{a,i}(y_a - \omega_a^{(\ell-1)})}{\sigma^2 + \overline{V}^{(\ell-1)}}\right) \left(1 + O\left(\frac{\ln n}{\sqrt{n}}\right)\right),
\end{aligned}
$$

where (a) relies on the assumption that $V_a^{(\ell-1)}$ concentrates around a deterministic quantity that doesn't depend on a and (b) relies on the assumption that $\sum_{a=1}^{m} \Phi_{a,i}^2 m_{i\to a}^{(\ell)} = \left(1 + O\left(\frac{\ln n}{\sqrt{n}}\right)\right) \frac{\alpha}{m} \sum_{a=1}^{m} m_{i\to a}^{(\ell)}$. Both assumptions follow from arguments above if Φ is Rademacher. Otherwise, if Φ is Gaussian, both assumptions follow from the rigorous state evolution analysis in [28].

Together, these equations define an approximation of RBP that has greatly reduced storage requirements. For the case where Φ is Rademacher, the above derivation shows it is asymptotically exact as well. For Gaussian $\tilde{\Phi}_{a,i}$, the state evolution analysis in the next section also implies that it is asymptotically exact.

The TAP form of the Relaxed Belief Propagation Algorithm:

$$\begin{aligned}
m_i^{(0)} &= \int f_{X_i}(x_i) x_i \, \mathrm{d}x \\
v_i^{(0)} &= \int f_{X_i}(x_i) \left(x_i - m_i^{(0)} \right)^2 \mathrm{d}x \\
\overline{V}^{(\ell)} &= \frac{1}{n} \sum_{i=1}^{n} v_i^{(\ell)} \\
\omega_a^{(\ell)} &= \sum_{i=1}^{n} \Phi_{a,i} m_i^{(\ell)} - \frac{y_a - \omega_a^{(\ell-1)}}{\sigma^2 + \overline{V}^{(\ell-1)}} \overline{V}^{(\ell-1)} \\
\tilde{v}^{(\ell)} &= \frac{\sigma^2 + \overline{V}^{(\ell-1)}}{\alpha} \\
\tilde{m}_i^{(\ell+1)} &= m_i^{(\ell)} + \frac{1}{\alpha} \sum_{a=1}^{m} \Phi_{a,i} \left(y_a - \omega_a^{(\ell-1)} \right) \\
m_i^{(\ell+1)} &= \eta \left(\tilde{m}_i^{(\ell)}, \tilde{v}^{(\ell)} \right) \\
v_i^{(\ell+1)} &= \theta \left(\tilde{m}_i^{(\ell)}, \tilde{v}^{(\ell)} \right)
\end{aligned}$$

5 Relationship to AMP

5.1 State Evolution

For sequences of graphs with tree-like neighborhoods, density evolution provides an asymptotically exact analysis of message-passing inference. For compressed sensing, this technique can be used to analyze RBP in the *large-sparse-limit*, where the node degrees grow slowly with the block length (e.g., like $\ln n$) [20]. This special limit is required because the growth of the node degrees is required for the central limit theorem but the degrees must grow slow enough that the sequence of graphs has tree-like neighborhoods.

Density evolution does not extend, however, to general graphs whose neighborhoods are not tree-like. For the dense factor graph implied by an i.i.d. Gaussian Φ matrix, it is somewhat surprising that one can rigorously analyze the average the evolution of message-passing decoding. The resulting analysis is called state evolution and the proof is quite different from standard density evolution [28]. From the previous section, it is clear that $\overline{V}^{(\ell)}$ is the average variance of the element-wise variable estimates. State evolution also proves that the quantity $V_a^{(\ell)} = \overline{V}^{(\ell)}$ is independent of a. If the true prior is used for reconstruction, then it also shows that $\overline{V}^{(\ell)}$ is also equal to the normalized mean-squared-error signal of signal estimates. For this quantity, state evolution predicts that $\overline{V}^{(\ell+1)} = \kappa \left(\overline{V}^{(\ell)} \right)$, where

$$\kappa(v) = \int f_X(x) \int \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \theta \left(x + z \sqrt{\frac{v + \sigma^2}{\alpha}}, \frac{v + \sigma^2}{\alpha} \right) \mathrm{d}z \mathrm{d}x.$$

This equation can be seen to start with the message variance v. Then, the variance is increased to $\frac{v+\sigma^2}{\alpha}$ by the bit-to-check and check-to-bit message passing. Finally, a random observation $Y = X + Z\sqrt{\frac{v+\sigma^2}{\alpha}}$, where Z is a standard Gaussian variable, is combined with the prior to compute the new mean-squared-error $\theta \left(Y, \sqrt{\frac{v+\sigma^2}{\alpha}} \right)$ of the signal estimates. The function $\kappa(v)$ gives the expected value of the new variance w.r.t. to the random observation of the signal in Gaussian noise.

The formula for $\kappa(v)$ can be expressed in a variety of equivalent ways:

$$
\begin{aligned}
\kappa(v) &= \mathrm{Var}\left(X \middle| X + Z\sqrt{\frac{v+\sigma^2}{\alpha}}\right) \\
&= \mathbb{E}\left[\left(X - \mathbb{E}\left[X \middle| X + Z\sqrt{\frac{v+\sigma^2}{\alpha}}\right]\right)^2\right] \\
&= \mathbb{E}\left[\left(X - \eta\left(X + Z\sqrt{\frac{v+\sigma^2}{\alpha}}, \frac{v+\sigma^2}{\alpha}\right)\right)^2\right] \\
&= \mathrm{MSE}\left(X \middle| X + Z\sqrt{\frac{v+\sigma^2}{\alpha}}\right),
\end{aligned}
$$

where $\mathrm{MSE}(X|Y)$ stands for the mean-squared error of the optimal estimator of X given Y.

5.2 AMP

In our notation, the AMP decoding evolution is given by

$$
\begin{aligned}
\underline{m}^{(0)} &= \underline{0} \\
\underline{\omega}^{(\ell)} &= \underline{y} - \Phi\underline{m}^{(\ell)} + \frac{k}{\alpha n}\underline{\omega}^{(\ell-1)} \\
\underline{m}^{(\ell+1)} &= \chi\left(\Phi^\top\underline{\omega}^{(\ell)} + \underline{m}^{(\ell)}; \theta^{(\ell)}\right),
\end{aligned}
$$

where $\chi(m;\theta) = \mathrm{sgn}(m)(|m|-\theta)^+$ is the soft-thresholding function and the parameter $\theta^{(\ell)}$ is chosen[3] so that $\left|\left\{i \in [n] \,|\, m_i^{(\ell+1)} \neq 0\right\}\right| = k$ for some fixed $k \in \{1, 2, \ldots, \alpha n - 1\}$.

Lemma 5.1. *A tuple $(\underline{m}, \underline{\omega}, \theta)$ is a solution to the LASSO problem (1) iff it is a fixed point of the AMP algorithm.*

Proof. If it is a fixed point, the fixed point equations are

$$
\begin{aligned}
\underline{\omega} &= \underline{y} - \Phi\underline{m} + \frac{k}{\alpha n}\underline{\omega} \\
\underline{m} &= \chi\left(\Phi^\top\underline{\omega} + \underline{m}; \theta\right).
\end{aligned}
$$

Solving the first equation for $\underline{\omega}$ and substituting into the second equation gives

$$
\underline{m} = \chi\left(\Phi^\top \frac{1}{1-\frac{k}{\alpha n}}\left(\underline{y} - \Phi\underline{m}\right) + \underline{m}; \theta\right)
$$

or, using $\chi(\alpha m; \alpha\theta) = \alpha\chi(m;\theta)$,

$$
\left(1 - \frac{k}{\alpha n}\right)\underline{m} = \chi\left(\Phi^\top\left(\underline{y} - \Phi\underline{m}\right) + \left(1 - \frac{k}{\alpha n}\right)\underline{m}; \theta\left(1 - \frac{k}{\alpha n}\right)\right).
$$

Letting $\lambda = \theta\left(1 - \frac{k}{\alpha n}\right)$, we see this is equivalent to

$$
\left[\Phi^\top\left(\underline{y} - \Phi\underline{m}\right)\right]_i \in \begin{cases} [-\lambda, \lambda] & \text{if } m_i = 0 \\ \{\lambda\} & \text{if } m_i > 0 \\ \{-\lambda\} & \text{if } m_i < 0, \end{cases}
$$

which is also the necessary and sufficient condition for a LASSO minimizer (2). Likewise, for a solution to the LASSO problem with k non-zero entries, the steps holds in reverse and allow one to construct a fixed point for AMP. □

[3]There are a variety of closely related rules that can be used here and we have chosen a particularly simple one.

5.3 Minimax Denoising and Soft-Thresholding

The factor-graph framework reduces the CS reconstruction problem to interference reduction followed by scalar estimation. If the prior is not known, one can emphasize robustness and solve the minimax problem to find the worst case prior in some class of distributions. For CS, a natural choice is the class of signal distributions, $\mathcal{F}_\epsilon$, such that $\Pr(X=0) \geq 1-\epsilon$. The minimax estimator for this class has been analyzed but the results do not have a simple expression. These ideas led [8] to use the soft-thresholding function $\chi(m;\theta) = \operatorname{sgn}(m)(|m|-\theta)^+$ in their derivation of AMP.

The MSE of soft-thresholding is given by

$$M(\theta, v, f_X) \triangleq \mathbb{E}\left[\left(X - \chi\left(X + Z\sqrt{v};\theta\right)\right)^2\right].$$

To understand the worst-case prior for this choice, let $\tilde{M}(\epsilon,\theta,v) \triangleq \sup_{f_X\in\mathcal{F}_\epsilon} M(\theta,v,f_X)$ and observe that

$$\begin{aligned}
\inf_{\theta} \tilde{M}(\epsilon,\theta,v) &= \inf_{\theta\geq 0} \sup_{f_X\in\mathcal{F}_\epsilon} M(\theta,v,f_X) \\
&= \inf_{\theta\geq 0} \sup_{f_X\in\mathcal{F}_\epsilon} M\left(\theta\sqrt{v},v,f_X\right) \\
&= \inf_{\theta\geq 0} \sup_{f_X\in\mathcal{F}_\epsilon} \mathbb{E}\left[\left(\sqrt{v}X - \chi\left(\sqrt{v}X + Z\sqrt{v};\theta\sqrt{v}\right)\right)^2\right] \\
&= v \inf_{\theta\geq 0} \sup_{f_X\in\mathcal{F}_\epsilon} \mathbb{E}\left[\left(X - \chi\left(X + Z;\theta\right)\right)^2\right] \\
&= v \inf_{\theta\geq 0} \tilde{M}(\epsilon,\theta,1).
\end{aligned}$$

This is automatically an upper bound on the MSE of the true minimax estimator (i.e., without assuming soft-thresholding). Also, it turns out that the soft-thresholding minimax problem can be solved [29] because the worst case distribution is discrete and known to be

$$\Pr(X=x) = \begin{cases} 1-\epsilon & \text{if } x=0 \\ \epsilon/2 & \text{if } x=\infty \\ \epsilon/2 & \text{if } x=-\infty. \end{cases}$$

Using this, it is easy to verify that

$$\tilde{M}(\epsilon,\theta,1) = \epsilon(1+\theta^2) + (1-\epsilon)\left[2(1+\theta^2)\int_\theta^\infty \phi(z)\mathrm{d}z - 2\theta\phi(\theta)\right].$$

Let $\tilde{M}(\epsilon) = \inf_{\theta\geq 0}\tilde{M}(\epsilon,\theta,1)$ and observe that state-evolution recursion for soft-thresholding with the worst-case prior satisfies

$$V^{(\ell+1)} = \tilde{M}(\epsilon)\frac{\sigma^2+V^{(\ell)}}{\alpha}$$

and converges to the fixed point

$$V^* = \frac{\tilde{M}(\epsilon)\sigma^2}{\alpha-\tilde{M}(\epsilon)}$$

as long as $\tilde{M}(\epsilon) < \alpha$.

A less pessimistic analysis results in the state-evolution recursion

$$\begin{aligned}
V^{(\ell+1)} &= \kappa\left(\frac{\sigma^2+V^{(\ell)}}{\alpha};\lambda^*\left(\frac{\sigma^2+V^{(\ell)}}{\alpha}\right)\right) \\
\kappa(v;\lambda) &= \mathbb{E}\left[\left(X - \chi\left(X + Z\sqrt{\frac{v+\sigma^2}{\alpha}},\lambda\right)\right)^2\right],
\end{aligned} \tag{17}$$

where $\lambda^*(v) = \arg\min_\lambda \kappa(v;\lambda)$ minimizes this variance. Of course, this recursion depends explicitly on the prior.

Remark 5.2. It is important to note that the matrix scaling $\frac{1}{\sqrt{n}}$ used in this handout matches [27] and is different from the $\frac{1}{\sqrt{m}}$ used in [8]. The former leads to roughly unit-norm rows while the latter leads to roughly unit-norm columns. For this reason, the state evolution equation (17) does not match the analogous equation in [8]. To recover their equation, one can simply replace our noise variance σ^2 by the quantity $\alpha\sigma^2$.

6 Closed-Form Expressions for $\eta(m, v)$ and $\theta(m, v)$

6.1 Bernoulli-Gauss Prior

For a Bernoulli-Gaussian prior, $f_X(x) = (1-p)\delta(x) + \frac{p}{\sigma_1}\phi\left(\frac{x-\mu_1}{\sigma_1}\right)$, the functions are given by

$$\eta(m,v) = \left(\frac{m\sigma_1^2 + \mu_1 v}{v+\sigma_1^2}\right)\left(1 + \frac{1-p}{p}\sqrt{\frac{v+\sigma_1^2}{v}}\exp\left(\frac{(m-\mu_1)^2}{2(v+\sigma_1^2)} - \frac{m^2}{2v}\right)\right)^{-1}$$

$$\beta(m,v) = \left(\frac{\sigma_1^4\left(m^2+v\right) + v\sigma_1^2(2m\mu_1+v) + \mu_1^2 v^2}{(v+\sigma_1^2)^2}\right)\left(1 + \frac{1-p}{p}\sqrt{\frac{v+\sigma_1^2}{v}}\exp\left(\frac{(m-\mu_1)^2}{2(v+\sigma_1^2)} - \frac{m^2}{2v}\right)\right)^{-1}$$

$$\theta(m,v) = \beta(m,v) - (\eta(m,v))^2.$$

Mathematica was used to assist with the integration to output the TEX code but these integrals can also be computed easily by hand.

6.2 Laplacian Prior

For a Laplacian prior, $f_X(x) = \frac{1}{\lambda}e^{-\lambda|x|}$, the functions are given by

$$\eta(m,v) = \frac{e^{-\frac{m^2}{2v}}\left(e^{\frac{(m-\lambda v)^2}{2v}}(m-\lambda v)\mathrm{erfc}\left(\frac{\lambda v-m}{\sqrt{2v}}\right) + e^{\frac{(\lambda v+m)^2}{2v}}(m+\lambda v)\mathrm{erfc}\left(\frac{\lambda v+m}{\sqrt{2v}}\right)\right)}{e^{\frac{1}{2}\lambda(\lambda v-2m)}\mathrm{erfc}\left(\frac{\lambda v-m}{\sqrt{2v}}\right) + e^{\frac{1}{2}\lambda(\lambda v+2m)}\mathrm{erfc}\left(\frac{\lambda v+m}{\sqrt{2v}}\right)}$$

$$\beta(m,v) = \frac{e^{-\frac{m^2}{2v}}\left(e^{\frac{(m-\lambda v)^2}{2v}}\left(\left((m-\lambda v)^2+v\right)\mathrm{erfc}\left(\frac{\lambda v-m}{\sqrt{2v}}\right) - e^{2\lambda m}\left((m+\lambda v)^2+v\right)\mathrm{erf}\left(\frac{\lambda v+m}{\sqrt{2v}}\right)\right) + e^{\frac{(\lambda v+m)^2}{2v}}\left((m+\lambda v)^2+v\right) - \frac{4\lambda v^{3/2}}{\sqrt{2\pi}}\right)}{e^{\frac{1}{2}\lambda(\lambda v-2m)}\mathrm{erfc}\left(\frac{\lambda v-m}{\sqrt{2v}}\right) + e^{\frac{1}{2}\lambda(\lambda v+2m)}\mathrm{erfc}\left(\frac{\lambda v+m}{\sqrt{2v}}\right)}$$

$$\theta(m,v) = \beta(m,v) - (\eta(m,v))^2.$$

By treating $x \geq 0$ and $x \leq 0$ separately, one gets integrals that are tedious but straightforward. Mathematica was used to assist with the integration to output the TEX code.

References

[1] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comp.*, vol. 20, no. 1, pp. 33–61, 1998.

[2] Y. Bresler, M. Gastpar, and R. Venkataramani, "Image compression on-the-fly by universal sampling in Fourier imaging systems," in *Proc. IEEE Inform. Theory Workshop*, pp. 48–48, 1999.

[3] M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation," *IEEE Signal Processing Mag.*, vol. 50, no. 6, pp. 1417–1428, 2002.

[4] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[5] E. Candes, 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, 2006.

[6] R. Tibshirani, “Regression shrinkage and selection via the lasso,” *J. Royal Stat. Soc. Ser. B*, pp. 267–288, 1996.

[7] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, “Least angle regression,” *Ann. Stats.*, vol. 32, no. 2, pp. 407–499, 2004.

[8] D. L. Donoho, A. Maleki, and A. Montanari, “Message-passing algorithms for compressed sensing,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 106, no. 45, pp. 18914–18919, 2009.

[9] D. L. Donoho, A. Maleki, and A. Montanari, “Message passing algorithms for compressed sensing: I. Motivation and construction,” in *Proc. IEEE Inform. Theory Workshop*, (Cairo, Egypt), Jan. 2010.

[10] Y. Wu and S. Verdú, “Rényi information dimension: Fundamental limits of almost lossless analog compression,” *IEEE Trans. Inform. Theory*, vol. 56, no. 8, pp. 3721–3748, 2010.

[11] Y. Wu and S. Verdú, “Optimal phase transitions in compressed sensing.” Arxiv preprint arXiv:1111.6822, Nov. 2011.

[12] W. Xu and B. Hassibi, “Efficient compressive sensing with deterministic guarantees using expander graphs,” in *Proc. IEEE Inform. Theory Workshop*, (Lake Tahoe, CA, USA), pp. 414–419, Sept. 2007.

[13] F. Zhang and H. D. Pfister, “On the iterative decoding of high rate LDPC codes with applications in compressed sensing,” in *Proc. 46th Annual Allerton Conf. on Commun., Control, and Comp.*, (Monticello, IL, USA), Sept. 2008.

[14] F. Zhang and H. D. Pfister, “Verification decoding of high-rate LDPC codes with applications in compressed sensing,” *IEEE Trans. Inform. Theory*, vol. 58, pp. 5042–5048, Aug. 2012.

[15] F. Krzakala, M. Mézard, F. Sausset, Y. Sun, and L. Zdeborová, “Statistical physics-based reconstruction in compressed sensing.” Arxiv preprint arXiv:1109.4424, Sept. 2011.

[16] D. Donoho, A. Javanmard, and A. Montanari, “Information-theoretically optimal compressed sensing via spatial coupling and approximate message passing.” Arxiv preprint arXiv:1112.0708, Dec. 2011.

[17] S. Kudekar and H. D. Pfister, “The effect of spatial coupling on compressive sensing,” in *Proc. Annual Allerton Conf. on Commun., Control, and Comp.*, (Monticello, IL, USA), pp. 347–353, Oct. 2010.

[18] S. Sarvotham, D. Baron, and R. G. Baraniuk, “Compressed sensing reconstruction via belief propagation,” Tech. Rep. ECE-06-01, Rice University, July 2006.

[19] 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, 2010.

[20] 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), pp. 194–198, Oct. 2006.

[21] T. Tanaka, “A statistical-mechanics approach to large-system analysis of CDMA multiuser detectors,” *IEEE Trans. Inform. Theory*, vol. 48, no. 11, pp. 2888–2910, 2002.

[22] Y. Kabashima, “A CDMA multiuser detection algorithm on the basis of belief propagation,” *J. Phys. A*, vol. 36, no. 43, pp. 11111–11121, 2003.

[23] A. Montanari and D. Tse, “Analysis of belief propagation for non-linear problems: The example of CDMA (or: How to prove Tanaka’s formula),” in *Proc. IEEE Inform. Theory Workshop*, (Punta del Este, Uruguay), pp. 160–164, IEEE, 2006.

[24] M. Bayati and A. Montanari, “The dynamics of message passing on dense graphs, with applications to compressed sensing.” [Online]. Available: http://arxiv.org/abs/1001.3448, 2010.

[25] L. He and L. Carin, "Exploiting structure in wavelet-based Bayesian compressive sensing," *IEEE Trans. Signal Process.*, vol. 57, no. 9, pp. 3488–3497, 2009.

[26] S. Som, L. C. Potter, and P. Schniter, "Compressive imaging using approximate message passing and a Markov-tree prior," in *Proc. Asilomar Conf. on Signals, Systems & Computers*, pp. 243–247, 2010.

[27] F. Krzakala, M. Mézard, F. Sausset, Y. Sun, and L. Zdeborová, "Probabilistic reconstruction in compressed sensing: algorithms, phase diagrams, and threshold achieving matrices," *Journal of Statistical Mechanics: Theory and Experiment*, no. 08, p. P08009, 2012.

[28] M. Bayati and A. Montanari, "The dynamics of message passing on dense graphs, with applications to compressed sensing," *IEEE Trans. Inform. Theory*, vol. 57, pp. 764–785, Feb. 2011.

[29] D. L. Donoho and I. M. Johnstone, "Minimax risk over ℓ_p-balls for ℓ_q-error," *Probability Theory and Related Fields*, vol. 99, no. 2, pp. 277–303, 1994.