Approximate Message Passing with Consistent Parameter Estimation and Applications to Sparse Learning

Ulugbek S. Kamilov
EPFL
ulugbek.kamilov@epfl.ch

Sundeep Rangan
Polytechnic Institute of New York University
srangan@poly.edu

Alyson K. Fletcher
University of California, Santa Cruz
afletcher@soe.ucsc.edu

Michael Unser
EPFL
michael.unser@epfl.ch

Abstract

We consider the estimation of an i.i.d. vector \( x \in \mathbb{R}^n \) from measurements \( y \in \mathbb{R}^m \) obtained by a general cascade model consisting of a known linear transform followed by a probabilistic componentwise (possibly nonlinear) measurement channel. We present a method, called adaptive generalized approximate message passing (Adaptive GAMP), that enables joint learning of the statistics of the prior and measurement channel along with estimation of the unknown vector \( x \). Our method can be applied to a large class of learning problems including the learning of sparse priors in compressed sensing or identification of linear-nonlinear cascade models in dynamical systems and neural spiking processes. We prove that for large i.i.d. Gaussian transform matrices the asymptotic componentwise behavior of the adaptive GAMP algorithm is predicted by a simple set of scalar state evolution equations. This analysis shows that the adaptive GAMP method can yield asymptotically consistent parameter estimates, which implies that the algorithm achieves a reconstruction quality equivalent to the oracle algorithm that knows the correct parameter values. The adaptive GAMP methodology thus provides a systematic, general and computationally efficient method applicable to a large range of complex linear-nonlinear models with provable guarantees.

1 Introduction

Consider the estimation of a random vector \( x \in \mathbb{R}^n \) from a measurement vector \( y \in \mathbb{R}^m \). As illustrated in Figure 1, the vector \( x \), which is assumed to have i.i.d. components \( x_i \sim P_X \), is passed through a known linear transform that outputs \( z = Ax \in \mathbb{R}^m \). The components of \( y \in \mathbb{R}^m \) are generated by a componentwise transfer function \( P_{Y|Z} \). This paper addresses the cases where the distributions \( P_X \) and \( P_{Y|Z} \) have some parametric uncertainty that must be learned so as to properly estimate \( x \).

This joint estimation and learning problem with linear transforms and componentwise nonlinearities arises in a range of applications, including empirical Bayesian approaches to inverse problems in signal processing, linear regression and classification [1,2], and, more recently, Bayesian compressed
Adaptive GAMP

Approximate message passing (AMP) methods for this estimation problem. The AMP techniques use Gaussian and quadratic approximations of loopy belief propagation (LBP) to provide estimation methods that are computationally efficient, general and analytically tractable. However, the AMP methods generally require that the distributions $P_X$ and $P_{Y|Z}$ are known perfectly. When the parameters $\lambda_x$ and $\lambda_z$ are unknown, various extensions have been proposed including combining AMP methods with Expectation Maximization (EM) estimation [9–12] and hybrid graphical models approaches [13]. In this work, we present a novel method for joint parameter and vector estimation called adaptive generalized AMP (adaptive GAMP), that extends the GAMP method of [14]. We present two major theoretical results related to adaptive GAMP: We first show that, similar to the analysis of the standard GAMP algorithm, the componentwise asymptotic behavior of adaptive GAMP can be exactly described by a simple scalar state evolution (SE) equations [14–18]. An important consequence of this result is a theoretical justification to the EM-GAMP algorithm in [9–12] which is a special case of adaptive GAMP with a particular choice of adaptation functions. Our second result demonstrates the asymptotic consistency of adaptive GAMP when adaptation functions correspond to the maximum-likelihood (ML) parameter estimation. We show that when the ML estimation is computed exactly, the estimated parameters converge to the true values and the performance of adaptive GAMP asymptotically coincides with the performance of the oracle GAMP algorithms that knows correct parameter values. Adaptive GAMP thus provides a computationally-efficient method for solving a wide variety of joint estimation and learning problems with a simple, exact performance characterization and provable conditions for asymptotic consistency.

All proofs and some technical details that have been omitted for space appear in the full paper [19] that also provides more background and simulations.

2 Adaptive GAMP

Approximate message passing (AMP) refers to a class of algorithms based on Gaussian approximations of loopy belief propagation (LBP) for the estimation of the vectors $x$ and $z$ according to the model described in Section 1. These methods originated from CDMA multiuser detection problems in [15, 20, 21]; more recently, they have attracted considerable attention in compressed sensing [17, 18, 22]. The Gaussian approximations used in AMP are closely related to standard expectation propagation techniques [23, 24], but with additional simplifications that exploit the linear coupling between the variables $x$ and $z$. The key benefits of AMP methods are their computational performance, their large domain of application, and, for certain large random $A$, their exact asymptotic performance characterizations with testable conditions for optimality [15–18]. This paper considers an adaptive version of the so-called generalized AMP (GAMP) method of [14] that extends the algorithm in [22] to arbitrary output distributions $P_{Y|Z}$.

The original GAMP algorithm of [14] requires that the distributions $P_X$ and $P_{Y|Z}$ are known. We propose an adaptive GAMP, shown in Algorithm 1, to allow for simultaneous estimation of the distributions $P_X$ and $P_{Y|Z}$ along with the estimation of $x$ and $z$. The algorithm assumes that distributions $P_X$ and $P_{Y|Z}$ have some parametric forms

$$P_X(x|\lambda_x), \quad P_{Y|Z}(y|z, \lambda_z),$$

(1)
for parameters $\lambda_x \in \Lambda_x$ and $\lambda_z \in \Lambda_z$ and for parameter set $\Lambda_x$ and $\Lambda_z$. Algorithm 1 produces a sequence of estimates $\hat{x}^t$ and $\hat{z}^t$ for $x$ and $z$ along with parameter estimates $\hat{\lambda}_x^t$ and $\hat{\lambda}_z^t$. The precise value of these estimates depends on several factors in the algorithm including the termination criteria and the choice of what we will call estimation functions $G_x^t$, $G_z^t$ and $G_s^t$, and adaptation functions $H_x^t$ and $H_z^t$.

Algorithm 1 Adaptive GAMP

**Require:** Matrix $A$, estimation functions $G_x^t$, $G_z^t$ and $G_s^t$ and adaptation functions $H_x^t$ and $H_z^t$.  

1: Initialize $t \leftarrow 0$, $s^{-1} \leftarrow 0$ and some values for $\hat{x}^0$, $\tau_{z}^0$.
2: **repeat**
   3:   {Output node update}
   4:   $\tau_{p}^t \leftarrow \|A\|_F^2 \tau_{p}^t / m$
   5:   $p^t \leftarrow A \hat{x}^t - s^{-1} \tau_{p}^t$
   6:   $\hat{\lambda}_x^t \leftarrow H_z^t(p^t, y, \tau_{p}^t)$
   7:   $\hat{z}^t_i \leftarrow G_x^t(p^t, y, \tau_{p}^t, \hat{\lambda}_x^t)$ for all $i = 1, \ldots, m$
   8:   $s^t_i \leftarrow G_s^t(p^t, y, \tau_{p}^t, \hat{\lambda}_z^t)$ for all $i = 1, \ldots, m$
   9:   $\tau_{s}^t \leftarrow -(1/m) \sum_i \partial G_s^t(p^t, y, \tau_{p}^t, \hat{\lambda}_z^t) / \partial p_i$
10: **end repeat**
11: {Input node update}
12: $1/\tau_{r}^t \leftarrow \|A\|_F^2 \tau_{r}^t / n$
13: $r^t = x^t + \tau_{r}^t A^T s^t$
14: $\hat{\lambda}_z^t \leftarrow H_x^t(r^t, \tau_{r}^t)$
15: $\hat{x}^{t+1}_j \leftarrow G_x^t(r^t_j, \tau_{r}^t, \hat{\lambda}_x^t)$ for all $j = 1, \ldots, n$
16: $\tau_{z}^{t+1} \leftarrow (\tau_{r}^t / n) \sum_j \partial G_x^t(r^t_j, \tau_{r}^t, \hat{\lambda}_x^t) / \partial r^t_j$
17: **until** Terminated

The choice of the estimation and adaptation functions allows for considerable flexibility in the algorithm. For example, it is shown in [14] that $G_x^t$, $G_z^t$, and $G_s^t$ can be selected such that the GAMP algorithm implements Gaussian approximations of either max-sum LBP or sum-product LBP that approximate the maximum-a-posteriori (MAP) or minimum-mean-squared-error (MMSE) estimates of $x$ given $y$, respectively. The adaptation functions can also be selected for a number of different parameter-estimation strategies. Because of space limitation, we present only the estimation functions for the sum-product GAMP algorithm from [14] along with an ML-type adaptation. Some of the analysis below, however, applies more generally.

As described in [14], the sum-product estimation can be implemented with the functions

$$G_x^t(r, \tau_r, \hat{\lambda}_x) := E[X|R = r, \tau_r, \hat{\lambda}_x], \tag{2a}$$
$$G_z^t(p, y, \tau_p, \hat{\lambda}_z) := E[Z|P = p, Y = y, \tau_p, \hat{\lambda}_z], \tag{2b}$$
$$G_s^t(p, y, \tau_p, \hat{\lambda}_z) := \frac{1}{\tau_p} \left( G_x^t(p, y, \tau_p, \hat{\lambda}_z) - p \right), \tag{2c}$$

where the expectations are with respect to the scalar random variables

$$R = X + V_x, \quad V_x \sim \mathcal{N}(0, \tau_r), \quad X \sim P_X(\hat{\lambda}_x), \tag{3a}$$
$$Z = P + V_z, \quad V_z \sim \mathcal{N}(0, \tau_p), \quad Y \sim P_Y|Z(\hat{\lambda}_z). \tag{3b}$$

The estimation functions (2) correspond to scalar estimates of random variables in additive white Gaussian noise (AWGN). A key result of [14] is that, when the parameters are set to the true values (i.e., $\hat{\lambda}_x = (\lambda_x, \lambda_z)$), the outputs $\hat{x}^t$ and $\hat{z}^t$ can be interpreted as sum products estimates of the conditional expectations $E(x|y)$ and $E(z|y)$. The algorithm thus reduces the vector-valued estimation problem to a computationally simple sequence of scalar AWGN estimation problems along with linear transforms.

The estimation functions $H_x^t$ and $H_z^t$ in Algorithm 1 produce the estimates for the parameters $\lambda_x$ and $\lambda_z$. In the special case when $H_x^t$ and $H_z^t$ produce fixed outputs

$$H_x^t(p^t, y^t, \tau_{p}^t) = \hat{\lambda}_x^t, \quad H_z^t(r^t, \tau_{r}^t) = \hat{\lambda}_z^t,$$
for pre-computed values of $X^i_t$ and $Y^i_t$, the adaptive GAMP algorithm reduces to the standard (non-adaptive) GAMP algorithm of [14]. The non-adaptive GAMP algorithm can be used when the parameters $\lambda_x$ and $\lambda_z$ are known.

When the parameters $\lambda_x$ and $\lambda_z$ are unknown, it has been proposed in [9–12] that they can be estimated via an EM method that exploits that fact that GAMP provides estimates of the posterior distributions of $x$ and $z$ given the current parameter estimates. As described in the full paper [19], this EM-GAMP method corresponds to a special case of the Adaptive GAMP method for a particular choice of the adaptation functions $H^t_x$ and $H^t_z$.

However, in this work, we consider an alternate parameter estimation method based on ML adaptation. The ML adaptation uses the following fact that we will rigorously justify below: For certain large random $A$, at any iteration $t$, the components of the vectors $r^t$ and the joint vectors $(p^t, y^t)$ will be distributed as

$$ R = \alpha_r X + V_x, \quad V_x \sim \mathcal{N}(0, \xi_r), \quad X \sim P_X(\lambda^*_x), $$

$$ Z = P + V_z, \quad (Z, P) \sim \mathcal{N}(0, \lambda^*_z), \quad P \sim P_{Y|Z}(\|Z, \lambda^*_z), $$

where $\lambda^*_x$ and $\lambda^*_z$ are the “true” parameters and the scalars $\alpha_r$ and $\xi_r$ and the covariance matrix $K_p$ are some parameters that depend on the estimation and adaptation functions used in the previous iterations. Remarkably, the distributions of the components of $r^t$ and $(p^t, y^t)$ will follow (4) even if the estimation functions in the iterations prior to $t$ used the incorrect parameter values. The adaptive GAMP algorithm can thus attempt to estimate the parameters via a maximum likelihood (ML) estimation:

$$ H^t_x(r^t, \tau^t_r) := \arg \max_{\lambda_x} \max_{(\alpha_r, \xi_r) \in S_x(\tau^t_r)} \left\{ \frac{1}{n} \sum_{j=0}^{n-1} \phi_x(r^t_j, \lambda_x, \alpha_r, \xi_r) \right\}, $$

$$ H^t_z(p^t, y^t, \tau^t_p) := \arg \max_{\lambda_z} \max_{K_p \in S_z(\tau^t_p)} \left\{ \frac{1}{m} \sum_{i=0}^{m-1} \phi_z(p^t_i, y_i, K_p) \right\}, $$

where $S_x$ and $S_z$ are sets of possible values for the parameters $\alpha_r$, $\xi_r$ and $K_p$. $\phi_x$ and $\phi_z$ are the log-likelihoods

$$ \phi_x(r, \lambda_x, \alpha_r, \xi_r) = \log P_x(r|\lambda_x, \alpha_r, \xi_r), $$

$$ \phi_z(p, y, \lambda_z, K_p) = \log P_{Y|Z}(p|y, \lambda_z, K_p) $$

and $p_R$ and $p_{P;Y}$ are the probability density functions corresponding to the distributions in (4).

### 3 Convergence and Asymptotic Consistency with Gaussian Transforms

#### 3.1 General State Evolution Analysis

Before proving the asymptotic consistency of the adaptive GAMP method with ML adaptation, we first prove a more general convergence result. Among other consequences, the result will justify the distribution model (4) assumed by the ML adaptation. Similar to the SE analyses in [14, 18] we consider the asymptotic behavior of the adaptive GAMP algorithm with large i.i.d. Gaussian matrices. The assumptions are summarized as follows. Details can be found in the full paper [19, Assumption 2].

**Assumption 1** Consider the adaptive GAMP algorithm running on a sequence of problems indexed by the dimension $n$, satisfying the following:

(a) For each $n$, the matrix $A \in \mathbb{R}^{m \times n}$ has i.i.d. components with $A_{ij} \sim \mathcal{N}(0, 1/m)$ and the dimension $m = m(n)$ is a deterministic function of $n$ satisfying $n/m \to \beta$ for some $\beta > 0$ as $n \to \infty$.

(b) The input vectors $x$ and initial condition $\hat{x}^0$ are deterministic sequences whose components converge empirically with bounded moments of order $s = 2k - 2$ as

$$ \lim_{n \to \infty} (x, \hat{x}^0)_{PL(s)} (X, \hat{X}^0), $$
to some random vector \((X, \hat{X}^0)\) for \(k = 2\). See [19] for a precise statement of this type of convergence.

(c) The output vectors \(z\) and \(y \in \mathbb{R}^m\) are generated by

\[ z = Ax, \quad y = h(z, w), \tag{8} \]

for some scalar function \(h(z, w)\) where the disturbance vector \(w\) is deterministic, but empirically converges as

\[ \lim_{n \to \infty} w^{PL(s)} = W, \tag{9} \]

with \(s = 2k - 2, k = 2\) and \(W\) is some random variable. We let \(P_{Y|Z}\) denote the conditional distribution of the random variable \(Y = h(Z, W)\).

(d) Suitable continuity assumptions on the estimation functions \(G_x^t, G_z^t\) and \(G_z^t\) and adaptation functions \(H_x^t\) and \(H_z^t\) – see [19] for details.

Now define the sets of vectors

\[ \theta_x^t := \{(x_j, \tilde{r}_j^t, \tilde{x}_j^{t+1})\}, \quad \theta_z^t := \{(z_i, \tilde{z}_i^t, y_i, p_i^t)\}, \quad j = 1, \ldots, n,\]

\[ \theta_x^t := \{(z_i, \tilde{z}_i^t, y_i, p_i^t)\}, \quad j = 1, \ldots, m. \tag{10} \]

The first vector set, \(\theta_x^t\), represents the components of the “true,” but unknown, input vector \(x\), its adaptive GAMP estimate \(\hat{X}^t\) as well as \(r^t\). The second vector, \(\theta_z^t\), contains the components of the “true,” but unknown, output vector \(z\), its GAMP estimate \(\hat{z}^t\) as well as \(p^t\) and the observed output \(y\).

The sets \(\theta_x^t\) and \(\theta_z^t\) are implicitly functions of the dimension \(n\). Our main result, Theorem 1 below, characterizes the asymptotic joint distribution of the components of these two sets as \(n \to \infty\). Specifically, we will show that the empirical distribution of the components of \(\theta_x^t\) and \(\theta_z^t\) converge to a random vector of the form

\[ \bar{\theta}_x^t := (X, R^t, \hat{X}^{t+1}), \quad \bar{\theta}_z^t := (Z, \hat{z}^t, Y, P^t), \tag{11} \]

where \(X\) is the random variable in the initial condition (7). \(R^t\) and \(\hat{X}^{t+1}\) are given by

\[ R^t = \alpha_t^t X + V^t, \quad V^t \sim \mathcal{N}(0, \xi_t^t), \quad \hat{X}^{t+1} = G_x^t(R^t, \tau_r^t, \bar{X}_x^t) \]

for some deterministic constants \(\alpha_t^t, \xi_t^t, \tau_r^t\) and \(\bar{X}_x^t\) that will be defined momentarily. Similarly, \((Z, P^t) \sim \mathcal{N}(0, K_x^t)\), and

\[ Y \sim P_{Y|Z}(\cdot | Z), \quad \hat{z}^t = G_z^t(P^t, Y, \tau_p^t, \bar{X}_z^t), \tag{13} \]

where \(W\) is the random variable in (9) and \(K_x^t\) and \(\bar{X}_z^t\) are also deterministic constants. The deterministic constants above can be computed iteratively with the following state evolution (SE) equations shown in Algorithm 2.

**Theorem 1** Consider the random vectors \(\theta_x^t\) and \(\theta_z^t\) generated by the outputs of GAMP under Assumption 1. Let \(\bar{\theta}_x^t\) and \(\bar{\theta}_z^t\) be the random vectors in (11) with the parameters determined by the SE equations in Algorithm 2. Then, for any fixed \(t\), almost surely, the components of \(\theta_x^t\) and \(\theta_z^t\) converge empirically with bounded moments of order \(k = 2\) as

\[ \lim_{n \to \infty} \theta_x^t^{PL(k)} = \bar{\theta}_x^t, \quad \lim_{n \to \infty} \theta_z^t^{PL(k)} = \bar{\theta}_z^t, \tag{17} \]

where \(\bar{\theta}_x^t\) and \(\bar{\theta}_z^t\) are given in (11). In addition, for any \(t\), the limits

\[ \lim_n \lambda_x^t = \bar{X}_x^t, \quad \lim_n \lambda_z^t = \bar{X}_z^t, \quad \lim_n \tau_r^t = \tau_r^t, \quad \lim_n \tau_p^t = \tau_p^t, \tag{18} \]

also hold almost surely.

Similar to several other analyses of AMP algorithms such as [14–18], the theorem provides a scalar equivalent model for the componentwise behavior of the adaptive GAMP method. That is, asymptotically the components of the sets \(\theta_x^t\) and \(\theta_z^t\) in (10) are distributed identically to simple scalar random variables. The parameters in these random variables can be computed via the SE equations.

5
Consider a family of distributions, \( \text{identifiability} \) the correct parameter estimates. This condition is essentially an function (5). Thus, the ML adaptation should work provided that the maximizations in (5) yield parameter consistency of Adaptive GAMP with ML adaptation. The key point is to realize that analysis of the EM-GAMP presented in \([11, 12]\). Here, we use the result to prove the asymptotic estimation and adaptation function. In particular, the result can be used to rigorously justify the SE analysis of the EM-GAMP presented in \([11, 12]\). Here, we use the result to prove the asymptotic parameter consistency of Adaptive GAMP with ML adaptation. The key point is to realize that the distributions (12) and (13) exactly match the distributions (4) assumed by the ML adaptation functions (5). Thus, the ML adaptation should work provided that the maximizations in (5) yield the correct parameter estimates. This condition is essentially an identifiability requirement that we make precise with the following definitions.

### Definition 1
Consider a family of distributions, \( \{ P_X(x|\lambda_x), \lambda_x \in \Lambda_x \} \), a set \( S_x \) of parameters \( (\alpha_r, \xi_r) \) of a Gaussian channel and function \( \phi_x(r, \lambda_x, \alpha_r, \xi_r) \). We say that \( P_X(x|\lambda_x) \) is identifiable with Gaussian outputs with parameter set \( S_x \) and function \( \phi_x \) if:

(a) The sets \( S_x \) and \( \Lambda_x \) are compact.
(b) For any “true” parameters \( \lambda^*_x \in \Lambda_x \), and \( (\alpha_r, \xi_r) \in S_x \), the maximization

\[
\hat{\lambda}_x = \arg \max_{\lambda_x \in \Lambda_x} \max_{(\alpha_r, \xi_r) \in S_x} \mathbb{E}[\phi_x(\alpha^*_r X + V, \lambda_x, \alpha_r, \xi_r)|\lambda^*_x, \alpha^*_r, \xi^*_r],
\]

is well-defined, unique and returns the true value, \( \hat{\lambda}_x = \lambda^*_x \). The expectation in (19) is with respect to \( X \sim P_X(x|\lambda^*_x) \) and \( V \sim \mathcal{N}(0, \xi^*_r) \).

---

**Algorithm 2 Adaptive GAMP State Evolution**

Given the distributions in Assumption 1, compute the sequence of parameters as follows:

- **Initialization:** Set \( t = 0 \) with

\[
K^0_x = \text{cov}(X, \hat{X}^0), \quad \tau^0_x = \tau^0_x, \quad (14)
\]

where the expectation is over the random variables \((X, \hat{X}^0)\) in Assumption 1(b) and \( \tau^0_x \) is the initial value in the GAMP algorithm.

- **Output node update:** Compute the variables associated with \( \overline{\alpha}^t_x \):

\[
\tau^t_r = \beta \tau^t_x, \quad K^t_p = \beta K^t_x, \quad \overline{X}^t_x = H^t_x(P^t, \tau^t_p), \quad (15a)
\]

\[
\tau^t_r = -E^{-1} \left[ \frac{\partial}{\partial r} G^t_s(P^t, Y, \tau^t_p, \overline{X}^t_x) \right], \quad \xi^t_r = (\tau^t_r)^2 E \left[ G^t_s(P^t, Y, \tau^t_p, \overline{X}^t_x) \right], \quad (15b)
\]

\[
\alpha^t_r = \tau^t_r E \left[ \frac{\partial}{\partial r} G^t_s(\hat{P}, h(z, W), \tau^t_p, \overline{X}^t_x) \right] \bigg|_{z=Z}, \quad (15c)
\]

where the expectations are over the random variables \((P^t, Y, W)\).

- **Input node update:** Compute the variables associated with \( \overline{\alpha}^t_x \):

\[
\overline{X}^t_x = H^t_x(R^t, \tau^t_r), \quad (16a)
\]

\[
\tau^{t+1}_x = \tau^t_x E \left[ \frac{\partial}{\partial r} G^t_x(R^t, \tau^t_r, \overline{X}^t_x) \right], \quad K^{t+1}_x = \text{cov}(X, \hat{X}^{t+1}), \quad (16b)
\]

where the expectation is over the random variable \((X, \hat{X}^{t+1})\).
from m parameters the performance of adaptive GAMP for estimation of a sparse Gauss-Bernoulli signal

GAMP method due to its ability to learn the prior. Here, we present a simple simulation to illustrate

Recent results suggest that there is considerable value in learning of priors \( P_X \) in the context of compressed sensing [25], which considers the estimation of sparse vectors \( x \) from underdetermined measurements \( (m < n) \). It is known that estimators such as LASSO offer certain optimal min-max performance over a large class of sparse distributions [26]. However, for many particular distributions, there is a potentially large performance gap between LASSO and MMSE estimator with the correct prior. This gap was the main motivation for [9, 10] which showed large gains of the EM-GAMP method due to its ability to learn the prior. Here, we present a simple simulation to illustrate the performance gain of adaptive GAMP and its asymptotic consistency. Specifically, Fig. 2 compares the performance of adaptive GAMP for estimation of a sparse Gauss-Bernoulli signal \( x \in \mathbb{R}^n \) from \( m \) noisy measurements

\[
y = Ax + w,
\]
where the additive noise $w$ is random with i.i.d. entries $w_i \sim N(0, \sigma^2)$. The signal of length $n = 400$ has 20\% nonzero components drawn from the Gaussian distribution of variance 5. Adaptive GAMP uses EM iterations, which are used to approximate ML parameter estimation, to jointly recover the unknown signal $x$ and the true parameters $\lambda_x = (\rho = 0.2, \sigma_x^2 = 5)$. The performance of adaptive GAMP is compared to that of LASSO with MSE optimal regularization parameter, and oracle GAMP that knows the parameters of the prior exactly. For generating the graphs, we performed 1000 random trials by forming the measurement matrix $A$ from i.i.d. zero-mean Gaussian random variables of variance $1/m$. In Figure 2(a), we keep the variance of the noise fixed to $\sigma^2 = 0.1$ and plot the average MSE of the reconstruction against the measurement ratio $m/n$. In Figure 2(b), we keep the measurement ratio fixed to $m/n = 0.75$ and plot the average MSE of the reconstruction against the noise variance $\sigma^2$. For completeness, we also provide the asymptotic MSE values computed via SE recursion. The results illustrate that GAMP significantly outperforms LASSO over the whole range of $m/n$ and $\sigma^2$. Moreover, the results corroborate the consistency of adaptive GAMP which achieves nearly identical quality of reconstruction with oracle GAMP. The performance results here and in [19] indicate that adaptive GAMP can be an effective method for estimation when the parameters of the problem are difficult to characterize and must be estimated from data.

5 Conclusions and Future Work

We have presented an adaptive GAMP method for the estimation of i.i.d. vectors $x$ observed through a known linear transforms followed by an arbitrary, componentwise random transform. The procedure, which is a generalization of EM-GAMP methodology of [9, 10], estimates both the vector $x$ as well as parameters in the source and componentwise output transform. In the case of large i.i.d. Gaussian transforms with ML parameter estimation, it is shown that the adaptive GAMP method is provably asymptotically consistent in that the parameter estimates converge to the true values. This convergence result holds over a large class of models with essentially arbitrarily complex parameterizations. Moreover, the algorithm is computationally efficient since it reduces the vector-valued estimation problem to a sequence of scalar estimation problems in Gaussian noise. We believe that this method is applicable to a large class of linear-nonlinear models with provable guarantees and that it can have applications in a wide range of problems. We have mentioned the use of the method for learning sparse priors in compressed sensing. Future work will include possible extensions to non-Gaussian matrices.

References


