A Complete Derivation of Sum-product Generalized Approximate Message Passing

Jianyi Yang
Beijing University of Posts and Telecommunications
Email: yangjianyig@bupt.edu.cn

Sept. 2017

Generalized Approximate Message Passing (GAMP) proposed by Sundeep Rangan is a computationally simple algorithm to solve the compressed sensing problem. Compared with original Approximate Message Passing (AMP), GAMP allows non-linear processing. Based on Rangan's work in [1], this article gives a complete derivation of Sum-product GAMP.

I. Model of Compressed Sensing

The problem of compressed sensing is expressed as

$$\min \|x\|_0 \text{, subject to } \boldsymbol{z}=\mathrm{f}\left(\boldsymbol{y}\right)=\mathrm{f}\left(\boldsymbol{Ax}\right), \tag{1}$$

where $\boldsymbol{x}\in\mathbb{C}^{N\times 1}$ is the sparse vector, $\boldsymbol{A}\in\mathbb{C}^{M\times N}$ is the linear mixing matrix with each element $a_{j,i}\sim\mathcal{N}\left(0,\frac{1}{\sqrt{M}}\right)$, $\mathrm{f}\left(\bullet\right)$ is a measuring function, which can be linear or non-linear. The linear mixing matrix $\boldsymbol{A}$ and measured vector $\boldsymbol{z}$ are known by us. We also know that the prior probability density of each element of the sparse vector $x_i\left(i=1,\cdots N\right)$ is $p_{X|Q}\left(x_i\mid q_i\right)$, and the output probability density is $p_{Z|Y}\left(z_j\mid y_j\right)$. The sparse vector $\boldsymbol{x}$ and the output vector $\boldsymbol{y}$ are unknown.

II. Factor Graph Representation

It is known to all that Sum-product AMP is based on the method of minimized mean square error (MMSE), which means the estimation of sparse vector is

$$\hat{\boldsymbol{x}}=\mathrm{E}_{p(\boldsymbol{x}|\boldsymbol{z})}\left[\boldsymbol{x}\mid\boldsymbol{z}\right]=\int\boldsymbol{x}p\left(\boldsymbol{x}\mid\boldsymbol{z}\right), \tag{2}$$

where the conditional probability in (2) is

$$p\left(\boldsymbol{x}\mid\boldsymbol{z}\right)=\frac{1}{Z}\prod_{j=1}^{M}p\left(\mathrm{z}_j\mid\boldsymbol{x}\right)\prod_{i=1}^{N}p\left(x_i\mid q_i\right), \tag{3}$$

where Z is a normalization variable which is uncorrelated with $\boldsymbol{x}$.

Now we transform this problem to multiple single-scale problems, namely

$$\hat{x}_i = \mathrm{E}_{p(x_i|\mathbf{z})}\left[x_i \mid \mathbf{z}\right] = \int x_i p\left(x_i \mid \mathbf{z}\right)\mathrm{d}x_i, i = 1, 2, \cdots N, \tag{4}$$

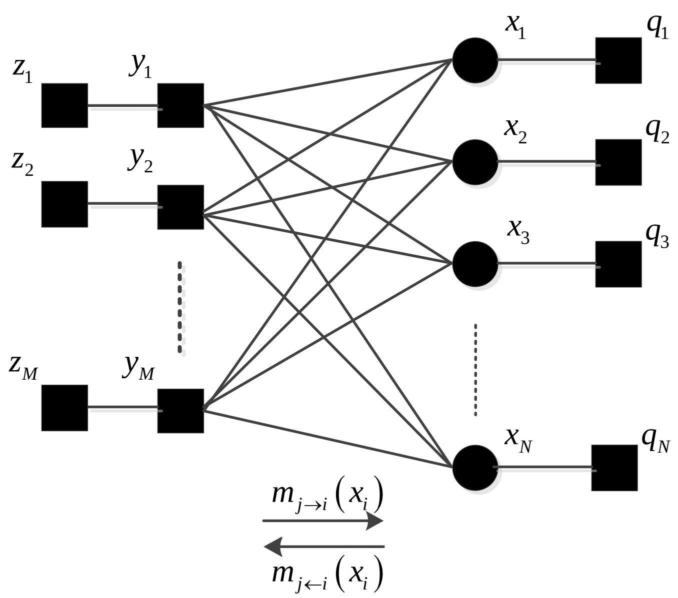

Fig. 1 The Factor Graph Representation of Compressed Sensing Model

where $p\left(x_i \mid \mathbf{z}\right) = \int_{\sim(x_i)} p\left(\mathbf{x} \mid \mathbf{z}\right)$ is a marginal probability. The complexity of computing these marginal probabilities directly is very high, but the sum-product algorithm in factor graphs with cycles is a simple method to solve this problem.

Substitute (3) into marginal probability, we can get

$$\begin{aligned} p\left(x_i \mid \mathbf{z}\right) &= \int_{\sim(x_i)} p\left(\mathbf{x} \mid \mathbf{z}\right) \\ &= \frac{1}{Z}\int_{\sim(x_i)} \prod_{j=1}^{M} p\left(\mathrm{z}_j \mid \mathbf{x}\right) \prod_{i=1}^{N} p\left(x_i \mid q_i\right) \\ &= \frac{1}{Z}\int_{\sim(x_i)} \prod_{j=1}^{M} p\left(z_j \mid y_j\right) \mathcal{X}_{\left\{y_j = (\mathbf{A}\mathbf{x})_j\right\}}\left(\mathbf{x}\right) \prod_{i=1}^{N} p\left(x_i \mid q_i\right) \end{aligned}, \tag{5}$$

where $\mathcal{X}_{\left\{y_j=(\mathbf{A}\mathbf{x})_j\right\}}\left(\mathbf{x}\right)$ is an indicative function, which equals 1 if $\mathbf{x} \in \left\{y_j=(\mathbf{A}\mathbf{x})_j\right\}$ and 0 if $\mathbf{x} \notin \left\{y_j=(\mathbf{A}\mathbf{x})_j\right\}$ 。

The factor graph representation of (5) is shown in Fig.1. The sum-product update rules are

$$m_{j\leftarrow i}^{t+1}\left(x_i\right) \cong p_{X|Q}\left(x_i \mid q_i\right) \prod_{l\neq j} m_{l\to i}^{t}\left(x_i\right) \tag{6}$$

$$m_{j\to i}^{t}\left(x_i\right) \cong \int_{\sim(x_i)} p_{Z|Y}\left(z_j \mid y_j\right) \prod_{l\neq i} m_{j\leftarrow l}^{t}\left(x_l\right) \tag{7}$$

where $m_{j\leftarrow i}^{t+1}\left(x_i\right)$ and $m_{j\to i}^{t}\left(x_i\right)$ are the message in tth iteration, and $\cong$ denotes identity between probability distributions up to a normalization constant.

III. Approximation of Message Passing

Before the approximation, we first give a lemma that is useful in our following derivation.

Lemma 1 Assume that the conditional probability of a random variable x is $p\left(x|r,\tau\right)=\frac{1}{Z\left(\tau,\ r\right)}q\left(x\right)\exp\left(-\frac{1}{2\tau}\left(x-r\right)^2\right)$, where $Z\left(\tau,\ r\right)$ is a normalization variable, and $Z\left(\tau,\ r\right)=\int q\left(x\right)\exp\left(-\frac{1}{2\tau}\left(x-r\right)^2\right)\mathrm{d}x$. If $\hat{x}$ and τ^x are respectively the expectation and the variance of x under $p\left(x|r,\tau\right)$, then $\frac{\partial}{\partial r}\hat{x}=\frac{\tau^x}{\tau}$.

Proof Calculate the derivative of $\hat{x}$, we get

$$
\begin{aligned}
\frac{\partial}{\partial r}\hat{x}=&\frac{1}{Z^2\left(\tau,\ r\right)}\int p_{X|Q}\left(x|q\right)\exp\left(-\frac{1}{2\tau}\left(x-r\right)^2\right)\frac{x-r}{\tau}\mathrm{d}x\bullet\hat{x}Z\left(\tau,\ r\right)\\
&+\frac{1}{Z\left(\tau,\ r\right)}\int xp_{X|Q}\left(x|q\right)\exp\left(-\frac{1}{2\tau}\left(x-r\right)^2\right)\frac{x-r}{\tau}\mathrm{d}x\\
=&-\frac{\hat{x}^2}{\tau}+\frac{r\hat{x}}{\tau}+\frac{\mathrm{E}\left[x^2\right]}{\tau}-\frac{r\hat{x}}{\tau}=\frac{\tau^x}{\tau}
\end{aligned}
$$

A. *From Variable Nodes to Factor Nodes*

Assume that $\hat{x}_{j\leftarrow i}^t$ and $\hat{\tau}_{j\leftarrow i}^t$ are expectation and variance of x_i under the massage form Variable Node *i* to Factor Nodes *j*, namely $m_{j\leftarrow i}^t\left(x_i\right)$. According to Central Limit Theorem, with the knowledge of x_i, y_j follows Gaussian distribution. Here, we additionally made the approximation $\hat{\tau}_{j\leftarrow i}^t\approx\hat{\tau}_i^t$, so we have

$$
p\left(y_j|x_i^t\right)=\mathcal{N}\left(y_j;\hat{p}_{j\leftarrow i}^t+a_{j,i}x_i^t,\tau_i^{t,p}\right)\ , \tag{8}
$$

where

$$
\hat{p}_{j\leftarrow i}^t=\sum_{l\neq i}a_{j,l}\hat{x}_{j\leftarrow l}^t \tag{9}
$$

$$
\tau_{j\leftarrow i}^{t,p}=\sum_{l\neq i}\left|a_{j,l}\right|^2\hat{\tau}_l^t\quad . \tag{10}
$$

In (10), we ignored a term of order $O\left(\left|a_{j,i}\right|^2\right)$.

If we define the function $\mathrm{H}\left(p,z,\tau\right)$ as

$$\mathrm{H}\left(p,z,\tau\right)=\int p_{Z|Y}\left(z\,|\,y\right)\mathcal{N}\left(y;\hat{p},\tau\right)\mathrm{d}\,y\,, \tag{11}$$

according to (7), we have

$$\begin{aligned} m_{j\to i}^{t}\left(x_i\right) &\cong \int_{\sim(x_i)} p_{Z|Y}\left(z_j\,|\,y_j\right)\prod_{l\neq i} m_{j\leftarrow l}^{t}\left(x_l\right)\mathrm{d}\,\boldsymbol{x} \\ &\overset{(a)}{=}\int p_{Z|Y}\left(z_j\,|\,y_j\right)p\left(y_j\,|\,x_i\right)\mathrm{d}\,y_j \\ &=\mathrm{H}\left(\hat{p}_{j\leftarrow i}^{t}+a_{j,i}x_i\,,z_j,\tau_{j\leftarrow i}^{t,p}\right) \end{aligned} \tag{12}$$

where (a) comes from replacing $\boldsymbol{x}$ with y_j.

If we define $\hat{p}_j^t$ and $\tau_j^{t,p}$ as

$$\hat{p}_j^t=\sum_{l\neq i}a_{j,l}\hat{x}_{j\leftarrow l}^{t}+a_{j,i}\hat{x}_i^t\quad, \tag{13}$$

$$\tau_j^{t,p}=\sum_{l}\left|a_{j,l}\right|^2\hat{\tau}_l^t\quad, \tag{14}$$

then it follows from (9) and (10) that

$$\hat{p}_{j\leftarrow i}^{t}=\hat{p}_j^t\text{-}a_{j,i}\hat{x}_i^t \tag{15}$$

$$\tau_{j\leftarrow i}^{t,p}=\tau_j^{t,p}-\left|a_{j,i}\right|^2\hat{\tau}_i^t\approx\tau_j^{t,p} \tag{16}$$

We neglect a term of order $O\left(\left|a_{j,i}\right|^2\right)$ in (16). Now (12) can be rewritten as

$$m_{j\to i}^{t}\left(x_i\right)\ \cong\mathrm{H}\left(\hat{p}_j^t+a_{j,i}\left(x_i\text{ -}\hat{x}_i^t\right),z_j,\tau_j^{t,p}\right) \tag{17}$$

Next, we approximate $m_{j\to i}^{t}\left(x_i\right)$ by a 2nd-order Taylor expansion:

$$\begin{aligned} m_{j\to i}^{t}\left(x_i\right) &\cong \exp\left(\log\left(\mathrm{H}\left(\hat{p}_j^t+a_{j,i}\left(x_i\text{ -}\hat{x}_i^t\right),z_j,\tau_j^{t,p}\right)\right)\right) \\ &\overset{(a)}{=}\exp\left(s_j^t a_{j,i}\left(x_i-\hat{x}_i^t\right)\text{-}\frac{\tau_j^{s,t}}{2}a_{j,i}^2\left(x_i-\hat{x}_i^t\right)^2+\mathrm{C}\right) \\ &=\exp\left(\left(s_j^t a_{j,i}+\tau_j^{s,t}a_{j,i}^2\hat{x}_i^t\right)x_i\text{-}\frac{\tau_j^{s,t}}{2}a_{j,i}^2 x_i^{\,2}+\mathrm{C}\right) \end{aligned} \tag{18}$$

where (a) comes from 2-order Taylor expansion of $\log\left(m_{j\to i}^{t}\left(x_i\right)\right)$ at $\hat{x}_i^t$ and C does not depend on x_i. If we define the output function as

$$g_{\text{out}}\left(\hat{p}, z, \tau^{p}\right) \equiv \frac{\partial}{\partial \hat{p}} \log \mathrm{H}\left(\hat{p}, z, \tau^{p}\right) \tag{19}$$

then s_j^t and $\tau_j^{s,t}$ can be expressed as

$$s_j^t = g_{\text{out}}\left(\hat{p}_j^t, z_j, \tau_j^{t,p}\right) \tag{20}$$

$$\tau_j^{s,t} = -\frac{\partial}{\partial \hat{p}} g_{\text{out}}\left(\hat{p}_j^t, z_j, \tau_j^{t,p}\right) \tag{21}$$

Substituting (11) into (19), we can further simplify the output function as

$$\begin{aligned} g_{\text{out}}\left(\hat{p}, z, \tau^{p}\right) &= \frac{\partial}{\partial \hat{p}} \log \mathrm{H}\left(\hat{p}, z, \tau^{p}\right) \\ &= \frac{1}{\mathrm{H}\left(\hat{p}, z, \tau^{p}\right)} \int \frac{y-\hat{p}}{\tau^{p}} p_{Z|Y}\left(z \mid y\right) \frac{1}{\sqrt{2\pi\tau^{p}}} \exp\left(\frac{1}{2\tau^{p}}\left(y-\hat{p}\right)^{2}\right) \mathrm{d}\, y \\ &= \frac{\mathrm{E}\left[y \mid \hat{p}, z, \tau^{p}\right]-\hat{p}}{\tau^{p}} \end{aligned}$$

According to Lemma 1, we can get：

$$-\frac{\partial}{\partial \hat{p}} g_{\text{out}}\left(\hat{p}, z, \tau^{p}\right) = \frac{\tau^{p} - \operatorname{var}\left(y \mid \hat{p}, z, \tau^{p}\right)}{\left(\tau^{p}\right)^{2}} 。$$

B. *From Factor Nodes to Variable Nodes*

Substituting (18) into (6), we obtain

$$\begin{aligned} m_{j\leftarrow i}^{t+1}\left(x_i\right) &\cong p_{X|Q}\left(x_i \mid q_i\right) \prod_{l\neq j} m_{l\rightarrow i}^{t}\left(x_i\right) \\ &= p_{X|Q}\left(x_i \mid q_i\right) \exp\left(\sum_{l\neq j}\left(s_l^t a_{l,i} + \tau_l^{s,t} a_{l,i}^2 \hat{x}_i^t\right) x_i - \sum_{l\neq j} \frac{\tau_l^{s,t}}{2} a_{l,i}^2 x_i^2 + \mathrm{C}\right) \\ &\overset{(a)}{=} \frac{1}{Z\left(\tau_{j\leftarrow i}^{r,t},\ \hat{r}_{j\leftarrow i}^{t}\right)} p_{X|Q}\left(x_i \mid q_i\right) \exp\left(-\frac{1}{2\tau_{j\leftarrow i}^{r,t}}\left(x_i - \hat{r}_{j\leftarrow i}^{t}\right)^2\right) \end{aligned} \tag{22}$$

where $Z\left(\tau_{j\leftarrow i}^{r,t},\ \hat{r}_{j\leftarrow i}^{t}\right) = \int p_{X|Q}\left(x_i \mid q_i\right) \exp\left(-\frac{1}{2\tau_{j\leftarrow i}^{r,t}}\left(x_i - \hat{r}_{j\leftarrow i}^{t}\right)^2\right) \mathrm{d}x_i$ is a normalization

variable uncorrelated with x_i, and

$$\tau_{j\leftarrow i}^{r,t} = 1 / \sum_{l\neq j} a_{l,i}^2 \tau_l^{s,\mathrm{t}} \tag{23}$$

$$\begin{aligned} \hat{r}_{j\leftarrow i}^{t} &= \tau_{j\leftarrow i}^{r,t} \sum_{l\neq j}\left(s_l^t a_{l,i} + \tau_l^{s,t} a_{l,i}^2 \hat{x}_i^t\right) \\ &= \hat{x}_i^t + \tau_{j\leftarrow i}^{r,t} \sum_{l\neq j} s_l^t a_{l,i} \end{aligned} \tag{24}$$

Now if we define the input function as

$$g_{\text{in}}(r,q,\tau)=\int x\frac{1}{Z(\tau,r)}p_{X|Q}(x\,|\,q)\exp\left(-\frac{1}{2\tau}(x-\hat{r})^2\right)\mathrm{d}\,x \quad , \tag{25}$$

the estimation of $x_{j\leftarrow i}$ can be updated as

$$\begin{aligned}\hat{x}_{j\leftarrow i}^{t+1} &= \mathrm{E}_{m_{j\leftarrow i}^{t+1}}\left[x_i\,|\,\hat{r}_{j\leftarrow i}^{t},q_i,\tau_{j\leftarrow i}^{r,t}\right]\\ &=\int x_i\frac{1}{Z\left(\tau_{j\leftarrow i}^{r,t},\ \hat{r}_{j\leftarrow i}^{t}\right)}p_{X|Q}(x_i\,|\,q_i)\exp\left(-\frac{1}{2\tau_{j\leftarrow i}^{r,t}}\left(x_i-\hat{r}_{j\leftarrow i}^{t}\right)^2\right)\mathrm{d}\,x_i\\ &= g_{\text{in}}\left(\hat{r}_{j\leftarrow i}^{t},q_i,\tau_{j\leftarrow i}^{r,t}\right)\end{aligned} \tag{26}$$

The last step of this derivation is the estimation of $\hat{x}_i^{t+1}$. We know that

$$\begin{aligned} m_i^{t+1}(x_i) &\cong p_{X|Q}(x_i\,|\,q_i)\prod_l m_{l\to i}^{t}(x_i)\\ &=\frac{1}{Z\left(\tau_i^{r,t},\ \hat{r}_i^{t}\right)}p_{X|Q}(x_i\,|\,q_i)\exp\left(-\frac{1}{2\tau_i^{r,t}}\left(x_i-\hat{r}_i^{t}\right)^2\right)\end{aligned} \tag{27}$$

where

$$\tau_i^{r,t}=1/\left(\sum_l a_{l,i}^2\tau_l^{s,t}\right) \tag{28}$$

$$\hat{r}_i^{t}=\hat{x}_i^{t}+\tau_i^{r,t}\sum_l s_l^t a_{l,i} \quad . \tag{29}$$

So, the estimation of $\hat{x}_i^{t+1}$ is

$$\hat{x}_i^{t+1}=g_{\text{in}}\left(\hat{r}_i^{t},q_i,\tau_i^{r,t}\right) \tag{30}$$

In order to find the relationship between $\hat{x}_i^{t+1}$ and $\hat{x}_{j\leftarrow i}^{t+1}$, we first make an approximation by neglecting a term of order $O\left(a_{j,i}^2\right)$. That is

$$\tau_{j\leftarrow i}^{r,t}=1/\left(\sum_l a_{l,i}^2\tau_l^{s,t}-a_{j,i}^2\tau_j^{s,t}\right)\approx 1/\left(\sum_l a_{l,i}^2\tau_l^{s,t}\right)=\tau_i^{r,t} \tag{31}$$

$$\hat{r}_{j\leftarrow i}^{t}\approx\hat{x}_i^{t}+\tau_i^{r,t}\sum_{l\neq j}s_l^t a_{l,i}=\hat{r}_i^{t}-\tau_i^{r,t}s_j^t a_{\mathrm{j},i} \tag{32}$$

Substituting (29) and (30) into (26), we have

$$\begin{aligned}\hat{x}_{j\leftarrow i}^{t+1}&=g_{\text{in}}\left(\hat{r}_i^{t}-\tau_i^{r,t}s_j^t a_{j,i},q_i,\tau_i^{r,t}\right)\\ &\overset{\text{(a)}}{\approx}\hat{x}_i^{t+1}-s_j^t a_{j,i}D_i^{t+1}\end{aligned} \quad , \tag{33}$$

where (a) is the 1st-order Taylor serials at s_j^t. So $D_i^{t+1}=\tau_i^{r,t}\frac{\partial}{\partial\hat{r}_i^t}g_{\text{in}}\left(\hat{r}_i^{t},q_i,\tau_i^{r,t}\right)$.

According to *Lemma 1*, $\frac{\partial}{\partial \hat{r}_i^t} g_{\text{in}}\left(\hat{r}_i^t, q_i, \tau_i^{r,t}\right) = \frac{\text{var}\left(x_i^{t+1} \mid \hat{r}_i^t, q_i, \tau_i^{r,t}\right)}{\tau_i^{r,t}} = \frac{\tau_i^{t+1}}{\tau_i^{r,t}}$, Thus,

$$D_i^{t+1} = \tau_i^{r,t} \frac{\partial}{\partial \hat{r}_i^t} g_{\text{in}}\left(\hat{r}_i^t, q_i, \tau_i^{r,t}\right) = \tau_i^{r,t} \frac{\tau_i^{t+1}}{\tau_i^{r,t}} = \tau_i^{t+1}, \tag{34}$$

namely:

$$\hat{x}_{j\leftarrow i}^{t+1} \approx \hat{x}_i^{t+1} - s_j^t a_{j,i} \tau_i^{t+1} \tag{35}$$

After substituting (34) into (13) and neglecting a term of $O\left(a_{j,i}^2\right)$, we can get

$$\hat{p}_j^t \approx \sum_i \left(a_{j,i}\hat{x}_i^t - a_{j,i}^2 \tau_i^t s_j^{t-1}\right) = \sum_i a_{j,i}\hat{x}_i^t - \tau_j^{p,t} s_j^{t-1} \tag{36}$$

Combining (36)(14)(20)(21)(28)(29)(30)(34), we can get GAMP algorithm.

IV. GAMP Algorithm Flow

GAMP
Input $\boldsymbol{A} \in \mathbb{C}^{M\times N}$, $\boldsymbol{q}$, $\boldsymbol{z}$, $p_{X\mid Q}\left(x_i \mid q_i\right)$, $p_{Z\mid Y}\left(z \mid y\right)$
Output $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{y}}$
Initialization $t=1; x_i^1 = \mathrm{E}_{X\mid Q}\left[x_i \mid q_i\right]$, $\tau_i^1 = \text{var}_{X\mid Q}\left[x_i \mid q_i\right], i = 1,2,\cdots,N$;
$\tau_j^{p,0} = 1, p_j^0 = z_j, s_j^0 = g_{\text{out}}\left(p_j^0, z_j, \tau_j^{p,0}\right) \quad j = 1,2,\cdots,M$
Repeat until a predefined number of iterations or other termination conditions are satisfied
for $i = 1,2,\cdots,N$
Linear Output
$\tau_j^{t,p} = \sum_l \left\lvert a_{j,l}\right\rvert^2 \hat{\tau}_l^t$
$\hat{p}_j^t = \sum_i a_{j,i}\hat{x}_i^t - \tau_j^{p,t} s_j^{t-1}$
$y_j^t = \sum_i a_{j,i}\hat{x}_i^t$
Non-linear Output
$s_j^t = g_{\text{out}}\left(\hat{p}_j^t, z_j, \tau_j^{t,p}\right)$
$\tau_j^{s,t} = -\frac{\partial}{\partial \hat{p}} g_{\text{out}}\left(\hat{p}_j^t, z_j, \tau_j^{t,p}\right)$
end

for $j = 1, 2, \cdots, M$

Linear Input

$$\tau_i^{r,t} = 1 / \left(\sum_l a_{l,i}^2 \tau_l^{s,\mathrm{t}} \right)$$

$$\hat{r}_i^t = \hat{x}_i^t + \tau_i^{r,t} \sum_l s_l^t a_{l,i}$$

Non-linear Input

$$\hat{x}_i^{t+1} = g_{\mathrm{in}}\left(\hat{r}_i^t, q_i, \tau_i^{r,t}\right)$$

$$\tau_i^{t+1} = \tau_i^{r,t} \frac{\partial}{\partial \hat{r}_i^t} g_{\mathrm{in}}\left(\hat{r}_i^t, q_i, \tau_i^{r,t}\right)$$

end

[1] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," preprint, 2012. Available: http://arxiv.org/ abs/1010.5141v2.