Fixed Points of Generalized Approximate Message Passing with Arbitrary Matrices

Sundeep Rangan, Philip Schniter, Erwin Riegler, Alyson Fletcher, Volkan Cevher

***Abstract*—The estimation of a random vector with independent components passed through a linear transform followed by a componentwise (possibly nonlinear) output map arises in a range of applications. Approximate message passing (AMP) methods, based on Gaussian approximations of loopy belief propagation, have recently attracted considerable attention for such problems. For large random transforms, these methods exhibit fast convergence and admit precise analytic characterizations with testable conditions for optimality, even for certain non-convex problem instances. However, the behavior of AMP under general transforms is not fully understood. In this paper, we consider the generalized AMP (GAMP) algorithm and relate the method to more common optimization techniques. This analysis enables a precise characterization of the GAMP algorithm fixed-points that applies to arbitrary transforms. In particular, we show that the fixed points of the so-called max-sum GAMP algorithm for MAP estimation are critical points of a constrained maximization of the posterior density. The fixed-points of the sum-product GAMP algorithm for estimation of the posterior marginals can be interpreted as critical points of a certain mean-field variational optimization.**

***Index Terms*—Belief propagation, ADMM, variational optimization, message passing.**

I. Introduction

Consider the constrained optimization problem

$$(\widehat{\mathbf{x}}, \widehat{\mathbf{z}}) := \arg\min_{\mathbf{x},\mathbf{z}} F(\mathbf{x}, \mathbf{z}) \quad \text{s.t. } \mathbf{z} = \mathbf{A}\mathbf{x}, \tag{1}$$

where $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{z} \in \mathbb{R}^m$, $\mathbf{A} \in \mathbb{R}^{m\times n}$ and the objective function admits a decomposition of the form

$$\begin{aligned} F(\mathbf{x}, \mathbf{z}) &:= f_x(\mathbf{x}) + f_z(\mathbf{z}) \\ f_x(\mathbf{x}) &= \sum_{j=1}^{n} f_{x_j}(x_j), \quad f_z(\mathbf{z}) = \sum_{i=1}^{m} f_{z_i}(z_i), \end{aligned} \tag{2}$$

for scalar functions $f_{x_j}(\cdot)$ and $f_{z_i}(\cdot)$. One example where this optimization arises is the estimation problem in Fig. 1. Here, a random vector $\mathbf{x}$ has independent components with densities $p_{x_j}(x_j($ 2001 0 d [()1(e)()00()12(n)11(i)1()()12()01(a)(u)12()1] 200 d [(s)221(a)(n)1(n)11(e)()2()1011()02(i)()1()(e)(r)12(l)1(a methods when compared against state-of-the-art conventional optimization techniques.

However, despite recent extensions to larger classes of random matrices [27], [28], the behavior of AMP methods under general $\mathbf{A}$ is not fully understood. The broad purpose of this paper is to show that certain forms of AMP algorithms can be seen as variants of more conventional optimization methods. This analysis will enable a precise characterization of the fixed points of the AMP methods that applies to arbitrary $\mathbf{A}$.

Our study focuses on a generalized AMP (GAMP) method proposed in [26] and rigorously analyzed in [29]. We consider this algorithm since many other variants of AMP are special cases of this general procedure. The GAMP method has two common versions: max-sum GAMP for the MAP estimation of and variances in lines 13, 14, 26 and 27 of the algorithm are taken with respect to the probability density functions:

$$p(\mathbf{x}|\mathbf{r},\boldsymbol{\tau}_r) \propto \exp\left[-f_x(\mathbf{x}) + \frac{1}{2}\|\mathbf{x}-\mathbf{r}\|^2_{\boldsymbol{\tau}_r}\right] \tag{8a}$$

$$p(\mathbf{z}|\mathbf{p},\boldsymbol{\tau}_p) \propto \exp\left[-f_z(\mathbf{z}) + \frac{1}{2}\|\mathbf{z}-\mathbf{p}\|^2_{\boldsymbol{\tau}_p}\right], \tag{8b}$$

where, for any vectors $\mathbf{v} \in \mathbb{R}^r$ and $\boldsymbol{\tau} \in \mathbb{R}^r$ with $\boldsymbol{\tau} > 0$,

$$\|\mathbf{v}\|^2_{\boldsymbol{\tau}} := \sum_{i=1}^{r} \frac{|v_i|^2}{\tau_i}.$$

Under separability assumption (2), these densities factor as

$$p(\mathbf{x}|\mathbf{r},\boldsymbol{\tau}_r) \propto \prod_{j=1}^{n} \exp\left[-f_{x_j}(x_j) - \frac{|x_j - r_j|^2}{2\tau_{r_j}}\right] \tag{9a}$$

$$p(\mathbf{z}|\mathbf{p},\boldsymbol{\tau}_p) \propto \prod_{i=1}^{m} \exp\left[-f\right.$$

III. Fixed-Points of Max-Sum

Secondly, the objective function $J_{\rm KL}$ from (21) is replaced by

$$J_{\rm \ P}(b_x, b_z, \overline{\boldsymbol{\tau}}_p) := D(b_x||e^{-f_x}) + D(b_z||e^{-f_z}) + H_{\rm ga\ ss}(b_z, \overline{\boldsymbol{\tau}}_p) \quad (23)$$

where $\overline{\boldsymbol{\tau}}_p$ is a positive vector and $H_{\rm ga\ ss}(b_z, \overline{\boldsymbol{\tau}}_p)$ is the following Gaussian upper bound on the entropy $H(b_z)$:

$$H_{\rm ga\ ss}(b_z, \overline{\boldsymbol{\tau}}_p) := \sum_{i=1}^{m} \left[\frac{1}{2\overline{\tau}_{p_i}} \text{var}(z_i|b_{z_i}) + \frac{1}{2}\log(2\pi\overline{\tau}_{p_i}) \right]. \quad (24)$$

The third and final approximation is that the constraint $b_z = T_{\mathbf{A}} b_x$ is replaced by the weaker *moment matching* constraint pair $\mathbb{E}(\mathbf{z}|b_z) = \mathbf{A}\mathbb{E}(\mathbf{x}|b_x)$ and $\overline{\boldsymbol{\tau}}_p = \mathbf{S}\text{var}(\mathbf{x}|b_x)$, where $\mathbf{S}$ is given in line 4 of Algorithm 1. The resulting optimization is

$$(\widehat{b}_x, \widehat{b}_z, \boldsymbol{\tau}_p) = \underset{b_x, b_z, \overline{\boldsymbol{\tau}}_p}{\arg\min}\, J_{\rm \ P}(b_x, b_z, \overline{\boldsymbol{\tau}}_p) \quad (25a)$$

$$\text{s.t.} \quad \mathbb{E}(\mathbf{z}|b_z) = \mathbf{A}\mathbb{E}(\mathbf{x}|b_x), \quad \overline{\boldsymbol{\tau}}_p = \mathbf{S}\text{var}(\mathbf{x}|b_x). \quad (25b)$$

Note that in (25b), the variance $\text{var}(\mathbf{x}|b_x)$ denotes the vector with components $\text{var}(x_j|b_{x_j})$, not a covariance matrix. The next lemma provides a certain Gaussian interpretation to the approximate optimization (25).

Lemma 1: For any positive vector $\overline{\boldsymbol{\tau}}_p$ and any density functions b_x and b_z, $J_{\rm \ P}(b_x, b_z, \overline{\boldsymbol{\tau}}_p)$ is an upper bound:

$$J_{\rm \ P}(b_x, b_z, \overline{\boldsymbol{\tau}}_p) \geq J_{\rm KL}(b_x, b_z), \quad (26)$$

with equality in the case that b_z is separable and Gaussian and $\overline{\boldsymbol{\tau}}_p = \text{var}(\mathbf{z}|b_z)$.

Proof: See Appendix C. ■

Thus, the optimization (25) can be interpreted as an approximation where the distributions are factorizable and the ou

Due to the separable structure of the objective function (28), the optimization (29) can be regarded as minimizing a separable function ${F^x}_{\mathrm{P}}(\mathbf{u}) + {F^z}_{\mathrm{P}}(\mathbf{v})$ with linear constraints (29b) between $\mathbf{u}$ and $\mathbf{v}$. In this context, the $\mathbf{x}$ and $\mathbf{z}$ minimizations in (32) and (33b) follow the format of the ADMM minimizations in Algorithm 3 for certain choices of the auxiliary functions. On the other hand, the optimization over $\boldsymbol{\tau}_x$ and $\boldsymbol{\tau}_p$ components follow the gradient-based method in the generalized ISTA method in Algorithm 2. So, the sum-product GAMP algorithm can be seen as a hybrid of the ISTA and ADMM methods for the optimization (29), which is equivalent to the variational optimization (25).

Unfortunately, this hybrid ISTA-ADMM method is non-standard and there is no existing convergence theory on the algorithm. However, Theorem 2 at least shows that if the sum-product GAMP algorithm converges, its fixed points correspond to critical points of optimization (29).

It is useful to briefly compare Theorem 2 with the variational interpretation of standard loopy BP. It is well-known [32] that the fixed points of standard loopy BP can be interpreted as distributions on the factor and variable nodes that minimize the so-called Bethe free energy subject to certain local consistency constraints. In comparison, GAMP appears to minimize a Gaussian approximation of the KL divergence subject to weaker moment matching constraints between the distributions on the variable nodes. In this manner, the fixed-points of GAMP appears closer in form of those of expectation propagation (EP) methods that can also be interpreted as saddle points of a certain free energy subject to moment matching [33]. However, the exact relation between EP and sum-product GAMP fixed points requires further study.

Conclusions

Although AMP methods admit precise analyses in the context of large random transform matrices $\mathbf{A}$, their behavior for general matrices is less well-understood. This limitation is unfortunate since many transforms arising in practical problems such as imaging and regression are not well-modeled as realizations of large random matrices. To help overcome these limitations, this paper draws connections between AMP and certain variants of standard optimization methods that employ adaptive vector-valued step-sizes. These connections enable a precise characterization of the fixed-points of both max-sum and sum-product GAMP for the case of arbitrary transform matrices $\mathbf{A}$. The convergence of AMP methods for general $\mathbf{A}$ is, however, still not fully understood. Simulations (not shown here) have indicated, for example, that under general choices of $\mathbf{A}$, AMP may diverge. We hope that the connections between AMP and standard optimization methods provided here help to better understand, and even improve, AMP convergence with general matrices.

Acknowledgements

The authors would
Ay65t.A384738]TJ /0 9.96264 Tf 8.44 Td [(i)0.9655-279.738]TJ /2857]2394 Tf 7.46199 2.76016 Td 0384738]TJ /R6552
stinJ T*[(i)0.ooer ar por ato-66 Tf 80.1602 0 Td [(A)5d [(A)5]TJ -246.(a)-1189 Tf 7.8 3.6 Td [(z)-673 [(c)-1.87208]

where (a) follows from substituting (2) and (10) into (16b) and eliminating the terms that do not depend on $\mathbf{z}$; (b) follows from the definition of $\mathbf{p}^t$ in line 8; and (c) follows from the definition of $\mathbf{z}^t$ in line 10. This proves (16b). The update (16a) can be proven similarly. To prove (16c), observe that

$$\mathbf{s}^t \overset{(a)}{=} \frac{1}{\boldsymbol{\tau}_p^t}(\mathbf{z}^t - \mathbf{p}^t) \overset{(b)}{=} \mathbf{s}^{t-1} + \frac{1}{\boldsymbol{\tau}_p^t}(\mathbf{z}^t - \mathbf{A}\mathbf{x}^t)$$

where (a) follows from the update of $\mathbf{s}^t$ in line 16 in Algorithm 1 (recall that the division is componentwise); and (b) follows from the update for $\mathbf{p}^t$ in line 8. We have thus proven the equivalence of the max-sum GAMP algorithm with the Lagrangian updates (16).

Now consider any fixed point $(\widehat{\mathbf{z}}, \widehat{\mathbf{x}}, \mathbf{s})$ of the max-sum GAMP algorithm. A fixed point of (16c), requires that

$$\widehat{\mathbf{z}} = \mathbf{A}\widehat{\mathbf{x}} \tag{36}$$

so the fixed point satisfies the constraint of the optimization (1). Now, using (36) and the fact that $\widehat{\mathbf{z}}$ is the minima of (16b), we have that

$$\frac{\partial}{\partial \mathbf{z}} L(\widehat{\mathbf{x}}, \widehat{\mathbf{z}}, \mathbf{s}) = 0.$$

Similarly, since $\mathbf{x}$ is the minima of (16a), we have that

$$\frac{\partial}{\partial \mathbf{x}} L(\widehat{\mathbf{x}}, \widehat{\mathbf{z}}, \mathbf{s}) = 0.$$

Thus, the fixed point $(\widehat{\mathbf{x}}, \widehat{\mathbf{z}}, \mathbf{s})$ is a critical point of the Lagrangian (10).

Finally, consider the quadratic terms $(\boldsymbol{\tau}_x, \boldsymbol{\tau}_r, \boldsymbol{\tau}_s)$ at the fixed point. From the updates of $\boldsymbol{\tau}_x$ and $\boldsymbol{\tau}_r$ in Algorithm 1 [see also (7)] and the definition of $\mathbf{d}_x$ in (15), we obtain

$$\frac{1}{\boldsymbol{\tau}_x} = \mathbf{d}_x + \frac{1}{\boldsymbol{\tau}_r} = \mathbf{d}_x + \mathbf{S}^T\boldsymbol{\tau}_s. \tag{37}$$

Similarly, the updates of $\boldsymbol{\tau}_s$ and $\boldsymbol{\tau}_p$ show that

$$\frac{1}{\boldsymbol{\tau}_s} = \frac{1}{\mathbf{d}_z} + \boldsymbol{\tau}_p = \frac{1}{\mathbf{d}_z} + \mathbf{S}\boldsymbol{\tau}_x. \tag{38}$$

Then, according to Definition 1, $\boldsymbol{\tau}_x$ and $\boldsymbol{\tau}_s$ are the approximate diagonals of $\mathbf{Q}_x$ and $\mathbf{Q}_z$ in (14), respectively.

Appendix C
Proof of Lemma 1

For any positive vector $\overline{\boldsymbol{\tau}}_p$ and density function b_z (even if it is not separable), we have the bound

$$H(b_z) \overset{(a)}{\leq} \sum_{i=1}^{m} H(b_{z_i})$$

$$\overset{(b)}{\leq} \frac{1}{2}\sum_{i=1}^{m}$$ [redacted]

$$\overset{(c)}{\leq} \frac{1}{2}\sum^{m}$$ [redacted]

(a) The probability density function $\widehat{b}$ is the solution to the constrained optimization:

$$\widehat{b} = \arg\min_b D(b||e^{-f}) \quad \text{s.t. } \mathbb{E}(u|b) = \overline{u},\ \text{var}(u|b) = \tau_u. \tag{44}$$

(b) There exists q and $\tau_q > 0$ such that the density function $\hat{b}$ is the solution to the unconstrained optimization

$$\widehat{b} = \arg\min_b \Big[D(b||e^{-f}) + \frac{1}{2\tau_q}\mathbb{E}((u-q)^2|b)\Big], \tag{45}$$

and $\mathbb{E}(u|\widehat{b}) = \overline{u}$, $\text{var}(u|\widehat{b}) = \tau_u$.

(c) There exists q and $\tau_q > 0$ such that the density function $\hat{b}$ is of the form

$$\widehat{b} \propto \exp\Big[-f(u) - \frac{1}{2\tau_q}(u-q)^2\Big], \tag{46}$$

and $\mathbb{E}(u|\widehat{b}) = \overline{u}$, $\text{var}(u|\widehat{b}) = \tau_u$.

Proof: This result is standard—similar calculations are in [34]. The equivalence between (a) and (b) can be shown via a Lagrangian argument and the equivalence between (b) and (c) can be found by taking the derivatives of the unconstrained objective (45) with respect to $b(u)$ for each u. ■

B. Equivalence of GAMP and ADMM-ISTA Iterations

We now use Lemma 3 to that the sum-product GAMP iterations are equivalent to the ADMM-ISTA iterations in Theorem 2. We begin by proving (33b). Let $\tilde{\mathbf{z}}^t$ equal the right-hand side of (33b). We want to show that $\tilde{\mathbf{z}}^t = \mathbf{z}^t$, where $\mathbf{z}^t$ is the output of line 13 of the sum-product GAMP algorithm. To show this, we first observe that

$$\begin{aligned}\tilde{\mathbf{z}}^t &\overset{(a)}{=} \arg\min_{\overline{\mathbf{z}}} \Big[F^z{}_{\mathrm{P}}(\overline{\mathbf{z}}, \boldsymbol{\tau}_p^t) + (\mathbf{s}^{t-1})^T\overline{\mathbf{z}} + \frac{1}{2}\|\overline{\mathbf{z}} - \mathbf{A}\mathbf{x}^t\|^2_{\boldsymbol{\tau}_p^t}\Big] \\ &\overset{(b)}{=} \arg\min_{\overline{\mathbf{z}}} \Big[F^z{}_{\mathrm{P}}(\overline{\mathbf{z}}, \boldsymbol{\tau}_p^t) + \frac{1}{2}\|\overline{\mathbf{z}} - \mathbf{p}^t\|^2_{\boldsymbol{\tau}_p^t}\Big],\end{aligned} \tag{47}$$

where (a) follows from substituting (31) and (27) into (33b) and eliminating the terms that do not depend on $\overline{\mathbf{z}}$, and (b) follows from substituting in the definition of $\mathbf{p}^t$ from line (8) and eliminating terms. Now, using the definition of $F^z{}_{\mathrm{P}}$ in (42b), it follows from (47) that

$$\tilde{\mathbf{z}}^t = \mathbb{E}(\mathbf{z}|\widehat{b}_z), \tag{48}$$

where $\widehat{b}_z$ is the density function on $\mathbf{z}$ that minimizes

$$\widehat{b}_z = \arg\min_{b_z} \Big[J^z{}_{\mathrm{P}}(b_z, \boldsymbol{\tau}_p^t) + \frac{1}{2}\left\|\mathbb{E}(\mathbf{z}|b_z) - \mathbf{p}^t\right\|^2_{\boldsymbol{\tau}_p^t}\Big]. \tag{49}$$

Now, this minimization can be simplified as

$$\begin{aligned}\widehat{b}_z &\overset{(a)}{=} \arg\min_{b_z} \Big[D(b_z||e^{-f_z}) + \sum_{i=1}^m \frac{\text{var}(z_i|b_{z_i})}{2\tau_{p_i}^t} \\ &\quad + \frac{1}{2}\left\|\mathbb{E}(\mathbf{z}|b_z) - \mathbf{p}^t\right\|^2_{\boldsymbol{\tau}_p^t}\Big] \\ &= \arg\min_{b_z} \sum_{i=1}^m \Big[D(b_{z_i}||e^{-f_{z_i}}) + \frac{\mathbb{E}\big(|z_i - p_i^t|^2 \,\big|\, b_{z_i}\big)}{2\tau_{p_i}^t}\Big]\end{aligned} \tag{50}$$

where (a) follows from substituting (41b) and (24) into (49) and removing terms that do not depend on b_z; and (b) follows from the separability assumption (22) and the fact that, for any density function b_{z_i},

$$\text{var}(z_i|b_{z_i}) + \big|\mathbb{E}(z_i|b_{z_i}) - p_i^t\big|^2 = \mathbb{E}\big(|z_i - p_i^t|^2 \,\big|\, b_{z_i}\big).$$

The minimization (50) is then separable, with solution

$$\widehat{b}_z(\mathbf{z}) = \prod_{i=1}^m \widehat{b}_{z_i}(z_i) \tag{51}$$

whose components are the solutions

$$\widehat{b}_{z_i}(z_i) = \arg\min_{b_{z_i}} \Big[D(b_{z_i}||e^{-f_{z_i}}) + \frac{1}{2\tau_{p_i}^t}\mathbb{E}\big(|z_i - p_i^t|^2|b_{z_i}\big)\Big]. \tag{52}$$

From Lemma 3, the solution to (52) can be restated as

$$\widehat{b}_{z_i}(z_i) \propto \exp\Big[-f_{z_i}(z_i) - \frac{|z_i - p_i^t|^2}{2\tau_{p_i}^t}\Big]$$

,(z)5.11187]TJJ -79R4221(i)0.965521(t)0.965521(h)-355.18(s)3.5R34 9.96264 0 0 -46687.08 5529518154 B75.668 ae /W 1 /H 1 /BPC 1
f t td-338.94(m)-4..25908(a)-391.345()0.963071(o)-5.88115(n)-1.66638(n)-34
=a(1 0101011 11 11 11 1000000100
n Tf T /TJ T Q gcm true /H /D Q q cmT TTf TT Tf T ttTJ T pTJ Tf T T TJJ TTf TT T Tf T tTJ T pTJ m true /H /D

critical point of the optimization (29), we will show that it is a critical point of the modified Lagrangian $L_{\text{P-mo}}$ and that it satisfies the constraint $\widehat{\mathbf{z}} = \mathbf{A}\widehat{\mathbf{x}}$.

First, the vector components of the sum-product GAMP fixed-point must be fixed-points of the Lagrangian updates (32) and (33). Thus, from (33a), we have that

$$\boldsymbol{\tau}_p = \mathbf{S}\boldsymbol{\tau}_x, \tag{57}$$

while from (33c) we have that

$$\widehat{\mathbf{z}} = \mathbf{A}\widehat{\mathbf{x}}, \tag{58}$$

and so the fixed point satisfies both constraints in the optimization (29).

Now, using (58) and the fact that $\widehat{\mathbf{z}}$ is the minima of (33b), we have that

$$\frac{\partial}{\partial \overline{\mathbf{z}}} L_{\text{P}}(\widehat{\mathbf{x}}, \widehat{\mathbf{z}}, \boldsymbol{\tau}_x, \boldsymbol{\tau}_p, \mathbf{s}) = 0. \tag{59}$$

Due to (57), equation (59) implies that

$$\frac{\partial}{\partial \overline{\mathbf{z}}} L_{\text{P-mo}}(\widehat{\mathbf{x}}, \widehat{\mathbf{z}}, \boldsymbol{\tau}_x, \mathbf{s}) = 0. \tag{60}$$

Similarly, since $\widehat{\mathbf{x}}$ is the minima of (32), we have that

$$\frac{\partial}{\partial \overline{\mathbf{x}}} L_{\text{P-mo}}(\widehat{\mathbf{x}}, \widehat{\mathbf{z}}, \boldsymbol{\tau}_x, \mathbf{s}) = 0. \tag{61}$$

The minimization (32) also implies that

$$\frac{\partial}{\partial \overline{\boldsymbol{\tau}}_x} L_{\text{P}}(\widehat{\mathbf{x}}, \widehat{\mathbf{z}}, \boldsymbol{\tau}_x, \boldsymbol{\tau}_p, \mathbf{s}) = -\frac{1}{2}\mathbf{S}^T\boldsymbol{\tau}_s. \tag{62}$$

Therefore,

$$\begin{aligned} &\frac{\partial}{\partial \overline{\boldsymbol{\tau}}_x} L_{\text{P-mo}}(\widehat{\mathbf{x}}, \widehat{\mathbf{z}}, \boldsymbol{\tau}_x, \mathbf{s}) \\ &\stackrel{(a)}{=} \frac{\partial}{\partial \overline{\boldsymbol{\tau}}_x} L_{\text{P}}(\widehat{\mathbf{x}}, \widehat{\mathbf{z}}, \boldsymbol{\tau}_x, \boldsymbol{\tau}_p, \mathbf{s}) + \mathbf{S}^T \frac{\partial}{\partial \overline{\boldsymbol{\tau}}_p} L_{\text{P}}(\widehat{\mathbf{x}}, \widehat{\mathbf{z}}, \boldsymbol{\tau}_x, \boldsymbol{\tau}_p, \mathbf{s}) \\ &\stackrel{(b)}{=} -\frac{1}{2}\mathbf{S}^T\boldsymbol{\tau}_s + +\frac{1}{2}\mathbf{S}^T\boldsymbol{\tau}_s = 0, \end{aligned} \tag{63}$$

where (a) follows from the definition of $L_{\text{P-mo}}$ in (56) and from (57), while (b) follows from (62) and (33d). The derivatives (61), (60) and (63), along with the constraints (57) and (58), show that the vectors $\widehat{\mathbf{x}}$, $\widehat{\mathbf{z}}$, $\boldsymbol{\tau}_x$ and $\boldsymbol{\tau}_p$ are critical points of the optimization (29). Finally, using Lemma 3 and arguments similar to those used in derivation of (54) and (55), it follows that the density functions $\widehat{b}_x$ and $\widehat{b}_z$ that minimize (28) are those given in (34).

REFERENCES

[1] J. A. Nelder and R. W. M. Wedderburn, "Generalized linear models," *J. Royal Stat. Soc. Series A*, vol. 135, pp. 370–385, 1972.

[2] P. McCullagh and J. A. Nelder, *Generalized linear models*, 2nd ed. Chapman & Hall, 1989.

[3] Y. C. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications*. New York: Cambridge Univ. Press, 2012.

[4] A. Chambolle, R. A. DeVore, N. Y. Lee, and B. J. Lucier, "Nonlinear wavelet image processing: Variational problems, compression, and noise removal through wavelet shrinkage," *IEEE Trans. Image Process.*, vol. 7, no. 3, pp. 319–335, Mar. 1998.

[5] I. Daubechies, M. Defrise, and C. D. Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413–1457, Nov. 2004.

[6] C. Vonesch and M. Unser, "Fast iterative thresholding algorithm for wavelet-regularized deconvolution," in *Proc. SPIE: Wavelet XII*, San Diego, CA, 2012.

[7] S. J. Wright, R. D. Nowak, and M. Figueiredo, "Sparse reconstruction by separable approximation," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2479–2493, Jul. 2009.

[8] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problem," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[9] Y. E. Nesterov, "Gradient methods for minimizing composite objective function," *CORE Report*, 2007.

[10] J. Bioucas-Dias and M. Figueiredo, "A new TwIST: Two-step iterative shrinkage/thresholding algorithms for image restoration," *IEEE Trans. Image Process.*, vol. 16, no. 12, pp. 2992 – 3004, Dec. 2007.

[11] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, 2010.

[12] J. Eckstein and D. Bertsekas, "On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Math. Program.*, vol. 5, pp. 293–318, 1992.

[13] T. Goldstein and S. Osher, "The split Bregman method for L1-regularized problems," *SIAM J. Imaging Sciences*, vol. 2, no. 2, 2009.

[14] X. Zhang, M. Burger, and S. Osher, "A unified primal-dual algorithm framework based on Bregman iteration," *SIAM J. Sci. Comput.*, vol. 46, pp. 20–46, 2011.

[15] P. L. Combettes and V. R. Wajs, "Signal recovery by proximal forward-backward splitting," *Multiscale Model. Simul.*, vol. 4, pp. 1168–1200, 2005.

[16] P. Tseng, "Applications of a splitting algorithm to decomposition in convex programming and variational inequalities," *SIAM J. Control and Optimization*, vol. 29, no. 1, pp. 119–138, Jan. 1991.

[17] M. Zhu and T. Chan, "An efficient primal-dual hybrid gradient algorithm for total variation image restoration," UCLA CAM, Tech. Rep. 08-34, 2008.

[18] J. E. Esser, "Primal dual algorithms for convex models and applications to image restoration, registration and nonlocal inpainting," Ph.D. dissertation, The University of California, Los Angeles, 2010.

[19] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *J. Math. Imaging Vis.*, vol. 40, pp. 120–145, 2011.

[20] B. He and X. Yuan, "Convergence analysis of primal-dual algorithms for a saddle-point problem: From contraction perspective," *SIAM J. Imaging Sci.*, vol. 5, no. 1, pp. 119–149, 2012.

[21] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proc. Nat. Acad. Sci.*, vol. 106, no. 45, pp. 18 914–18 919, Nov. 2009.

[22] ——, "Message passing algorithms for compressed sensing I: motivation and construction," in *Proc. Info. Theory Workshop*, Jan. 2010.

[23] ——, "Message passing algorithms for compressed sensing II: analysis and validation," in *Proc. Info. Theory Workshop*, Jan. 2010.

[24] M. Bayati and A. Montanari, "The dynamics of message passing on dense graphs, with applications to compressed sensing," *IEEE Trans. Inform. Theory*, vol. 57, no. 2, pp. 764–785, Feb. 2011.

[25] S. Rangan, "Estimation with random linear mixing, belief propagation and compressed sensing," in *Proc. Conf. on Inform. Sci. & Sys.*, Princeton, NJ, Mar. 2010, pp. 1–6.

[26] ——, "Generalized approximate message passing for estimation with random linear mixing," in *Proc. IEEE Int. Symp. Inform. Theory*, Saint Petersburg, Russia, Jul.–Aug. 2011, pp. 2174–2178.

[27] G. Caire, S. Shamai, A. Tulino, and S. Verdú, "Support recovery in compressed sensing: Information-theoretic bounds," in *Proc. UCSD Workshop Inform. Theory & Its Applications*, La Jolla, CA, Jan. 2011.

[28] M. Bayati, M. Lelarge, and A. Montanari, "Universality in polytope phase transitions and iterative algorithms," in *Proc. ISIT*, Jul. 2012, pp. 1643 –1647.

[29] A. Javanmard and A. Montanari, "State evolution for general approximate message passing algorithms, with applications to spatial coupling," arXiv:1211.5164 [math.PR]., Nov. 2012.

[30] Y. Weiss and W. T. Freeman, "On the optimality of solutions of the max-product belief-propagation algorithm in arbitrary graphs," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 736–744, Feb. 2001.

[31] T. Heskes, "Stable fixed points of loopy belief propagation are minima of the Bethe free energy," in *Proc. Neural Information Process. Syst.*, Vancouver, Canada, Dec. 2003.

[32] J. S. Yedidia, W. T. Freeman, and Y. Weiss, "Understanding belief propagation and its generalizations," in *Exploring Artificial Intelligence*

in the New Millennium. San Francisco, CA: Morgan Kaufmann Publishers, 2003, pp. 239–269.

[33] M. Opper and O. Winther, "Expectation consistent approximate inference," *J. Mach. Learning Res.*, vol. 1, pp. 2177–2204, 2005.

[34] M. Wainwright and M. Jordan, "Graphical models, exponential families, and variational inference," *Found. Trends Mach. Learn.*, vol. 1, 2008.

[35] J. Boutros and G. Caire, "Iterative multiuser joint decoding: Unified framework and asymptotic analysis," *IEEE Trans. Inform. Theory*, vol. 48, no. 7, pp. 1772–1793, Jul. 2002.

[36] T. Tanaka and M. Okada, "Approximate belief propagation, density evolution, and neurodynamics for CDMA multiuser detection," *IEEE Trans. Inform. Theory*, vol. 51, no. 2, pp. 700–706, Feb. 2005.

[37] D. Guo and C.-C. Wang, "Asymptotic mean-square optimality of belief propagation for sparse linear systems," in *Proc. IEEE Inform. Theory Workshop*, Chengdu, China, Oct. 2006, pp. 194–198.

[38] A. Montanari, "Graphical model concepts in compressed sensing," in *Compressed Sensing: Theory and Applications*, Y. C. Eldar and G. Kutyniok, Eds. Cambridge Univ. Press, Jun. 2012, pp. 394–438.

[39] T. P. Minka, "A family of algorithms for approximate Bayesian inference," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, 2001.

[40] M. Seeger, "Bayesian inference and optimal design for the sparse linear model," *J. Machine Learning Research*, vol. 9, pp. 759–813, Sep. 2008.

[41] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Royal Stat. Soc., Ser. B*, vol. 58, no. 1, pp. 267–288, 1996.

[42] W. Hachem, P. Loubaton, and J. Najim, "Deterministic equivalents for certain functionals of large random matrices," *Ann. Applied Probability*, vol. 17, no. 3, pp. 875–930, Jun. 2007.