Concise Derivation for Generalized Approximate Message Passing Using Expectation Propagation

Qiuyun Zou, Haochuan Zhang, Chao-Kai Wen, Shi Jin, and Rong Yu

***Abstract*—Generalized approximate message passing (GAMP) is an efficient algorithm for the estimation of independent identically distributed random signals under generalized linear model. The sum-product GAMP has long been recognized as an approximate implementation of the sum-product loopy belief propagation. In this letter, we propose to view the message passing in a new perspective of expectation propagation (EP). Comparing with the previous methods that were based on Taylor expansions, the proposed EP method could unify the derivations for the real and the complex GAMP, with a difference only in the setup of Gaussian densities.**

***Index Terms*—Generalized approximate message passing, expectation propagation, Gaussian reproduction property.**

I. Introduction

APPROXIMATE message passing (AMP) was proposed in [1] to recover sparse signals from linear measurements that were corrupted by additive white Gaussian noise. Owing to its efficiency (roughly of quadratic complexity in the problem size) and effectiveness (asymptotically exact in certain cases), AMP has since found various applications in engineering. After AMP, an extension termed generalized AMP (GAMP) [2] was developed, which considered a much broader scope, the generalized linear model (GLM), allowing the componentwise output mapping to take an arbitrary form. Motivated by GAMP, the so-called BiGAMP was proposed by [3]. BiGAMP further extended the linear inverse scope to bi-linear inverse, in which both the input signal and the measurement matrix are to be estimated. In these early works, signals were considered to be real. However, for applications like wireless communications, magnetic resonance imaging, and radar imaging, a complex form would be more desired. In this context, complex AMP and GAMP were proposed in [4]–[7] among others. Although the sum-product AMP/GAMP has long been recognized as an approximate implementation of the sum-product loopy belief propagation (LBP), we find that, following this direction, the derivation for the complex case may face some new challenges in the expansion and justification of Taylor series.

In an attempt to recover complex GAMP without Taylor expansions, this letter offers a new perspective of viewing the message passing as in an expectation propagation (EP) [8] manner. To be specific, each variable and factor node in a factor graph (FG) updates its message according to scalar EP rules. The resulting messages are then in a Gaussian form, and after certain approximation to the means and variances, the complex sum-product GAMP could be recovered. It is worthy of noting that the idea of using EP was introduced by [9]–[11] to derive real and complex AMP. The method there, however, relied on a key assumption of Gaussian likelihood, which prohibits a direct extension to complex GAMP. To remove such a limitation, we borrow an idea from [3], where a key step is to apply central limit theorem in the Gaussian approximation of one part of the message. We also note that there exists other ways to recover the complex GAMP. For instance, GAMP can be treated as a degenerate Gr-AMP [12], which admits also a concise derivation (but be aware their factor graph was different from ours as they required additional nodes $\delta(\cdot)$). As another example, complex GAMP could be approximated by expanding the complex model into pairs of real variables, then applying the HyGAMP [13] with (real, imaginary) pairs in the prior and likelihood, leading to a GAMP-like algorithm that passes 2×2 covariance matrices, and finally taking the SHyGAMP [14] approach to approximate those covariance matrices by scaled identities.

Comparing with these methods, our derivation contributes through the following aspects: first, offering a new perspective to view the GAMP message passing in an EP manner; second, extending prior work to establish a more general link that connects EP and GAMP.

II. System Model and GAMP Recap

System Model: Consider the generalized linear model [2]

$$\boldsymbol{x} \rightarrow \boxed{\boldsymbol{H}} \xrightarrow{\boldsymbol{z}} \boxed{p(\boldsymbol{y}|\boldsymbol{z})} \rightarrow \boldsymbol{y} \tag{1}$$

where the input $\boldsymbol{x} \in \mathbb{C}^N$, randomly drawn from the distribution $p(\boldsymbol{x}) = \prod_i p(x_i)$, is *linearly mixed* by a deterministic measurement matrix $\boldsymbol{H} \in \mathbb{C}^{M \times N}$ to obtain $\boldsymbol{z} = \boldsymbol{H}\boldsymbol{x}$. Then $\boldsymbol{z}$ is passed through a *componentwise mapping* channel whose transition probability function is $p(\boldsymbol{y}|\boldsymbol{z}) = \prod_{a=1}^{M} p(y_a|z_a)$. The final

Manuscript received September 8, 2018; accepted October 14, 2018. Date of publication October 18, 2018; date of current version November 1, 2018. The work of Q. Zou and H. Zhang was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 61501127; in part by the NSF of Guangdong under Grant 2016A030313705; and in part by the Special Fund for Applied Science and Technology of Guangdong under Grant 2015B010129001, Grant 2014B090907010, and Grant 2015B010106010. The work of C.-K. Wen was supported by the ITRI in Hsinchu, Taiwan. The work of S. Jin was supported by the NSFC for Distinguished Young Scholars of China under Grant 61625106. The work of R. Yu was supported by the NSFC under Grant 61422201, Grant 61370159, and Grant 61503083. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Qing Ling. *(Corresponding author: Haochuan Zhang.)*

Q. Zou, H. Zhang, and R. Yu are with Guangdong University of Technology, Guangzhou 510006, China (e-mail: qiuyunzou@qq.com; haochuan.zhang@qq.com; yurong@ieee.org).

C.-K. Wen is with the National Sun Yat-sen University, Kaohsiung 80424, Taiwan (e-mail: chaokai.wen@mail.nsysu.edu.tw).

S. Jin is with Southeast University, Nanjing 210096, China (e-mail: jinshi@seu.edu.cn).

Color versions of one or more of the figures in this letter are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2018.2876806

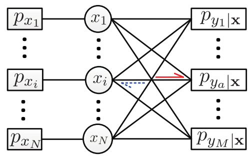

Fig. 1. Factor graph associated with the marginalization of interest.

observation is $\boldsymbol{y} \in \mathbb{C}^M$. The GLM degenerates to standard linear model when $p(y_a|z_a)$ is particularized to $y_a = z_a + n_a$ with n_a being a Gaussian noise. We assume the measurement matrix $\boldsymbol{H}$ to be sub-Gaussian and large, with $N, M \to \infty$ and fixed M/N. Moreover, the prior $p(\boldsymbol{x}) = \prod_{i=1}^N p(x_i)$, and the likelihood is $p(\boldsymbol{y}|\boldsymbol{x}) = \prod_{a=1}^M p(y_a|\boldsymbol{x})$. We are interested in the minimum mean square error (MMSE) estimate here [15]: $\hat{m}_i = \mathbb{E}_{x_i|\boldsymbol{y}}[x_i]$, for $i = 1, \ldots, N$, where the expectation is taken w.r.t. the marginal posterior $p(x_i|\boldsymbol{y}) = \int_{\boldsymbol{x}_{\setminus i}} p(\boldsymbol{x}|\boldsymbol{y}) \mathrm{d}\boldsymbol{x}_{\setminus i}$, with $\boldsymbol{x}_{\setminus i}$ denoting all elements in $\boldsymbol{x}$ except the i-th one, and $p(\boldsymbol{x}|\boldsymbol{y})$ denoting the joint posterior $p(\boldsymbol{x}|\boldsymbol{y}) = \frac{p(\boldsymbol{x})p(\boldsymbol{y}|\boldsymbol{x})}{p(\boldsymbol{y})} \propto [\prod_{i=1}^N p(x_i)][\prod_{a=1}^M p(y_a|\boldsymbol{x})]$. It is worthy of noting that the goal of the sum-product GAMP [2] is to approximate the marginal posterior density by a (scaled) product of the prior and a Gaussian distribution. In this context, the MMSE estimate (i.e., the posterior mean) is a byproduct. To see how the GAMP approximation could be obtained, our review starts from FG and LBP, then moves on to the relaxed belief propagation (RBP), and finally applies the Thouless-Anderson-Palmer (TAP) approach to recover GAMP.

FG & LBP: A FG is a bipartite graph representing the factorization of a density function. Fig. 1 illustrates the FG associated with the marginalization problem of interest here. In the FG, circles (termed variable nodes) represent the variables x_i's, and squares (termed factor nodes) represent the density functions (prior and likelihood). For a FG with cycles (like the one here), a sum-product LBP approximates [1] the marginal posteriors by passing messages between different nodes iteratively. Two kinds of messages are defined: one, denoted by $\mu_{i\leftarrow a}^t(x_i)$, represents the message from the a-th factor node $p(y_a|\boldsymbol{x})$ to the i-th variable node x_i at iteration t, and the other, denoted by $\mu_{i\to a}^t(x_i)$, represents the message in the opposite direction. The message updating rules of the sum-product LBP are as follows $(t = 1, \ldots, T)$:

$$\mu_{i\leftarrow a}^{t+1}(x_i) \propto \int p(y_a|\boldsymbol{x}) \prod_{j\neq i} \mu_{j\leftarrow a}^t(x_j) \mathrm{d}\boldsymbol{x}_{\setminus i} \tag{2a}$$

$$\mu_{i\to a}^{t+1}(x_i) \propto p(x_i) \prod_{b\neq a} \mu_{i\leftarrow b}^{t+1}(x_i) \tag{2b}$$

with the output marginal being $\mu_{i\leftarrow a}^T(x_i) \cdot \mu_{i\to a}^T(x_i)$.

RBP: To reduce the complexity and overhead of LBP in tracking messages, RBP [16], [17] was proposed, which rewrites the iterative equations of messages (density functions) as iteration of the means and variances of those densities, in the large system size limit. To see this, one first approximates the message $\mu_{i\leftarrow a}^t(x_i)$ as Gaussian density, and then proves that the other message $\mu_{i\to a}^t(x_i)$ also contributes only through its mean and variance.

TAP: Still, RBP needs to keep track of $M \times N$ messages. The number soars up quickly as $M, N \to \infty$. To reduce it, first-order Taylor expansions on the means and variances are taken, yielding only $M + N$ new approximates, see e.g., [2, (118)–(120)] and [18, (160)-(161)]. Substituting the new approximates back to the RBP and neglect certain high-order terms, a new iteration is formed, which is the desired GAMP algorithm [2]. Because the fixed point version of the new iteration dates back to the work of Thouless, Anderson, and Palmer [19] on mean field spin glass, this method is also called the TAP approach.

Limitations: We reviewed above the LBP-based method developed by [2] for GAMP's derivation in real. A direct extension to complex is, however, facing new challenges arising from Taylor expansions on complex-valued functions of complex arguments. Below is an example of this [20]:

$$f(c+\triangle) - f(c) = \sum_{p=1}^{m} \sum_{n=0}^{p} \frac{(\triangle^*)^n \triangle^{p-n}}{n!(p-n)!} \left. \frac{\partial^p f(z)}{\partial z^{p-n} \partial (z^*)^n} \right|_{z=c} + |\triangle|^m \varepsilon(\triangle) \tag{3}$$

where $(\cdot)^*$ denotes the conjugate operation, and f is a complex to complex mapping. Comparing to Taylor expansion in real, the above equation needs to handle the following challenging tasks: computing conjugate derivatives, combining conjugate variables, and proving the negligibility of complex residuals, among others. Although it is possible to treat the complex argument as a two-dimensional vector and apply the vector-form Taylor expansion, the challenges above won't simply disappear. They just take a new form and will be reflected in the cross differential terms. These cross terms did not exist in the discussion for real, so all analysis and justifications similar to [3, Table II] need to start over again.

Algorithm 1: Complex GAMP [5].

1. Initialize: $t = 1$, $s_a^0 = 0$, $\hat{m}_i^1 = \int x_i p(x_i) \mathrm{d}x_i$, $\hat{v}_i^1 = \int |x_i - \hat{m}_i^1|^2 p(x_i) \mathrm{d}x_i$.
2. Iteration (for $t = 1, \ldots, T$)

$$V_a^t = \sum_i |H_{ai}|^2 \hat{v}_i^t \tag{4a}$$

$$Z_a^t = \sum_i H_{ai} \hat{m}_i^t - s_a^{t-1} V_a^t \tag{4b}$$

$$\hat{z}_a^t = \mathbb{E}_{\zeta_a^t}[\zeta_a^t] \tag{4c}$$

$$\tilde{v}_a^t = \mathbb{V}_{\zeta_a^t}[\zeta_a^t] \tag{4d}$$

$$s_a^t = (\hat{z}_a^t - Z_a^t)/V_a^t \tag{4e}$$

$$\tau_a^t = (V_a^t - \tilde{v}_a^t)/(V_a^t)^2 \tag{4f}$$

$$\Sigma_i^{t+1} = \left[\sum_a |H_{ai}|^2 \tau_a^t\right]^{-1} \tag{4g}$$

$$R_i^{t+1} = \hat{m}_i^t + \Sigma_i^{t+1} \sum_a H_{ai}^* s_a^t \tag{4h}$$

$$\hat{m}_i^{t+1} = \mathbb{E}_{\xi_i^{t+1}}[\xi_i^{t+1}] \tag{4i}$$

$$\hat{v}_i^{t+1} = \mathbb{V}_{\xi_i^{t+1}}[\xi_i^{t+1}] \tag{4j}$$

3. Output: $(\hat{m}_i^{T+1}, \hat{v}_i^{T+1})$ as (estimate, MSE)

---

[1] Noted that LBP only approximates the marginal posterior, and sometimes this approximation could be bad. Also, LBP itself is tractable only in two cases: i) when all priors or likelihoods are Gaussian, or, ii) when all priors and likelihoods are discrete. In any other case, the complexity of the exact LBP will grow exponentially.

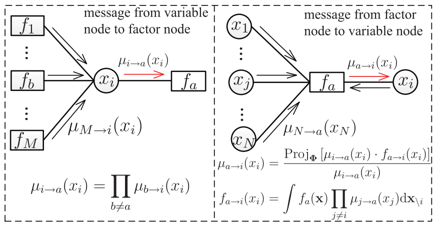

Fig. 2. EP rules for updating messages [11].

III. New Derivation

In this section we provide a new way to recover complex GAMP by viewing the message passing as in an EP manner. We present first in Algorithm 1 the complex GAMP result of [5]. Then, we introduce some basics about EP and show how it could be incorporated. Finally, we derive complex GAMP step by step, starting from the EP message passing.

EP Message Updating: EP [8] approximates the cumulants of the marginal posterior by replacing, in an alternating way, some part of the factorable distribution with Gaussian densities. EP is closely related to Kullback-Leibler (KL) divergence, which measures the "difference" from one distribution to another. Let $p(x)$ and $q(x)$ denote two densities, then the KL divergence from $p(x)$ to $q(x)$ is denoted as $\mathcal{D}[q(x)\|p(x)]$. For EP, the projection of one density onto the Gaussian family is a key concept, which is denoted as:[2]

$$\mathrm{Proj}[q(x)] \triangleq \arg\min_{p\in\Phi} \mathcal{D}[q(x)\|p(x)] = \mathcal{N}(x|m,v), \tag{5}$$

where Φ is the family of Gaussian densities, and

$$m = \int xq(x)\,\mathrm{d}x, \qquad v = \int |x-m|^2 q(x)\,\mathrm{d}x. \tag{6}$$

General rules for EP's message updating are given in Fig. 2. Let all nodes in the FG of Fig. 1 performs these EP operations, their message updating then could become [11]:

$$\mu_{i\leftarrow a}^{t+1}(x_i) \propto \frac{\mathrm{Proj}\left[\mu_{i\to a}^{t}(x_i)\cdot f_{i\leftarrow a}^{t}(x_i)\right]}{\mu_{i\to a}^{t}(x_i)} \tag{7}$$

$$\mu_{i\leftarrow i}^{t+1}(x_i) \propto \prod_{b} \mu_{i\leftarrow b}^{t+1}(x_i) \tag{8}$$

$$\mu_{i\to i}^{t+1}(x_i) \propto \frac{\mathrm{Proj}\left[p(x_i)\cdot \mu_{i\leftarrow i}^{t+1}(x_i)\right]}{\mu_{i\leftarrow i}^{t+1}(x_i)} \tag{9}$$

$$\mu_{i\to a}^{t+1}(x_i) \propto \mu_{i\to i}^{t+1}(x_i)\cdot \prod_{b\neq a}\mu_{i\leftarrow b}^{t+1}(x_i) \tag{10}$$

where $\mu_{i\to i}^{t}(x_i)$ is the message from p_{x_i} to x_i at iteration t, $\mu_{i\leftarrow i}^{t}(x_i)$ is that in the opposite direction, $\mu_{i\to a}^{t}(x_i)$ is the message from x_i to $p(y_a|\boldsymbol{x})$, $\mu_{i\leftarrow a}^{t}(x_i)$ is that in the opposite direction, and $f_{i\leftarrow a}^{t}(x_i) \triangleq \int p(y_a|\boldsymbol{x})\prod_{j\neq i}\mu_{j\to a}^{t}(x_j)\mathrm{d}\boldsymbol{x}_{\setminus i}$. These rules could be condensed into the following form:

$$\mu_{i\leftarrow a}^{t+1}(x_i) \propto \frac{\mathrm{Proj}\left[\mu_{i\to a}^{t}(x_i)\cdot f_{i\leftarrow a}^{t}(x_i)\right]}{\mu_{i\to a}^{t}(x_i)} \tag{11a}$$

$$\mu_{i\to a}^{t+1}(x_i) \propto \frac{\mathrm{Proj}[p(x_i)\cdot \prod_b \mu_{i\leftarrow b}^{t+1}(x_i)]}{\mu_{i\leftarrow a}^{t+1}(x_i)} \tag{11b}$$

with $\mu_{i\leftarrow a}^{T}(x_i)\cdot \mu_{i\to a}^{T}(x_i)$ being the marginal output. Different from LBP, all messages here in the EP-based method are Gaussian densities.

Step-by-Step Derivation: We start from (11a).

Step 1: We simplify $f_{i\leftarrow a}^{t}(x_i)$ in (11a). The challenge here comes from the integral, and our solution is to apply central limit theorem similar to [3, (13)–(19)]. We rewrite first:

$$f_{i\leftarrow a}^{t}(x_i) \propto \int p\left(y_a \,|\, \overbrace{H_{ai}x_i + \sum_{j\neq i} H_{aj}x_j}^{z_a}\right) \prod_{j\neq i}\mu_{j\to a}^{t}(x_j)\mathrm{d}\boldsymbol{x}_{\setminus i}$$

We then associate x_j with a random variable (RV) $\xi_{j\to a}^{t}$ that follows $\mu_{j\to a}^{t}(x_j)$, and associate z_a with another RV η_a^t. For η_a^t, we have $\eta_a^t = H_{ai}x_i + \sum_{j\neq i} H_{aj}\xi_{j\to a}^{t}$. Given a sufficiently large N, the RV η_a^t then follows a Gaussian distribution (central limit theorem) with mean and variance as

$$\mathbb{E}[\eta_a^t] = H_{ai}x_i + Z_{i\leftarrow a}^{t}, \qquad Z_{i\leftarrow a}^{t} \triangleq \sum_{j\neq i} H_{aj} m_{j\to a}^{t},$$

$$\mathbb{V}[\eta_a^t] = V_{i\leftarrow a}^{t}, \qquad V_{i\leftarrow a}^{t} \triangleq \sum_{j\neq i} |H_{aj}|^2 v_{j\to a}^{t},$$

where $m_{j\to a}^{t}$ and $v_{j\to a}^{t}$ are the mean and variance of $\xi_{j\to a}^{t} \sim \mu_{j\to a}^{t}(x_j)$. In this context, $f_{i\leftarrow a}^{t}(x_i)$ could be rewritten as

$$f_{i\leftarrow a}^{t}(x_i) \propto \int p(y_a|z_a)\mathcal{N}(z_a|H_{ai}x_i - Z_{i\leftarrow a}^{t}, V_{i\leftarrow a}^{t})\mathrm{d}z_a. \tag{12}$$

Step 2: We compute $\mathrm{Proj}\left[\mu_{i\to a}^{t}(x_i) f_{i\leftarrow a}^{t}(x_i)\right]$ in (11a). Since all messages $\mu(\cdot)$ are Gaussian, $\mu_{i\to a}^{t}(x_i)$ could be denoted as $\mathcal{N}(x_i|m_{i\to a}^{t}, v_{i\to a}^{t})$. Multiplying this with $f_{i\leftarrow a}^{t}(x_i)$ from (12) and applying the Gaussian reproduction property,[3] we see

$$\begin{aligned}
&\mu_{i\to a}^{t}(x_i)\cdot f_{i\leftarrow a}^{t}(x_i)\\
&\propto \int p(y_a|z_a)\mathcal{N}(z_a|H_{ai}x_i - Z_{i\leftarrow a}^{t}, V_{i\leftarrow a}^{t})\mu_{i\to a}^{t}(x_i)\mathrm{d}z_a\\
&\propto \int p(y_a|z_a)\mathcal{N}(z_a|Z_a^t, V_a^t)\mathcal{N}(x_i|\tilde{m}_{i\leftarrow a}^{t}(z_a), \tilde{v}_{i\leftarrow a}^{t})\mathrm{d}z_a
\end{aligned}$$

where $\tilde{m}_{i\leftarrow a}^{t}(z_a)$ is to highlight the dependence on z_a, and

$$V_a^t = \sum_j |H_{aj}|^2 v_{j\to a}^{t}, \quad Z_a^t = \sum_j H_{aj} m_{j\to a}^{t}, \tag{13}$$

$$\tilde{v}_{i\leftarrow a}^{t} = \left(\frac{1}{v_{i\to a}^{t}} + \frac{|H_{ai}|^2}{V_{i\leftarrow a}^{t}}\right)^{-1}, \tag{14}$$

$$\tilde{m}_{i\leftarrow a}^{t}(z_a) = \tilde{v}_{i\leftarrow a}^{t}\left[\frac{m_{i\to a}^{t}}{v_{i\to a}^{t}} + \frac{H_{ai}^{*}(z_a - Z_{i\leftarrow a}^{t})}{V_{i\leftarrow a}^{t}}\right]. \tag{15}$$

The projection then reads

$$\begin{aligned}
&\mathrm{Proj}\left[\mu_{i\to a}^{t}(x_i) f_{i\leftarrow a}^{t}(x_i)\right]\\
&= \mathrm{Proj}\left[\frac{\int p(y_a|z_a)\mathcal{N}(z_a|Z_a^t, V_a^t)\mathcal{N}(x_i|\tilde{m}_{i\leftarrow a}^{t}(z_a), \tilde{v}_{i\leftarrow a}^{t})\mathrm{d}z_a}{\int p(y_a|z_a)\mathcal{N}(z_a|Z_a^t, V_a^t)\mathrm{d}z_a}\right]\\
&= \mathrm{Proj}\left\{\mathbb{E}_{\zeta_a^t}\left[\mathcal{N}(x_i|\tilde{m}_{i\leftarrow a}^{t}(\zeta_a^t), \tilde{v}_{i\leftarrow a}^{t})\right]\right\}
\end{aligned} \tag{16}$$

where $\zeta_a^t \sim p_{\zeta_a^t}(z_a) = \frac{p(y_a|z_a)\mathcal{N}(z_a|Z_a^t, V_a^t)}{\int p(y_a|z)\mathcal{N}(z|Z_a^t, V_a^t)\mathrm{d}z}$. Denote

$$\tilde{z}_a^t \triangleq \mathbb{E}_{\zeta_a^t}[\zeta_a^t], \qquad \tilde{v}_a^t \triangleq \mathbb{V}_{\zeta_a^t}[\zeta_a^t], \tag{17}$$

[2]We note that an interpretation of GAMP that leverages the KL divergence and variational optimization was published in [21]. This prior work directly builds on the idea of similar Gaussian projection.

[3]Gaussian reproduction property [22]: $\mathcal{N}(x|a,A)\cdot\mathcal{N}(x|b,B) = \mathcal{N}(0|a-b, A+B)\cdot\mathcal{N}(x|c,C)$ with $C=(1/A+1/B)^{-1}$ and $c = C\cdot(a/A+b/B)$.

then the projection in (16) could be computed via (5):

$$\mathrm{Proj}\left\{\mathbb{E}_{\zeta_a^t}\left[\mathcal{N}(x_i|\tilde{m}_{i\leftarrow a}^t(\zeta_a^t),\tilde{v}_{i\leftarrow a}^t)\right]\right\}=\mathcal{N}\left(x_i|m_{i,a}^t,v_{i,a}^t\right) \tag{18}$$

$$m_{i,a}^t = \mathbb{E}_{\zeta_a^t}\left[\tilde{m}_{i\leftarrow a}^t(\zeta_a^t)\right] = \frac{m_{i\to a}^t V_{i\leftarrow a}^t + v_{i\to a}^t H_{ai}^*(\tilde{z}_a^t - Z_{i\leftarrow a}^t)}{V_a^t} \tag{19}$$

$$v_{i,a}^t = \mathbb{E}_{\zeta_a^t}\left[\tilde{v}_{i\leftarrow a}^t + |\tilde{m}_{i\leftarrow a}^t(\zeta_a^t)|^2\right] - |\mathbb{E}_{\zeta_a^t}\left[\tilde{m}_{i\leftarrow a}^t(\zeta_a^t)\right]|^2 = \frac{v_{i\to a}^t V_{i\leftarrow a}^t V_a^t + (v_{i\to a}^t)^2 |H_{ai}|^2 \tilde{v}_a^t}{(V_a^t)^2} \tag{20}$$

Step 3: we now compute $\mu_{i\leftarrow a}^{t+1}(x_i)$ in (11a) using the Gaussian reproduction property and (18). The result is as follows

$$\mu_{i\leftarrow a}^{t+1}(x_i) \propto \mathcal{N}(x_i|m_{i\leftarrow a}^{t+1}, v_{i\leftarrow a}^{t+1}) \tag{21}$$

$v_{i\leftarrow a}^{t+1} = (\frac{1}{v_{i,a}^t} - \frac{1}{v_{i\to a}^t})^{-1} = \frac{(V_a^t)^2 - v_{i\to a}^t |H_{ai}|^2 (V_a^t - \tilde{v}_a^t)}{|H_{ai}|^2 (V_a^t - \tilde{v}_a^t)}$, $m_{i\leftarrow a}^{t+1} = v_{i\leftarrow a}^{t+1}(\frac{m_{i,a}^t}{v_{i,a}^t} - \frac{m_{i\to a}^t}{v_{i\to a}^t}) = \frac{H_{ai}^*(\tilde{z}_a^t - Z_a^t)V_a^t + |H_{ai}|^2 m_{i\to a}^t (V_a^t - \tilde{v}_a^t)}{|H_{ai}|^2 (V_a^t - \tilde{v}_a^t)}$. These mean and variance can be further simplified by ignoring terms of order $O(|H_{ai}|^2)$ as follows (similar to [2, (118)])

$$v_{i\leftarrow a}^{t+1} = \frac{1}{|H_{ai}|^2 \tau_a^t}, \qquad \tau_a^t \triangleq \frac{V_a^t - \tilde{v}_a^t}{(V_a^t)^2}, \tag{22}$$

$$m_{i\leftarrow a}^{t+1} = \frac{H_{ai}^* s_a^t + |H_{ai}|^2 \tau_a^t m_{i\to a}^t}{|H_{ai}|^2 \tau_a^t}, \qquad s_a^t \triangleq \frac{\tilde{z}_a^t - Z_a^t}{V_a^t}. \tag{23}$$

Step 4: we compute the projection of (11b)

$$\mathrm{Proj}[p(x_i)\prod_a \mu_{i\leftarrow a}^{t+1}(x_i)] \propto \mathrm{Proj}[p(x_i)\mathcal{N}(x_i|R_i^{t+1},\Sigma_i^{t+1})] \tag{24}$$

$$\propto \mathcal{N}(x_i|\hat{m}_i^{t+1}, \hat{v}_i^{t+1}) \tag{25}$$

where (24) is obtained by applying the Gaussian reproduction property, (25) follows directly from the projection's definition, and

$$\Sigma_i^{t+1} = \left[\sum_{a=1}^{M} |H_{ai}|^2 \tau_a^t\right]^{-1}, \tag{26}$$

$$R_i^{t+1} = \Sigma_i^{t+1} \sum_{a=1}^{M} \left(H_{ai}^* s_a^t + |H_{ai}|^2 \tau_a^t m_{i\to a}^t\right), \tag{27}$$

$$\xi_i^{t+1} \sim p_{\xi_i^{t+1}}(x_i) = \frac{p(x_i)\mathcal{N}(x_i|R_i^{t+1},\Sigma_i^{t+1})}{\int p(x)\mathcal{N}(x|R_i^{t+1},\Sigma_i^{t+1})\,\mathrm{d}x}, \tag{28}$$

$$\hat{m}_i^{t+1} \triangleq \mathbb{E}_{\xi_i^{t+1}}[\xi_i^{t+1}], \qquad \hat{v}_i^{t+1} \triangleq \mathbb{V}_{\xi_i^{t+1}}[\xi_i^{t+1}]. \tag{29}$$

Step 5: We compute (11b) by applying the Gaussian reproduction property to (25) and (21), and neglecting infinitesimals

$$\mu_{i\to a}^{t+1}(x_i) \propto \mathcal{N}\left(x_i|m_{i\to a}^{t+1}, v_{i\to a}^{t+1}\right) \tag{30}$$

$$v_{i\to a}^{t+1} = \left(\frac{1}{\hat{v}_i^{t+1}} - \frac{1}{v_{i\leftarrow a}^{t+1}}\right)^{-1} = \left(\frac{1}{\hat{v}_i^{t+1}} - |H_{ai}|^2 \tau_a^t\right)^{-1} = \hat{v}_i^{t+1} \tag{31}$$

$$m_{i\to a}^{t+1} = v_{i\to a}^{t+1}\left(\frac{\hat{m}_i^{t+1}}{\hat{v}_i^{t+1}} - \frac{m_{i\leftarrow a}^{t+1}}{v_{i\leftarrow a}^{t+1}}\right) = \hat{m}_i^{t+1} - H_{ai}^* s_a^t \hat{v}_i^{t+1} \tag{32}$$

Step 6: We simplify V_a^t and Z_a^t by (31) and (32)

$$V_a^t = \sum_{i=1}^{N} |H_{ai}|^2 \hat{v}_i^t \tag{33}$$

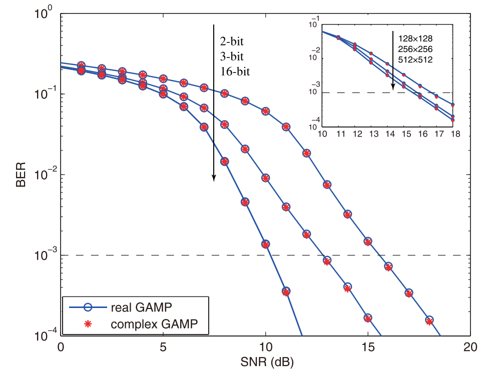

Fig. 3. Bit error rate performance of massive MIMO with ADC quantization.

$$Z_a^t = \sum_{i=1}^{N} H_{ai}\hat{m}_i^t - s_a^{t-1} V_a^t \tag{34}$$

We also simplify R_i^{t+1} by substituting (26) and (32) into (27)

$$R_i^{t+1} = \hat{m}_i^t + \Sigma_i^{t+1} \sum_{a=1}^{M} H_{ai}^* s_a^t. \tag{35}$$

By that, we have extracted $(M+N)$ iterating variables, Σ_i^{t+1}, R_i^{t+1}, Z_a^t and V_a^t, from the $M \times N$ messages, $\mu_{i\leftarrow a}^t(x_i)$ and $\mu_{i\to a}^t(x_i)$, and a new iteration as seen in Algorithm 1 could be established. From this process, one could see that Taylor expansions were completely avoided in the EP-based method. Moreover, the method could switch smoothly between the real case and the complex, with the only difference in defining the Gaussian density as real or complex.

IV. Numerical Examples

To verify complex GAMP in Algorithm 1, we provide below some numerical examples from wireless communications, where complex GAMP is use to offer a posterior marginal estimate of the transmitted symbol, considering the deployment of massive multiple-input multiple-output (MIMO) with low-resolution analog-to-digital converters (ADCs) at the receiver. As a benchmark, we also run the real GAMP, whose problem size has been doubled since each complex signal has to be expanded as a two-dimensional real vector. The general setup follows [23], [24], and we use 10^4 channel realizations for the averaging of each error rate. We vary the ADC resolution from 2-bit, 3-bit, to 16-bit for a 512×512 MIMO. We also fix the ADC resolution at 2-bit to see the impact of antenna numbers: 128×128, 256×256, and 512×512. These results are given in Fig. 3, where the complex GAMP matches very well with the real GAMP.

V. Conclusion

This letter proposed an EP-based method for the derivation of GAMP in complex. Compared with previous work that was LBP-based, our method did not require the operation of Taylor expansions, thus saving the trouble of expanding and analyzing the series. Our method also unifies the derivation for the real and complex GAMP, with a minor difference only in the definition of Gaussian densities. To sum up, this letter offers a new link that connects EP to the GAMP.

References

[1] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proc. Nat. Acad. Sci. USA*, vol. 106, no. 45, pp. 18914–18919, 2009.

[2] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," 2010, arXiv:1010.5141.

[3] J. T. Parker, P. Schniter, and V. Cevher, "Bilinear generalized approximate message passing—Part I: Derivation," *IEEE Trans. Signal Process.*, vol. 62, no. 22, pp. 5839–5853, Nov. 2014. [Online]. Available: https://doi.org/10.1109/TSP.2014.2357776

[4] P. Schniter, "Turbo reconstruction of structured sparse signals," in *Proc. 44th Annu. Conf. Inf. Sci. Syst.*, 2010, pp. 1–6.

[5] P. Schniter, "A message-passing receiver for BICM-OFDM over unknown clustered-sparse channels," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 8, pp. 1462–1474, Dec. 2011.

[6] J. P. Vila and P. Schniter, "Expectation-maximization Gaussian-mixture approximate message passing," *IEEE Trans. Signal Process.*, vol. 61, no. 19, pp. 4658–4672, Oct. 2013.

[7] P. Schniter and S. Rangan, "Compressive phase retrieval via generalized approximate message passing," *IEEE Trans. Signal Process.*, vol. 63, no. 4, pp. 1043–1055, Feb. 2015.

[8] T. P. Minka, "A family of algorithms for approximate Bayesian inference," Ph.D. dissertation, Dept. Elect. Eng. Comput. Sci., Massachusetts Inst. Technol., Cambridge, MA, USA, 2001.

[9] S. Wu, L. Kuang, Z. Ni, J. Lu, D. Huang, and Q. Guo, "Low-complexity iterative detection for large-scale multiuser MIMO-OFDM systems using approximate message passing," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 902–915, Oct. 2014.

[10] X. Meng, S. Wu, L. Kuang, and J. Lu, "An expectation propagation perspective on approximate message passing," *IEEE Signal Process. Lett.*, vol. 22, no. 8, pp. 1194–1197, Aug. 2015.

[11] X. Meng, S. Wu, L. Kuang, and J. Lu, "Concise derivation of complex Bayesian approximate message passing via expectation propagation," 2015, arXiv:1509.08658.

[12] X. Meng, S. Wu, and J. Zhu, "A unified Bayesian inference framework for generalized linear models," *IEEE Signal Process. Lett.*, vol. 25, no. 3, pp. 398–402, Mar. 2018.

[13] S. Rangan, A. K. Fletcher, V. K. Goyal, E. Byrne, and P. Schniter, "Hybrid approximate message passing," *IEEE Trans. Signal Process.*, vol. 65, no. 17, pp. 4577–4592, Sep. 2017.

[14] E. Byrne and P. Schniter, "Sparse multinomial logistic regression via approximate message passing," *IEEE Trans. Signal Process.*, vol. 64, no. 21, pp. 5485–5498, Nov. 2016.

[15] E. T. Jaynes, *Probability Theory: The Logic of Science*. Cambridge, U.K.: Cambridge Univ. Press, 2003.

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

[17] S. Rangan, "Estimation with random linear mixing, belief propagation and compressed sensing," in *Proc. 44th Annu. Conf. Inf. Sci. Syst.*, 2010, pp. 1–6.

[18] M. R. Andersen, "Sparse inference using approximate message passing," M.S. thesis, Dept. Appl. Math. Comput. Sci., Tech. Univ. Denmark, Kongens Lyngby, Denmark, 2014.

[19] D. J. Thouless, P. W. Anderson, and R. G. Palmer, "Solution of 'solvable model of a spin glass,"' *Philos. Mag.*, vol. 35, no. 3, pp. 593–601, 1977.

[20] J. Eriksson, E. Ollila, and V. Koivunen, "Statistics for complex random variables revisited," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2009, pp. 3565–3568.

[21] S. Rangan, P. Schniter, E. Riegler, A. K. Fletcher, and V. Cevher, "Fixed points of generalized approximate message passing with arbitrary matrices," *IEEE Trans. Inf. Theory*, vol. 62, no. 12, pp. 7464–7474, Dec. 2016.

[22] C. E. Rasmussen and C. K. Williams, *Gaussian Process for Machine Learning*. Cambridge, MA, USA: MIT Press, 2006.

[23] H. He, C. K. Wen, and S. Jin, "Generalized expectation consistent signal recovery for nonlinear measurements," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2017, pp. 2333–2337.

[24] H. He, C. K. Wen, and S. Jin, "Bayesian optimal data detector for hybrid mmwave MIMO-OFDM systems with low-resolution ADCs," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 3, pp. 469–483, Jun. 2018.