WEIGHTED-DAMPED APPROXIMATE MESSAGE PASSING FOR COMPRESSED SENSING

Shengchu Wang, Yunzhou Li, Zhen Gao and Jing Wang

Wireless and Mobile Communications R&D Center, Tsinghua University, Beijing, China
wsc11@mails.tsinghua.edu.cn, {liyunzhou, zgao, wangj} @tsinghua.edu.cn

ABSTRACT

Approximate Message Passing (AMP) simplified from Loopy Belief Propagation (LBP), is an important algorithm for sparse signal reconstruction in Compressed Sensing (CS). To improve the performance of current AMP algorithms, a weighted-damped AMP algorithm (WDAMP) is derived from a weighted version of BP that adopt probability damping technique. Simulation results show that WDAMP outperforms normal AMP for both 1-D and 2-D signal reconstruction. For 1-D signal reconstruction, probability damping brings most of the improvement. For 2-D signal reconstruction, weighting technique makes the major contribution. In summary, WDAMP outperforms conventional AMP.

Index Terms*—*Approximate Message Passing, Belief Propagation, Compressed Sensing, Tree-reweighted

1. INTRODUCTION

In Compressed Sensing (CS), the key problem is how to reconstruct the sparse signal as precisely as possible from noisy measurements obtained via an underdetermined linear observation equation [1], [2]. This problem can be reformulated as Bayesian inference on cyclic factor graph model [3], which can be solved through Loopy Belief Propagation (LBP) [4]. However, the computational complexity and memory requirement of LBP is too high for practical application. In order to solve this problem, Approximate Message Passing (AMP) simplified from LBP has been proposed in [3], [5] and [6].

Recently, LBP is improved as Tree-reweighted belief propagation (TRW-BP) in [7], and enhanced by the probability damping technique in [8]. Based on them, a weighted damped BP (WDBP) is constructed in this paper. WDBP adopts probability damping technique and replaces the factor appearance probability in TRW-BP [9] with positive weights satisfying an equality constraint. Simplified from WDBP, weighted-damped AMP is presented.

Simulation results show that WDAMP outperforms normal AMP for both 1-D and 2-D signal reconstruction. For 1-D signal reconstruction, probability damping brings most of the improvement. For 2-D signal reconstruction, weighting technique makes the most contribution. In summary, WDAMP outperforms conventional AMP.

2. PROBLEM MODEL

The mathematical model for CS can be expressed as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}, \tag{1}$$

in which $y \in \mathbb{R}^M$ is measurement vector, $\mathbf{x} \in \mathbb{R}^N$ is the identically independent distributed (*i.i.d*) sparse signal vector with K non-zero elements, $\mathbf{H} \in \mathbb{R}^{M\times N}$ is the column-wise normalized Gaussian measurement matrix, and $\mathbf{w} \in \mathbb{R}^M$ is the zero-mean white Gaussian noise with variance σ^2. It's assumed that $K < M < N$. In this paper, K and σ^2 are assumed to be known. The well-known CS problem is about how to reconstruct $\mathbf{x}$ from $\mathbf{y}$ as exactly as possible based on the underdetermined equation (1).

In Bayesian inference, signal prior probability distribution function (PDF) is needed. In this paper, the PDF of the *i.i.d* element in $\mathbf{x}$ is assumed to be Bernoulli-Gauss [6], which is expressed as

$$\varphi(x) = (1-p)\delta(x) + p\phi(x, \mu_x, \sigma_x^2) \tag{2}$$

in which p (fixed as K/N in this paper) is probability of the non-zero elements in $\mathbf{x}$, $\delta(x)$ is the Dirac function, and $\phi(x, \mu_x, \sigma_x^2)$ is the standard Gaussian distribution function with mean μ_x and variance σ_x^2.

3. WEIGHTED DAMPED APPROXIMATE MESSAGE PASSING

In this section, the weighted loopy belief propagation algorithm that adopts probability damping technique (WDBP) is designed firstly. Then the WDAMP is derived based on WDBP.

The work was supported by National 863 (No. 2012AA011402), National 973 (No.2012CB316000), NFS China (60832008).

3.1. Design of weighted damped loopy belief propagation

The i th element of $\mathbf{x}$ can be approximate as $\int dx_i x_i p(x_i \mid \mathbf{y})$, which is the Maximum A-Posteriori (MAP) estimate for x_i. The calculation of $p(x_i \mid \mathbf{y})$ is intractable. Fortunately, LBP can provide a useful approximate on $p(x_i \mid \mathbf{y})$. In this paper, a new weighted-damped BP (WDBP) is derived under the theoretical framework of LBP.

Based on (1), the posterior distribution of $\mathbf{x}$ over $\mathbf{y}$ is $p(\mathbf{x} \mid \mathbf{y})$, whose calculation is intractable again. $p(\mathbf{x} \mid \mathbf{y})$ will be approximated as $b(\mathbf{x}, \mathbf{y})$ under the objective of minimizing their Kullback-Leibler (KL) divergence. $b(\mathbf{x}, \mathbf{y})$ belongs to one local polytope defined as [9]

$$L(G) = \{b(\mathbf{x},\mathbf{y}) \mid b_i(x_i) \ge 0, b_\mu(y_\mu, \mathbf{x}) \ge 0, \int_{x_i} dx_i b_i(x_i) = 1, \int_{\mathbf{x}} d\mathbf{x} b_\mu(y_\mu, \mathbf{x}) = 1, \int \prod_{j \ne i} dx_j b_\mu(y_\mu, \mathbf{x}) = b_i(x_i)\}, \quad (3)$$

where $b_i(x_i)$ and $b_\mu(y_\mu, \mathbf{x})$ are normalized, non-negative, and mutually consistent marginal distributions, but not necessarily correspond to a valid global distributions [10], $i \in [N]$, $\mu \in [M]$, $[N] = \{1, 2, ..., N\}$, $[M] = \{1, 2, ..., M\}$.

The optimization problem can be formulated as

$$\max_{b \in L(G)} \{\mathrm{H}(b) + \sum_{i=1}^{N} \sum_{x_i} b_i(x_i) \log \varphi_i(x_i) + \sum_{\mu=1}^{M} \sum_{x_{[N]}} b_\mu(y_\mu, \mathbf{x}) \log \psi_\mu(y_\mu, \mathbf{x})\}, \quad (4)$$

where $\mathrm{H}(b)$ is defined as the entropy of the distribution $b(\mathbf{x})$, and $\varphi_i(x_i)$ is defined as (2), and $\psi_\mu(y_\mu, \mathbf{x})$ is

$$\psi_\mu(y_\mu, \mathbf{x}) = \frac{1}{\sqrt{2\pi}\sigma} \exp\{-\frac{(y_\mu - \sum_{k \in [N]} H_{\mu k} x_k)}{2\sigma^2}\}. \quad (5)$$

Similar as that in [4], $\mathrm{H}(b)$ can be approximated by the Bethe free energy defined as

$$\mathrm{H}_{Bethe}(b) = \sum_{i=1}^{N} \mathrm{H}(b_i) - \sum_{\mu=1}^{M} I_\mu(b_\mu), \quad (6)$$

in which $H(b_i)$ is the entropy of $b_i(x_i)$ and $I_\mu(b_\mu)$ is a mutual information term defined as

$$I_\mu(b_\mu) = \int_{x_{[N]}} b_\mu(y_\mu, \mathbf{x}) \log \frac{b_\mu(y_\mu, \mathbf{x})}{\prod_{i \in [N]} b_i(x_i)}. \quad (7)$$

By substituting $\mathrm{H}(b)$ in (4) with (6), combining the new objective function with the constraints in (3) through Larangian multipliers, and setting the derivative as 0, the formulas of LBP can be derived (refer to [4] for more details), which will provide approximations of $\{p(x_i \mid \mathbf{y})\}_{i \in [N]}$.

In this paper, a weighted version of Bethe free energy is proposed as

$$\mathrm{H}_{weight}(b) = \sum_{i=1}^{N} \mathrm{H}(b_i) - \sum_{\mu=1}^{M} \rho_\mu I_\mu(b_\mu). \quad (8)$$

In TRW-BP, $\{\rho_\mu\}$ are defined as factor appearance probabilities [9]. In this paper, $\{\rho_\mu\}$ are relaxed as the weights satisfying

$$\rho_\mu > 0, \sum_{\mu=1}^{M} \rho_\mu = M. \quad (9)$$

Similar as the derivation of LBP above, replace $\mathrm{H}(b)$ in (4) with (8), the weighted BP formulas can be derived as

$$m_{i \to \mu}(x_n) = \frac{1}{Z^{i \to \mu}} \varphi_i(x_i) m_{\mu \to i}^{\rho_\mu - 1}(x_i) \prod_{\nu \in [M] \setminus \mu} m_{\nu \to i}^{\rho_\nu}(x_i) \quad (10)$$

$$m_{\mu \to i}(x_i) = \frac{1}{Z^{\mu \to i}} \int \prod_{j \ne i} dx_j \psi_\mu^{1/\rho_\mu}(x_{[N]}) \prod_{j \in [N] \setminus i} m_{j \to \mu}(x_j), \quad (11)$$

$$m_i(x_i) = \frac{1}{Z^i} \varphi_i(x_i) \prod_{\mu \in [M]} m_{\mu \to i}^{\rho_\mu}(x_i), \quad (12)$$

where $Z^{i \to \mu}$, $Z^{\mu \to i}$ and Z^i are normalization factors, (Refer to the appendix of [9] for the proof). Based on the factor graph theory [10], $i \in [N]$ and $\mu \in [M]$ are variable and factor nodes respectively. (10) and (11) are the $2MN$ messages exchanging between M factor nodes and N variable nodes. The message passing procedures will be repeated many times until some termination conditions are satisfied. (12) are the beliefs of variables $\{x_i\}_{i \in [N]}$. Once $\{\rho_\mu\}$ are set as constant 1, (10) ~ (12) will degenerate to be the conventional LBP. Based on the "probability damping" introduced in [8], (12) can be modified as

$$m_i^{t+1}(x_i) = \frac{1}{\tilde{Z}^i} [m_i^t(x_i)]^\gamma [\varphi_i(x_i) \prod_{\mu \in [M]_i} m_{\mu \to i}^{\rho_\mu}(x_i)]^{1-\gamma}, \quad (13)$$

in which t indicates the current iteration and γ is the damping parameter (superscript t starts from 0). (10), (11) and (13) are the key formulas for WDBP. In this paper, $m_i^0(x_i) = 1$, and $m_i^{t+1}(x_i)$ will be recalculated iteratively for the estimation of the mean and variance of the variable x_i, which will be applied to update the prior signal distribution parameters.

3.2. Derivation of WDAMP

Since multiple integrals are involved in (11), the complexity of WDBP is too high for practical application, the key derivation of WDAMP in this subsection starts from the simplification of (11). .

Based on Hubbard-Stratonovich transform and Taylor expansion, (11) can be simplified as [6]

$$m_{\mu \to i}(x_i) = \frac{1}{\tilde{Z}^{i \to \mu}} \exp\{-\frac{A_{\mu \to i}}{2} x_i^2 + B_{\mu \to i} x_i\}, \quad (14)$$

where

$$A_{\mu\to i}=\frac{H_{\mu i}^2}{\rho_\mu\sigma^2+\sum_{j\neq i}H_{\mu j}^2 v_{j\to\mu}},\ B_{\mu\to i}=\frac{H_{\mu i}(y_\mu-\sum_{j\neq i}H_{\mu j}a_{j\to\mu})}{\rho_\mu\sigma^2+\sum_{j\neq i}H_{\mu j}^2 v_{j\to\mu}},$$

and $a_{i\to\mu}=\int dx_i x_i m_{i\to\mu}(x_i),\ v_{i\to\mu}=\int dx_i x_i^2 m_{i\to\mu}(x_i)-a_{i\to\mu}^2$.

Substitute (14) into (10) and (13), then

$$m_{i\to\mu}^{t+1}(x_i)=\mathrm{M}\Big(x;\mu_x^t,(\sigma_x^t)^2,\frac{1}{\sum_{\nu=1}^M\rho_\nu A_{\nu\to i}-A_{\mu\to i}},\frac{\sum_{\nu=1}^M\rho_\nu B_{\nu\to i}-B_{\mu\to i}}{\sum_{\nu=1}^M\rho_\nu A_{\nu\to i}-A_{\mu\to i}}\Big) \tag{15}$$

$$m_i^{t+1}(x_i)=[m_i^t(x_i)]^\gamma[\mathrm{M}(x_i;\mu_x^t,(\sigma_x^t)^2,(\Sigma_i^{t+1})^2,R_i^{t+1})]^{1-\gamma} \tag{16}$$

in which we define

$$\mathrm{M}(x;\mu_x,\sigma_x^2,\Sigma^2,R)=\frac{1}{\sqrt{2\pi}\Sigma\hat{Z}}[(1-p)\delta(x)+p\phi(x,\mu_x,\sigma_x^2)]\mathrm{e}^{-(x-R)^2/2\Sigma^2}\,, \tag{17}$$
$$=(1-\lambda)\delta(x)+\lambda\phi(x,\tilde{\mu}_x,\tilde{\sigma}_x^2)$$

$$\lambda=1-\frac{(1-p)}{\sqrt{2\pi}\Sigma\hat{Z}}e^{-R^2/2\Sigma^2},\ \tilde{\sigma}_x^2=\frac{\sigma_x^2\Sigma^2}{\sigma_x^2+\Sigma^2},\ \tilde{\mu}_x=\frac{\sigma_x^2R+\Sigma^2\mu_x}{\sigma_x^2+\Sigma^2}$$

$$\hat{Z}=[(1-p)e^{-R^2/2\Sigma^2}+p\tilde{\sigma}_x e^{(\tilde{\mu}_x^2-G)/2\tilde{\sigma}_x^2}/\sigma_x]/\sqrt{2\pi}\Sigma \tag{18}$$
$$G=(\Sigma^2\mu_x^2+R^2\sigma_x^2)/(\Sigma^2+\sigma_x^2)$$

and

$$\begin{cases}(\Sigma_i^{t+1})^2=\dfrac{1}{\sum_\mu\rho_\mu A_{\mu\to i}},\\ R_i^{t+1}=\dfrac{\sum_\mu\rho_\mu B_{\mu\to i}}{\sum_\mu\rho_\mu A_{\mu\to i}}.\end{cases} \tag{19}$$

Until now, (14) ~ (16) compose a new version of WDBP equations, which have been simplified significantly relative to (10), (11) and (13). However, they are still too complex and can be simplified further.

If we define

$$m_i^t(x_i)=(1-\lambda_i^t)\delta(x_i)+\lambda_i^t\phi(x_i,\mu_i^t,(\sigma_i^2)^t)\,, \tag{20}$$

the second term of (16) can be expressed, based on (17), as

$$\mathrm{M}(x_i;:)=(1-\lambda_\#)\delta(x_i)+\lambda_\#\phi(x_i,\mu_\#,\sigma_\#^2)\,, \tag{21}$$

where parameters $\{\lambda_\#,\sigma_\#^2,\mu_\#\}$ can be determined by (18) explicitly. Substituting (20) and (21) into (16), $m_i^{t+1}(x_i)$ can be approximated as $(1-\lambda_i^{t+1})\delta(x)+\lambda_i^{t+1}\phi(x,\mu_i^{t+1},(\sigma_i^2)^{t+1})$, in which

$$\lambda_i^{t+1}=1-A/Z,\ (\sigma_i^{t+1})^2=\frac{(\sigma_i^t)^2\sigma_\#^2}{(1-\gamma)(\sigma_i^t)^2+\gamma\sigma_\#^2} \tag{22}$$

$$\mu_i^{t+1}=\frac{\gamma\sigma_\#^2\mu_i^t+(1-\gamma)(\sigma_i^t)^2\mu_\#}{(1-\gamma)(\sigma_i^t)^2+\gamma\sigma_\#^2} \tag{23}$$

where $Z=A+\sigma_i^{t+1}(\frac{\lambda_i^t}{\sigma_i^t})^\gamma(\frac{\lambda_\#}{\sigma_\#})^{1-\gamma}\exp\{\frac{(\mu_i^{t+1})^2-G}{2(\sigma_i^{t+1})^2}\}$,

$$A=(\lambda_i^t)^\gamma(1-\lambda_\#)^{1-\gamma}/(\sqrt{2\pi}\sigma_i^t)^\gamma e^{-\gamma(\mu_i^t)^2/2(\sigma_i^t)^2}$$
$$+(1-\lambda_i^t)^\gamma\lambda_\#^{1-\gamma}/(\sqrt{2\pi}\sigma_\#)^{1-\gamma}e^{-(1-\gamma)\mu_\#^2/2\sigma_\#^2}+(1-\lambda_i^t)^\gamma(1-\lambda_\#)^{1-\gamma}$$

$$G=\frac{\gamma\sigma_\#^2(\mu_i^t)^2+(1-\gamma)(\sigma_i^t)^2\mu_\#^2}{\gamma\sigma_\#^2+(1-\gamma)(\sigma_i^t)^2}.$$

If we define the mean and variance of variable x_i as

$$a_i^{t+1}=\int dx_i x_i m_i^{t+1}(x_i) \tag{24}$$

$$v_i^{t+1}=\int dx_i x_i^2 m_i^{t+1}(x_i)-(a_i^{t+1})^2 \tag{25}$$

and define

$$\omega_\mu=\sum_i H_{\mu i}a_{i\to\mu},\ V_\mu=\sum_i H_{\mu i}^2 v_{i\to\mu} \tag{26}$$

(26), (19), (24) and (25) can be simplified, following the procedures (33) ~ (44) in [6], as

$$V_\mu^{t+1}=\sum_i H_{\mu i}^2 v_i^t \tag{27}$$

$$\omega_\mu^{t+1}=\sum_i H_{\mu i}a_i^t-\frac{(y_\mu-\omega_\mu^t)}{\rho_\mu\sigma^2+V_\mu^t}\sum_i H_{\mu i}^2 v_i^t \tag{28}$$

$$(\Sigma_i^{t+1})^2=(\sum_\mu\frac{\rho_\mu H_{\mu i}^2}{\rho_\mu\sigma^2+V_\mu^{t+1}})^{-1} \tag{29}$$

$$R_i^{t+1}=a_i^t+\frac{\sum_\mu[\rho_\mu H_{\mu i}(y_\mu-\omega_\mu^{t+1})/(\rho_\mu\sigma^2+V_\mu^{t+1})]}{\sum_\mu[\rho_\mu H_{\mu i}^2/(\rho_\mu\sigma^2+V_\mu^{t+1})]} \tag{30}$$

$$a_i^{t+1}=\lambda_i^{t+1}\mu_i^{t+1},\ v_i^{t+1}=\lambda_i^{t+1}((\mu_i^{t+1})^2+(\sigma_i^{t+1})^2)-(a_i^{t+1})^2 \tag{31}$$

in which the iteration is started from $t=0$, $V_\mu^0=0$, $v_i^0=1$, $a_i^0=0$ for $\mu\in[M]$ and $i\in[N]$. The related parameters in (31) are calculated based on (22) and (23). Because $\mathbf{H}$ is a column normalized matrix, $H_{\mu i}^2$ in (27) ~ (30) can be approximated as $1/M$. The complexity of the algorithm can be decreased by half with the cost of minor performance loss.

3.2. Parameter Update and Weight Choice

The signal parameters are updated as follows

$$\begin{cases}\mu_x^t=\sum_i a_i^t/K,\\ (\sigma_x^2)^t=\sum_i(v_i^t+(a_i^t)^2)/K\end{cases}, \tag{32}$$

which are initialized as $\mu_x^0=0$ and $(\sigma_x^2)^0=1$. Until now, (27) ~ (32) represent the whole procedures of WDAMP, which will be executed iteratively until the reconstructed signal satisfy some predefined precision requirements or reach the maximum iteration number. (31) will output the final estimated mean and variance of $\mathbf{x}$.

How to find the optimal weight set $\{\rho_\mu\}$ under the constraint of (9) is the key for WDAMP. In this paper, the optimal weights are obtained by trial. WDAMP will be repeated many times, and every time the weights are generated uniformly among $[0,1]$, and normalized to satisfy (9). Calculate the error $\|\mathbf{y}-\mathbf{H}\hat{\mathbf{x}}\|_2$ ($\hat{\mathbf{x}}$ is the reconstructed signal), and the minimum error will correspond to the optimal weights. The theoretical derivation of the optimal weights is our future work.

4. SIMULATION RESULTS

In simulation, SNR is defined as $10\log_{10}(\|\mathbf{y}\|_2^2/(M\sigma^2))$ dB, and mean square error (MSE) is defined as $\|\mathbf{x}-\hat{\mathbf{x}}\|_2/\|\mathbf{x}\|_2$. The maximum iteration for AMP is 50. In WDAMP, 30 trials are used for the weights selection. In the figures below, "AMP" represents normal AMP [6], and "damp-AMP" is the special case of WDAMP when all the weights are 1.

4.1. 1-D Gaussian signal reconstruction

In this test, the non-zero elements of the 1-D signal are zero-mean Gaussian distributed with variance of 1. The length of the signal is 800, the number of non-zero atoms is 200. The damping parameter γ is set to be 0.3. In the first experiment, the signal reconstruction MSE of AMP, damp-AMP and WDAMP are compared for different measurements. SNR is fixed as 25dB. From Figure 1, the adoption of probability damping improves the performance of AMP significantly, and the weighting provides further improvement. This improvement diminishes when the number of measurements is larger than 420.

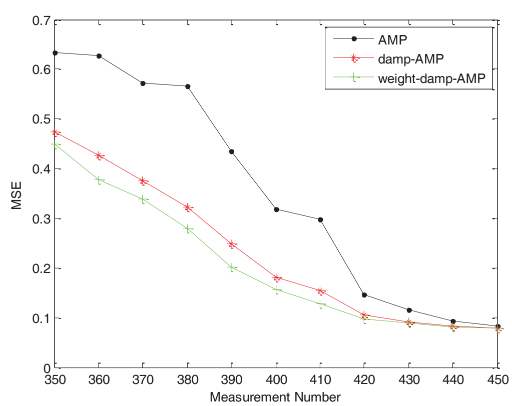

Fig.1. Signal Reconstruction under different M

Figure 3 compared NMSE performance under different SNR for fixed number of measurements (M=380). We see that, when SNE increase from 15dB to 25dB, both the improvements brought by probability damping solely and by weighting-damping jointly increase. WDAMP outperform normal AMP a lot at high SNR.

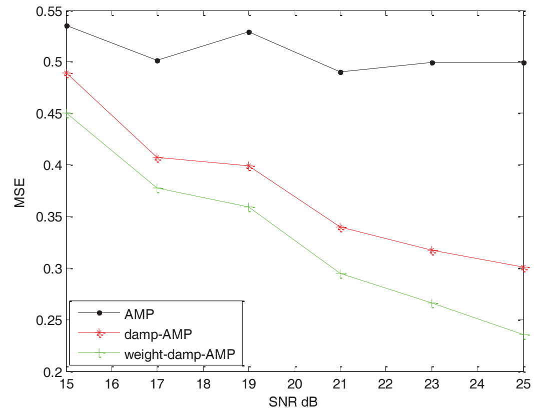

Fig.2. Signal reconstruction under different SNR

4.2. 2-D image reconstruction

For 2-D image reconstruction, 64x64 Mondrian image is used, and level-4 Haar wavelet transform is adopted. The SNR is fixed to be 25 dB. The number of non-zeros is calculated according to the rule that the largest K decomposition coefficients will maintain 99% energy of the total coefficients. From Figure 3, we see that, with the increase of the measurement number, the performance gain brought by weighting mechanism is steady, but the gain brought by damping technique decreases.

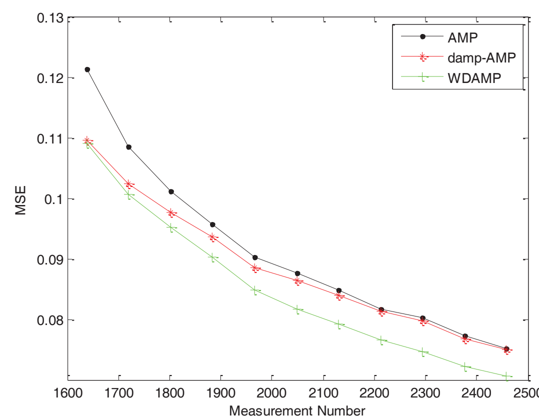

Fig.3. Image reconstruction performance

5. CONCLUSION

In this paper, based on the tree-reweighted loopy belief propagation and probability damping technique, a new weighted-damped LBP is constructed, from which a new version of approximate message passing algorithm (WDAMP) is formulated. Simulation results show that WDAMP outperforms conventional AMP on both 1-D and 2-D signal reconstruction.

6. REFERENCES

[1] D. L. Donoho, “Compressed Sensing,” *IEEE Trans. On Info. Theory*, vol. 52, no. 4, pp. 1289-1306, Apr. 2006.

[2] E. J. Candès, J. Romberg, and T. Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” *IEEE Trans. Info. Theory,* vol. 52, no 2, pp. 489-509, Feb. 2006.

[3] D. L. Donoho, A. Maleki, and A. Montanari, “How to design message passing algorithms for compressed sensing,” 2011 [Online]. Available: http://www.ece.rice.edu/mam15/bpist.pdf

[4] Jonathan S. Yedidia, William T. Freeman, Yair Weiss, “Constructing Free-Energy Approximation and Generalized Belief Propagation Algorithms,” *IEEE Trans. Info. Theory*, vol. 51, pp. 2282-2312, July 2005.

[5] D. L. Donoho, A. Maliki, and A. Montanari, “Message-passing algorithms for compressed sensing,” *Proc. Nat. Acad. Sci.*, vol. 106, no. 45, pp. 18914-18919, 2009.

[6] F. Krzakala, M. M´ezard, F. Sausset, Y. Sun, and L. Zdeborov´a, “Probabilistic reconstruction in compressed sensing: algorithms, phase diagrams, and threshold achieving matrices,” *Journal of Statistical Mechanics: Theory and Experiment*, (2012)P08009, 2012.

[7] Vladimir Kolmogorov, “Convergent tree-reweighted message passing for energy minimization,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 28, pp. 1568-1583, Oct. 2006.

[8] Macro Pretti, “A message-passing algorithm with damping,” *Journal of Statistical Mechanics: Theory and Experiment*, (2005)P11008, 2005.

[9] Henk Wymeersch, Federico Penna and Vladimir Savic, “Uniformly reweighted belief propagation: a factor graph approach,” *IEEE International Symposium on Information Theory*, pp. 2000-2004, July 2011.

[10] Frank R. K., Brendan J. K. and Hans-Andrea L., “Factor graphs and the sum-product algorithm,” *IEEE Trans. Info. Theory*, vol. 47, no. 2, pp. 498-519, Feb. 2001