Convergent Approximate Message Passing by Alternating Constrained Minimization of Bethe Free Energy

Dirk Slock
EURECOM
Communication Systems Department
Sophia Antipolis, France
Email: dirk.slock@eurecom.fr

***Abstract*—Approximate Message Passing (AMP) allows for Bayesian inference in linear models with non identically independently distributed (n.i.i.d.) Gaussian priors and measurements of the linear mixture outputs with n.i.i.d. Gaussian noise. It represents an efficient technique for approximate inference which becomes accurate when both rows and columns of the measurement matrix can be treated as sets of independent vectors and both dimensions become large. It has been shown that the fixed points of AMP correspond to extrema of a large system limit of the Bethe Free Energy (LSL-BFE), which represents a meaningful approximation optimization criterion regardless of whether the measurement matrix exhibits the independence properties. However, the convergence of AMP can be notoriously problematic for certain measurement matrices and the only sure fix so far is damping (by a difficult to determine amount). In this paper we revisit the AMP algorithm by rigorously applying an alternating constrained minimization strategy to an appropriately reparameterized LSL-BFE with matched variable and constraint partitioning. This guarantees convergence, and due to convexity in the Gaussian case, to the global optimum. We show that the AMP estimates converge to the Linear Minimum Mean Squared Error (LMMSE) estimates, regardless of the behavior of the variances. In the LSL, the variances also converge to the LMMSE values, and hence to the correct values.**

I. Introduction

In the Gaussian case, the signal model for the recovery of a sparse signal vector $\boldsymbol{x}$ can be formulated as, $\mathbf{z} = \mathbf{A}\boldsymbol{x}$, $\boldsymbol{y} = \mathbf{z} + \boldsymbol{v}$, where $\boldsymbol{y}$ are the observations or data, $\mathbf{A}$ is called the measurement or the sensing matrix which is known and is of dimension $M \times N$ with typically $M < N$. In the sparse model case, $\boldsymbol{x}$ contains only K non-zero (or significant) entries, with $K < M < N$. In Bayesian inference, the Sparse Bayesian Learning (SBL) algorithm was first proposed by [1], [2]. SBL is based on a two or three layer hierarchical prior on the sparse coefficients $\boldsymbol{x}$. The priors for the hyperparameters (precision parameters) are chosen such that the marginal prior for $\boldsymbol{x}$ induces sparsity, allowing the majority of the coefficients to tend towards zero. It is worth mentioning that [3] provides a detailed overview of the various sparse signal recovery algorithms which fall under l_1 or l_2 norm minimization approaches such as Basis Pursuit, LASSO etc and SBL methods. The authors justify the superior recovery performance of SBL compared to the above mentioned conventional methods. Nevertheless, the matrix inversion involved in the Linear Minimum Mean Squared Error (LMMSE) step in SBL at each iteration makes it computationally complex even for moderately large data sets. This complexity is the motivation behind approximate inference methods.

Belief Propagation (BP) based SBL algorithms [4] are computationally more efficient. A more detailed discussion on the various approximate inference methods for SBL appears in [5]. Various studies on the convergence analysis of Gaussian BP (GaBP) can be found in [6]–[9]. Although BP achieves great empirical success [10], not enough rigorous work exists to characterize the convergence behavior of BP in loopy networks. In [11] a convergence condition for GaBP is provided which requires the underlying distribution to be walk-summable. Their convergence analysis is based on the Gaussian Markov random field (GMRF) based decomposition, in which the underlying distribution is expressed in terms of the pairwise connections between the variables.

The Approximate Message Passing (AMP) algorithm has been introduced to further reduce complexity of GaBP. In Generalized AMP (GAMP), the vector x can have non-Gaussian priors and the measurement process can be more general than with additive Gaussian noise. However, the convergence of AMP can be problematic for certain measurement matrices $\mathbf{A}$. Many variations have been introduced to help AMP converge, such as adding ADMM, exploiting part of the singular value decomposition of the measurement matrix in Vector AMP (VAMP) (but which does not allow to handle n.i.i.d. priors conveniently), sequential updating in Swept AMP (SwAMP) which works almost always, and especially by introducing damping with typically difficult to determine damping requirements.

A. *Contributions of this paper*

- We propose a version of AMP with guaranteed convergence, by rigorously applying an alternating constrained minimization strategy with matched variable and constraint partitioning. We apply this strategy to an appropriately augmented Lagrangian of the constrained Large System Limit of the Bethe Free Energy.
- The new AMP, dubbed AMBAMP below, requires to solve for the mean constraint Lagrange multipliers $\mathbf{s}$ appearing in the posterior mean $\widehat{\mathbf{z}}(\mathbf{s})$ to make it satisfies

this mean constraint. This can be solved explicitely in the Gaussian case considered here.

- We indicate that asymptotically, for a statistical n.i.i.d. element model for $\mathbf{A}$, the variance computations in AMBAMP are exact and we indicate an asymptotically correct iterative scheme to compute them.
- We also show (perhaps for the first time) that the estimates (means) produced by AMP in the Gaussian case converge to the actual Linear Mininum Mean Squared Error estimates, regardless of whether the variances converge to the correct LMMSE values or not.

II. Approximate Message Passing

The data model considered in AMP is essentially a linear mixing model

$$\mathbf{z} = \mathbf{A}\,\boldsymbol{x}\;,\; p_{\boldsymbol{x}}(\boldsymbol{x})\;,\; p_{\boldsymbol{y}|\mathbf{z}}(\boldsymbol{y}|\mathbf{z}) \tag{1}$$

with (possibly) non identically independently distributed (n.i.i.d.) prior $p_{\boldsymbol{x}}(\boldsymbol{x}) = \prod_{i=1}^{N} p_{x_i}(x_i)$ and n.i.i.d. measurements $p_{\boldsymbol{y}|\mathbf{z}}(\boldsymbol{y}|\mathbf{z}) = \prod_{k=1}^{M} p_{y_k|z_k}(y_k|z_k)$. In Bayesian estimation we are interested in the posterior, which is given by

$$\begin{aligned} &p_{\boldsymbol{x},\mathbf{z}|\boldsymbol{y}}(\boldsymbol{x},\mathbf{z}|\boldsymbol{y}) = p_{\boldsymbol{x}}(\boldsymbol{x})\, p_{\boldsymbol{y}|\mathbf{z}}(\boldsymbol{y}|\mathbf{z})\; \mathbb{1}_{\{\mathbf{z}=\mathbf{A}\boldsymbol{x}\}}\; / \, p_{\boldsymbol{y}}(\boldsymbol{y}) \\ &= \tfrac{1}{Z(\boldsymbol{y})} e^{-\sum_{i=1}^{N} f_{x_i}(x_i) - \sum_{k=1}^{M} f_{z_k}(z_k)}\; \mathbb{1}_{\{\mathbf{z}=\mathbf{A}\boldsymbol{x}\}} \end{aligned} \tag{2}$$

where we have the negative loglikelihoods for prior and measurements

$$f_{x_i}(x_i) = -\ln p_{x_i}(x_i)\;,\; f_{z_k}(z_k) = -\ln p_{y_k|z_k}(y_k|z_k) \tag{3}$$

where the equality in case of $f_{z_k}(z_k)$ is up to constants that may depend on $\boldsymbol{y}$ (and which are absorbed in the normalization constant $Z(\boldsymbol{y})$). We shall consider here the Gaussian case, with n.i.i.d. Gaussian $p_{\boldsymbol{x}}(\boldsymbol{x})$ and $p_{\boldsymbol{y}|\mathbf{z}}(\boldsymbol{y}|\mathbf{z}) = p_{\boldsymbol{v}}(\boldsymbol{y}-\mathbf{z})$, so we have (neglecting constants)

$$f_{x_i}(x_i) = \frac{x_i^2}{2\sigma^2_{x_i}}\;,\; f_{z_k}(z_k) = \frac{(z_k - y_k)^2}{2\sigma^2_{v_k}}\;. \tag{4}$$

We shall need below the vectors $\boldsymbol{\sigma}^2_x = [\sigma^2_{x_1}\cdots\sigma^2_{x_N}]^T$, $\boldsymbol{\sigma}^2_v = [\sigma^2_{v_1}\cdots\sigma^2_{v_M}]^T$. The problem in Bayesian estimation is the computation of the constant $Z(\boldsymbol{y})$ in (2) and of the posterior means and variances. Belief propagation is a message passing technique that allows to compute the posterior marginals. However, due to loops in the factor graph, loopy belief propagation may have convergences issues and is furthermore still relatively complex. AMP is an approximate belief propagation technique which is motivated by asymptotic considerations in which the rows and columns of the measurement matrix $\mathbf{A}$ are considered as random and independent, in which case AMP can actually produce the correct posterior marginals. In any case, AMP computes a separable approximate posterior of the form

$$q_{\boldsymbol{x},\mathbf{z}}(\boldsymbol{x},\mathbf{z}) = q_{\boldsymbol{x}}(\boldsymbol{x})\, q_{\mathbf{z}}(\mathbf{z}) = \prod_{i=1}^{N} q_{x_i}(x_i)\;\prod_{k=1}^{M} q_{z_k}(z_k) \tag{5}$$

in which the dependence on $\boldsymbol{y}$ has been omitted. Note that as in Expectation Propagation, these posterior marginal (approximations) will be of the form

$$q_{x_i}(x_i) = m_{x_i}(x_i)\, p_{x_i}(x_i)\;,\; q_{z_k}(z_k) = m_{z_k}(z_k)\, p_{y_k|z_k}(y_k|z_k) \tag{6}$$

where not all dependence on $\boldsymbol{y}$ is shown explicitly. The factors $m_{x_i}(x_i)$, $m_{z_k}(z_k)$ are the extrinsic informations on the respective variables. The AMP algorithm [12], [13] appears in the table for Algorithm 1. We only consider here Sum-Product AMP (for MMSE estimation, as opposed to Max-Sum AMP for MAP estimation).

Algorithm 1 AMP

```
Require: $\boldsymbol{y}$, $\mathbf{A}$, $\mathbf{S} = \mathbf{A}.\mathbf{A}$, $f_{\boldsymbol{x}}(\boldsymbol{x})$, $f_{\mathbf{z}}(\mathbf{z})$
Initialize: $t = 0$, $\widehat{\boldsymbol{x}}^t$, $\boldsymbol{\tau}^t_x$, $\mathbf{s}^{t-1} = \mathbf{0}$
repeat
    [Output node update]
    $\boldsymbol{\tau}^t_p = \mathbf{S}\,\boldsymbol{\tau}^t_x$
    $\boldsymbol{p}^t = \mathbf{A}\,\widehat{\boldsymbol{x}}^t - \mathbf{s}^{t-1}.\boldsymbol{\tau}^t_p$
    $\widehat{\mathbf{z}}^t = \boldsymbol{p}^t.\boldsymbol{\sigma}^2_v./(\boldsymbol{\sigma}^2_v + \boldsymbol{\tau}^t_p) + \boldsymbol{y}.\boldsymbol{\tau}^t_p./(\boldsymbol{\sigma}^2_v + \boldsymbol{\tau}^t_p)$
    $\boldsymbol{\tau}^t_z = \boldsymbol{\sigma}^2_v.\boldsymbol{\tau}^t_p./(\boldsymbol{\sigma}^2_v + \boldsymbol{\tau}^t_p)$
    $\mathbf{s}^t = (\widehat{\mathbf{z}}^t - \boldsymbol{p}^t)./\boldsymbol{\tau}^t_p$
    $\boldsymbol{\tau}^t_s = (\mathbf{1} - \boldsymbol{\tau}^t_z./\boldsymbol{\tau}^t_p)./\boldsymbol{\tau}^t_p$
    [Input node update]
    $\boldsymbol{\tau}^t_r = \mathbf{1}./(\mathbf{S}^T\boldsymbol{\tau}^t_s)$
    $\mathbf{r}^t = \widehat{\boldsymbol{x}}^t + \boldsymbol{\tau}^t_r.\mathbf{A}^T\mathbf{s}^t$
    $\widehat{\boldsymbol{x}}^{t+1} = \mathbf{r}^t.\boldsymbol{\sigma}^2_x./(\boldsymbol{\sigma}^2_x + \boldsymbol{\tau}^t_r)$
    $\boldsymbol{\tau}^t_x = \boldsymbol{\tau}^t_r.\boldsymbol{\sigma}^2_x./(\boldsymbol{\sigma}^2_x + \boldsymbol{\tau}^t_r)$
until Convergence
```

III. AMBAMP

AMB is short for ACM-LSL-BFE: Alternating Constrained Minimization of the Large System Limit of the Bethe Free Energy. As we shall see, AMBAMP uses most of the same updates as AMP, but AMP does not rigorously follow the principle of alternating minimization (block coordinate descent) esp. in the presence of constraints. It has been shown that any fixed point of the AMP algorithm is a critical point of the following constrained minimization of a Large System Limit (LSL) of the Bethe Free Energy (BFE) (see [13] and references therein):

$$\begin{aligned} \min_{q_{\boldsymbol{x}},q_{\mathbf{z}},\boldsymbol{\tau}_p}\; & J_{LSL-BFE}(q_{\boldsymbol{x}}, q_{\mathbf{z}}, \boldsymbol{\tau}_p) \\ s.t.\; & \mathbb{E}(\mathbf{z}|q_{\mathbf{z}}) = \mathbf{A}\;\mathbb{E}(\boldsymbol{x}|q_{\boldsymbol{x}}) \\ & \boldsymbol{\tau}_p = \mathbf{S}\,\mathrm{var}(\boldsymbol{x}|q_{\boldsymbol{x}}) \end{aligned} \tag{7}$$

where the LSL BFE is given by

$$\begin{aligned} &J_{LBFE}(q_{\boldsymbol{x}}, q_{\mathbf{z}}, \boldsymbol{\tau}_p) = D(q_{\boldsymbol{x}}||e^{-f_{\boldsymbol{x}}}) + D(q_{\mathbf{z}}||e^{-f_{\mathbf{z}}}) + H_G(q_{\mathbf{z}}, \boldsymbol{\tau}_p), \\ &\text{with } H_G(q_{\mathbf{z}}, \boldsymbol{\tau}_p) = \tfrac{1}{2}\sum_{k=1}^{M}\left[\frac{\mathrm{var}(z_k|q_{z_k})}{\tau_{p_k}} + \ln(2\pi\,\tau_{p_k})\right] \end{aligned} \tag{8}$$

and where $D(q||p) = \mathbb{E}(\ln(\frac{q}{p})\,|q)$ is the Kullback-Leibler distance (KLD) and $H_G(q_{\mathbf{z}}, \boldsymbol{\tau}_p)$ is a sum of a KLD and an entropy of Gaussians with identical means but different variances. Here $\mathbb{E}(\mathbf{z}|q_{\mathbf{z}})$ denotes the expectation of the vector

$\mathbf{z}$ using the distribution $q_{\mathbf{z}}$, and $\boldsymbol{\tau}_p = [\tau_{p_1} \cdots \tau_{p_N}]^T$ denotes a vector of variances.

The LSL BFE optimization problem (8) can be reformulated with the following augmented Lagrangian

$$\begin{aligned}
&\min_{q_{\boldsymbol{x}}, q_{\mathbf{z}}, \boldsymbol{\tau}_p, \boldsymbol{u}} \max_{\mathbf{s}, \boldsymbol{\tau}_s} L(q_{\boldsymbol{x}}, q_{\mathbf{z}}, \boldsymbol{\tau}_p, \boldsymbol{u}, \mathbf{s}, \boldsymbol{\tau}_s) \text{ with} \\
&L = D(q_{\boldsymbol{x}}||e^{-f_{\boldsymbol{x}}}) + D(q_{\mathbf{z}}||e^{-f_{\mathbf{z}}}) + H_G(q_{\mathbf{z}}, \boldsymbol{\tau}_p) \\
&+\mathbf{s}^T(\mathbb{E}(\mathbf{z}|q_{\mathbf{z}}) - \mathbf{A}\,\mathbb{E}(\boldsymbol{x}|q_{\boldsymbol{x}})) - \tfrac{1}{2}\boldsymbol{\tau}_s^T(\boldsymbol{\tau}_p - \mathbf{S}\,\text{var}(\boldsymbol{x}|q_{\boldsymbol{x}})) \\
&+\tfrac{1}{2}\|\,\mathbb{E}(\boldsymbol{x}|q_{\boldsymbol{x}}) - \boldsymbol{u}\|^2_{\boldsymbol{\tau}_r} + \tfrac{1}{2}\|\,\mathbb{E}(\mathbf{z}|q_{\mathbf{z}}) - \mathbf{A}\,\boldsymbol{u}\|^2_{\boldsymbol{\tau}_p}
\end{aligned} \tag{9}$$

where $\mathbf{s}$, $\boldsymbol{\tau}_s$ are Lagrange multipliers, and $\boldsymbol{\tau}_r = \mathbf{1./}(\mathbf{S}^T\boldsymbol{\tau}_s)$ is just a short-hand notation for quantities that depend on $\boldsymbol{\tau}_s$. We also use the notations: $\|\boldsymbol{u}\|^2_{\boldsymbol{\tau}} = \sum_i u_i^2/\tau_i$, element-wise multiplication as in $\mathbf{s.}\boldsymbol{\tau}$ and element-wise division as in $\mathbf{1./}\boldsymbol{\tau}$, where $\mathbf{1}$ is a vector of 1's.

We interpret the constraints as follows:

$\mathbb{E}(\mathbf{z}|q_{\mathbf{z}}) = \mathbf{A}\,\mathbb{E}(\boldsymbol{x}|q_{\boldsymbol{x}})$ is interpreted as a constraint on $\mathbb{E}(\mathbf{z}|q_{\mathbf{z}})$, and

$\boldsymbol{\tau}_p = \mathbf{S}\,\text{var}(\boldsymbol{x}|q_{\boldsymbol{x}})$ (which is a vector of the individual variances) is interpreted as a constraint on $\boldsymbol{\tau}_p$. To interpret constraints as constraints on a subset of the variables, such subset should be rich enough to allow to satisfy the constraints. Due to the updating order, the other variables will be fixed actually as can be seen further. So the alternating optimization of (9), which corresponds to alternating minimization of the constrained problem (8), should be carried out in the following way. In the partitioning of the variables to be updated, the Lagrange multipliers for the constraints in which a given subset of variables is involved, should be optimized at the same time as that subset of variables. Such alternating optimization policy guarantees the cost function to decrease at each update, and hence to converge, to at least a local optimum. We propose to follow the following updating order

$$\{q_{\mathbf{z}}, \mathbf{s}\} \to \{\boldsymbol{u}\} \to \{\boldsymbol{\tau}_p, \boldsymbol{\tau}_s\} \to \{q_{\boldsymbol{x}}\}\,. \tag{10}$$

In other words, at iteration t we have the following sequence

$$\{q^t_{\mathbf{z}}, \mathbf{s}^t\} = \arg\min_{q_{\mathbf{z}}} \max_{\mathbf{s}} L(q^{t-1}_{\boldsymbol{x}}, q_{\mathbf{z}}, \boldsymbol{\tau}^{t-1}_p, \boldsymbol{u}^{t-1}, \mathbf{s}, \boldsymbol{\tau}^{t-1}_s) \tag{11}$$

$$\{\boldsymbol{u}^t\} = \arg\min_{\boldsymbol{u}} L(q^{t-1}_{\boldsymbol{x}}, q^t_{\mathbf{z}}, \boldsymbol{\tau}^{t-1}_p, \boldsymbol{u}, \mathbf{s}^t, \boldsymbol{\tau}^{t-1}_s) \tag{12}$$

$$\{\boldsymbol{\tau}^t_p, \boldsymbol{\tau}^t_s\} = \arg\min_{\boldsymbol{\tau}_p} \max_{\boldsymbol{\tau}_s} L(q^{t-1}_{\boldsymbol{x}}, q^t_{\mathbf{z}}, \boldsymbol{\tau}_p, \boldsymbol{u}^t, \mathbf{s}^t, \boldsymbol{\tau}_s) \tag{13}$$

$$\{q^t_{\boldsymbol{x}}\} = \arg\min_{q_{\boldsymbol{x}}} L(q_{\boldsymbol{x}}, q^t_{\mathbf{z}}, \boldsymbol{\tau}^t_p, \boldsymbol{u}^t, \mathbf{s}^t, \boldsymbol{\tau}^t_s) \tag{14}$$

A. Update of $\{q_{\mathbf{z}}, \mathbf{s}\}$

The most tricky part is the update of $\{q_{\mathbf{z}}, \mathbf{s}\}$. To that end, consider

$$\begin{aligned}
&L(q^{t-1}_{\boldsymbol{x}}, q_{\mathbf{z}}, \boldsymbol{\tau}^{t-1}_p, \boldsymbol{u}^{t-1}, \mathbf{s}, \boldsymbol{\tau}^{t-1}_s) \\
&= D(q_{\mathbf{z}}||e^{-f_{\mathbf{z}}}) + \tfrac{1}{2}\text{var}(\mathbf{z}|q_{\mathbf{z}})\mathbf{./}\boldsymbol{\tau}^{t-1}_p \\
&+\mathbf{s}^T\,\mathbb{E}(\mathbf{z}|q_{\mathbf{z}}) + \tfrac{1}{2}\|\,\mathbb{E}(\mathbf{z}|q_{\mathbf{z}}) - \mathbf{A}\,\boldsymbol{u}^{t-1}\|^2_{\boldsymbol{\tau}^{t-1}_p} + const. \\
&= D(q_{\mathbf{z}}||e^{-f_{\mathbf{z}}}) + \tfrac{1}{2}\,\mathbb{E}(\mathbf{z}^T\mathbf{z}|q_{\mathbf{z}})\mathbf{./}\boldsymbol{\tau}^{t-1}_p \\
&-(\mathbb{E}(\mathbf{z}|q_{\mathbf{z}}))^T((\mathbf{A}\,\boldsymbol{u}^{t-1})\mathbf{./}\boldsymbol{\tau}^{t-1}_p - \mathbf{s}) + const. \\
&= D(q_{\mathbf{z}}||e^{-f_{\mathbf{z}}}) + \tfrac{1}{2}\,\mathbb{E}(\|\mathbf{z} - \boldsymbol{p}^t(\mathbf{s})\|^2_{\boldsymbol{\tau}^{t-1}_p}|q_{\mathbf{z}}) + const.
\end{aligned} \tag{15}$$

where *const.* denotes constants w.r.t. $\mathbf{z}$ and where we introduced

$$\boldsymbol{p}^t(\mathbf{s}) = \mathbf{A}\,\boldsymbol{u}^{t-1} - \mathbf{s.}\boldsymbol{\tau}^{t-1}_p\,. \tag{16}$$

This cost function is separable, hence we can continue per component. Note that the last expression in (15) can be interpreted as a single KLD:

$$\begin{aligned}
&\min_{q_{z_k}} D(q_{z_k}||g^t_{z_k}/Z^t_{z_k}) \;\Rightarrow\; q^t_{z_k} = g^t_{z_k}/Z^t_{z_k} \\
&Z^t_{z_k}(s_k) = \textstyle\int g^t_{z_k}(z_k; s_k)\,dz_k\,,\; -\ln g^t_{z_k}(z_k; s_k) = \\
&f_{z_k}(z_k) + \tfrac{1}{\tau^{t-1}_{p_k}}(\tfrac{z_k^2}{2} - z_k\,\mathbf{A}_{k,:}\,\boldsymbol{u}^{t-1}) + z_k\,s_k
\end{aligned} \tag{17}$$

where $\mathbf{A}_{k,:}$ denotes row k of $\mathbf{A}$. Note that the partition function $Z^t_{z_k}$ acts as cumulant generating function:

$$\begin{aligned}
&-\frac{\partial \ln Z^t_{z_k}}{\partial s_k} = \mathbb{E}(z_k|q^t_{z_k}) = \mathbb{E}(z_k|p^t_k(s_k), \tau^{t-1}_{p_k}) = \widehat{z}^t_k(s_k) \\
&\frac{\partial^2 \ln Z^t_{z_k}}{\partial s_k^2} = \text{var}(z_k|p^t_k(s_k), \tau^{t-1}_{p_k}) = \tau^t_{z_k}(s_k) \\
&-\frac{\partial^3 \ln Z^t_{z_k}}{\partial s_k^3} = \mathbb{E}(z_k - \mathbb{E}\,z_k)^3
\end{aligned} \tag{18}$$

To satisfy the mean constraint in (7), we require s^t_k to satisfy

$$\widehat{z}^t_k = \widehat{z}^t_k(s^t_k) = \mathbf{A}_{k,:}\widehat{\boldsymbol{x}}^{t-1}\,,\; \tau^t_{z_k} = \tau^t_{z_k}(s^t_k)\,. \tag{19}$$

The second derivative in (18) being positive shows that $\widehat{z}^t_k(s_k)$ is a monotonically increasing function. Actually, we shall see that in the Gaussian case here, $\widehat{z}^t_k(s_k)$ is linear and $\tau^t_{z_k}$ is constant as a function of s_k. Hence the following first-order Taylor series expansion is exact:

$$\widehat{z}^t_k(s^t_k) = \widehat{z}^t_k(s^{t-1}_k) - \tau^t_{z_k}(s^{t-1}_k)\,(s^t_k - s^{t-1}_k) = \mathbf{A}_{k,:}\widehat{\boldsymbol{x}}^{t-1} \tag{20}$$

from which we get

$$s^t_k = s^{t-1}_k - \frac{1}{\tau^t_{z_k}}(\mathbf{A}_{k,:}\widehat{\boldsymbol{x}}^{t-1} - \widehat{z}^t_k(s^{t-1}_k)) \tag{21}$$

which is very similar to a Method of Moments (MM) update of the Lagrange multiplier (which allows the MM to converge). Actually, we can recognize from the last expression in (15) that the extrinsic for $\mathbf{z}$ is Gaussian:

$$\ln m^t_{\mathbf{z}}(\mathbf{z}) = -\frac{1}{2}\,\|\mathbf{z} - \boldsymbol{p}^t(\mathbf{s})\|^2_{\boldsymbol{\tau}^{t-1}_p} \tag{22}$$

which together with the Gaussian prior leads to a Gaussian posterior via the Gauss-Markov theorem

$$\begin{aligned}
&q^t_{\mathbf{z}}(\mathbf{z}) = \mathcal{N}(\mathbf{z}; \widehat{\mathbf{z}}^t, diag(\boldsymbol{\tau}^t_z)) \quad \text{with} \\
&\widehat{\mathbf{z}}^t(\mathbf{s}) = \boldsymbol{p}^t(\mathbf{s})\mathbf{.}\boldsymbol{\sigma}^2_v\mathbf{./}(\boldsymbol{\sigma}^2_v + \boldsymbol{\tau}^{t-1}_p) + \boldsymbol{y}\mathbf{.}\boldsymbol{\tau}^{t-1}_p\mathbf{./}(\boldsymbol{\sigma}^2_v + \boldsymbol{\tau}^{t-1}_p) \\
&\boldsymbol{\tau}^t_z = \boldsymbol{\sigma}^2_v\mathbf{.}\boldsymbol{\tau}^{t-1}_p\mathbf{./}(\boldsymbol{\sigma}^2_v + \boldsymbol{\tau}^{t-1}_p)
\end{aligned} \tag{23}$$

where $\boldsymbol{p}^t(\mathbf{s})$ is defined in (16). This shows that the posterior mean is linear in $\mathbf{s}$ and that the posterior variance is constant w.r.t. $\mathbf{s}$. The posterior variance expression leads to line 8 in Algorithm 2. Line 4 follows from (16) with $\boldsymbol{p}^t = \boldsymbol{p}^t(\mathbf{s}^{t-1})$. Line 5 follows from (23). Line 6 corresponds to the mean constraint in (19). Line 7 finally follows from (21).

B. Update of $\boldsymbol{u}$

From (9), (12), we get

$$\begin{aligned}&L(q_{\boldsymbol{x}}^{t-1}, q_{\mathbf{z}}^t, \boldsymbol{\tau}_p^{t-1}, \boldsymbol{u}, \mathbf{s}^t, \boldsymbol{\tau}_s^{t-1}) \\ &= \tfrac{1}{2}\|\widehat{\boldsymbol{x}}^{t-1} - \boldsymbol{u}\|^2_{\boldsymbol{\tau}_r} + \tfrac{1}{2}\|\mathbf{A}\,\widehat{\boldsymbol{x}}^{t-1} - \mathbf{A}\,\boldsymbol{u}\|^2_{\boldsymbol{\tau}_p} + const.\end{aligned} \tag{24}$$

where $const.$ denotes constants w.r.t. $\boldsymbol{u}$. The minimizer is obviously

$$\boldsymbol{u}^t = \widehat{\boldsymbol{x}}^{t-1}\,. \tag{25}$$

Note that due to the update of (only) $\{q_{\mathbf{z}}, \mathbf{s}\}$ just before, we have $\widehat{\mathbf{z}}^t = \mathbf{A}\,\widehat{\boldsymbol{x}}^{t-1}$ which greatly simplifies this update of $\boldsymbol{u}$. In contrast to [14] where a complex update of $\boldsymbol{u}$ is required which is not compatible with the fast AMP style algorithms.

C. Update of $\{\boldsymbol{\tau}_p, \boldsymbol{\tau}_s\}$

Due to the preceding update of $\boldsymbol{u}$, the two quadratic terms shown in (24) are now zero. As a result, the dependence of those terms on $\boldsymbol{\tau}_p$, $\boldsymbol{\tau}_s$ via the weights disappears for the update of $\{\boldsymbol{\tau}_p, \boldsymbol{\tau}_s\}$ which comes next. Hence the terms of interest in (12) for (10) are now

$$\begin{aligned}&L(q_{\boldsymbol{x}}^{t-1}, q_{\mathbf{z}}^t, \boldsymbol{\tau}_p, \boldsymbol{u}^t, \mathbf{s}^t, \boldsymbol{\tau}_s) \\ &= H_G(q_{\mathbf{z}}^t, \boldsymbol{\tau}_p) - \tfrac{1}{2}\boldsymbol{\tau}_s^T(\boldsymbol{\tau}_p - \mathbf{S}\,\boldsymbol{\tau}_x^{t-1}) + const. = const. + \\ &\tfrac{1}{2}\sum_{k=1}^{M}\left[\frac{\tau_{z_k}^t}{\tau_{p_k}} + \ln(2\pi\,\tau_{p_k})\right] - \frac{1}{2}\sum_{k=1}^{M}\tau_{s_k}(\tau_{p_k} - \mathbf{S}_{k,:}\,\boldsymbol{\tau}_x^{t-1})\end{aligned} \tag{26}$$

where $const.$ denotes constants w.r.t. $\{\boldsymbol{\tau}_p, \boldsymbol{\tau}_s\}$. The optimization yields

$$\frac{\partial L}{\partial \tau_{s_k}} = 0 \;\Rightarrow\; \tau_{p_k}^t = \mathbf{S}_{k,:}\,\boldsymbol{\tau}_x^{t-1} \tag{27}$$

$$\frac{\partial L}{\partial \tau_{p_k}} = \frac{1}{2}\left(-\frac{\tau_{z_k}^t}{\tau_{p_k}^2} + \frac{1}{\tau_{p_k}} - \tau_{s_k}\right) = 0$$

$$\Rightarrow\; \tau_{s_k}^t = \frac{1}{\tau_{p_k}^t}\left(1 - \frac{\tau_{z_k}^t}{\tau_{p_k}^t}\right) \tag{28}$$

$$\tag{29}$$

D. Update of $q_{\boldsymbol{x}}$

For the update of $q_{\boldsymbol{x}}$ in (14) finally, consider the relevant terms in the augmented Lagrangian (and remember that $\boldsymbol{\tau}_r^t = \mathbf{1}./(\mathbf{S}^T\boldsymbol{\tau}_s^t)$ or $\mathbf{1}./\boldsymbol{\tau}_r^t = \mathbf{S}^T\boldsymbol{\tau}_s^t$) (13)

$$\begin{aligned}&L(q_{\boldsymbol{x}}, q_{\mathbf{z}}^t, \boldsymbol{\tau}_p^t, \boldsymbol{u}^t, \mathbf{s}^t, \boldsymbol{\tau}_s^t) \\ &= D(q_{\boldsymbol{x}}\|e^{-f_{\boldsymbol{x}}}) - \mathbf{s}^{t\,T}\mathbf{A}\;\mathbb{E}(\boldsymbol{x}|q_{\boldsymbol{x}}) + \tfrac{1}{2}\boldsymbol{\tau}_s^{t\,T}\mathbf{S}\,\mathrm{var}(\boldsymbol{x}|q_{\boldsymbol{x}}) \\ &+\tfrac{1}{2}\|\,\mathbb{E}(\boldsymbol{x}|q_{\boldsymbol{x}}) - \boldsymbol{u}^t\|^2_{\boldsymbol{\tau}_r^t} + const. \\ &= D(q_{\boldsymbol{x}}\|e^{-f_{\boldsymbol{x}}}) + \tfrac{1}{2}\,(\mathbf{1}./\boldsymbol{\tau}_r^t)^T\;\mathbb{E}(\boldsymbol{x}.\boldsymbol{x}|q_{\boldsymbol{x}}) - \mathbf{s}^{t\,T}\mathbf{A}\;\mathbb{E}(\boldsymbol{x}|q_{\boldsymbol{x}}) \\ &-(\boldsymbol{u}^t./\boldsymbol{\tau}_r^t))^T\,\mathbb{E}(\boldsymbol{x}|q_{\boldsymbol{x}}) + const. \\ &= D(q_{\boldsymbol{x}}\|e^{-f_{\boldsymbol{x}}}) + \tfrac{1}{2}\,(\mathbf{1}./\boldsymbol{\tau}_r^t)^T\;\mathbb{E}(\boldsymbol{x}.\boldsymbol{x}|q_{\boldsymbol{x}}) \\ &-(\boldsymbol{u}^t + \boldsymbol{\tau}_r^t.\mathbf{A}^T\mathbf{s}^t)^T(\mathbb{E}(\boldsymbol{x}|q_{\boldsymbol{x}})./\boldsymbol{\tau}_r^t) + const. \\ &= D(q_{\boldsymbol{x}}\|e^{-f_{\boldsymbol{x}}}) + \tfrac{1}{2}\;\mathbb{E}(\|\boldsymbol{x} - \mathbf{r}^t\|^2_{\boldsymbol{\tau}_r^t}|q_{\boldsymbol{x}}) + const.\end{aligned} \tag{30}$$

where $const.$ denotes constants w.r.t. $\boldsymbol{x}$ and

$$\mathbf{r}^t = \boldsymbol{u}^t + \boldsymbol{\tau}_r^t.\mathbf{A}^T\mathbf{s}^t\,. \tag{31}$$

Algorithm 2 AMBAMP

```
Require: $\boldsymbol{y}$, $\mathbf{A}$, $\mathbf{S} = \mathbf{A}.\mathbf{A}$, $f_{\boldsymbol{x}}(\boldsymbol{x})$, $f_{\mathbf{z}}(\mathbf{z})$
Initialize: $t = 0$, $\widehat{\boldsymbol{x}}^{t-1}$, $\boldsymbol{\tau}_x^{t-1}$, $\boldsymbol{u}^{t-1}$, $\boldsymbol{\tau}_p^{t-1}$, $\mathbf{s}^{t-1} = \mathbf{0}$
repeat
    [Output node update]
    $\boldsymbol{p}^t = \mathbf{A}\,\boldsymbol{u}^{t-1} - \mathbf{s}^{t-1}.\boldsymbol{\tau}_p^{t-1}$
    $\widehat{\mathbf{z}}^t(\mathbf{s}^{t-1}) = (\boldsymbol{p}^t.\boldsymbol{\sigma}_v^2 + \boldsymbol{y}.\boldsymbol{\tau}_p^{t-1})./(\boldsymbol{\sigma}_v^2 + \boldsymbol{\tau}_p^{t-1})$
    $\widehat{\mathbf{z}}^t = \widehat{\mathbf{z}}^t(\mathbf{s}^t) = \mathbf{A}\,\widehat{\boldsymbol{x}}^{t-1}$
    $\boldsymbol{\tau}_z^t = \boldsymbol{\sigma}_v^2.\boldsymbol{\tau}_p^{t-1}./(\boldsymbol{\sigma}_v^2 + \boldsymbol{\tau}_p^{t-1})$
    $\mathbf{s}^t = \mathbf{s}^{t-1} - (\widehat{\mathbf{z}}^t - \widehat{\mathbf{z}}^t(\mathbf{s}^{t-1}))./\boldsymbol{\tau}_z^t$
    $\boldsymbol{u}^t = \widehat{\boldsymbol{x}}^{t-1}$
    [Variance matching]
    $\boldsymbol{\tau}_p^t = \mathbf{S}\,\boldsymbol{\tau}_x^{t-1}$
    $\boldsymbol{\tau}_s^t = (\mathbf{1} - \boldsymbol{\tau}_z^t./\boldsymbol{\tau}_p^t)./\boldsymbol{\tau}_p^t$
    $\boldsymbol{\tau}_r^t = \mathbf{1}./(\mathbf{S}^T\boldsymbol{\tau}_s^t)$
    [Input node update]
    $\mathbf{r}^t = \boldsymbol{u}^t + \boldsymbol{\tau}_r^t.\mathbf{A}^T\mathbf{s}^t$
    $\widehat{\boldsymbol{x}}^t = \mathbf{r}^t.\boldsymbol{\sigma}_x^2./(\boldsymbol{\sigma}_x^2 + \boldsymbol{\tau}_r^t)$
    $\boldsymbol{\tau}_x^t = \boldsymbol{\tau}_r^t.\boldsymbol{\sigma}_x^2./(\boldsymbol{\sigma}_x^2 + \boldsymbol{\tau}_r^t)$
until Convergence
```

This cost function is separable. We get per component

$$\begin{aligned}&\min_{q_{x_k}} D(q_{x_k}\|g_{x_k}^t/Z_{x_k}^t) \;\Rightarrow\; q_{x_k}^t = g_{x_k}^t/Z_{x_k}^t \\ &Z_{x_k}^t = \textstyle\int g_{x_k}^t(x_k)\,dx_k\,, \\ &-\ln g_{x_k}^t(x_k) = f_{x_k}(x_k) + \tfrac{1}{2\tau_{r_k}^t}[(x_k - r_k)^2 - r_k^2]\,.\end{aligned} \tag{32}$$

Actually, we can recognize from the last expression in (30) that the extrinsic for $\boldsymbol{x}$ is again Gaussian:

$$\ln m_{\boldsymbol{x}}^t(\boldsymbol{x}) = -\frac{1}{2}\,\|\boldsymbol{x} - \mathbf{r}^t\|^2_{\boldsymbol{\tau}_r^t} \tag{33}$$

which together with the Gaussian prior for $\boldsymbol{x}$ leads to a Gaussian posterior via the Gauss-Markov theorem

$$\begin{aligned}&q_{\boldsymbol{x}}^t(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x}; \widehat{\boldsymbol{x}}^t, diag(\boldsymbol{\tau}_x^t)) \quad \text{with} \\ &\widehat{\boldsymbol{x}}^t = \mathbf{r}^t.\boldsymbol{\sigma}_x^2./(\boldsymbol{\sigma}_x^2 + \boldsymbol{\tau}_r^t) \\ &\boldsymbol{\tau}_x^t = \boldsymbol{\tau}_r^t.\boldsymbol{\sigma}_x^2./(\boldsymbol{\sigma}_x^2 + \boldsymbol{\tau}_r^t)\end{aligned} \tag{34}$$

where $\mathbf{r}^t$ is defined in (31).

IV. AMBAMP Large System Analysis

In the Gaussian case, we have investigated Large System Analysis in [15] using large random matrix theory. In this analysis, the entries of $\mathbf{A}$ are considered n.i.i.d. with zero mean and variances according to $\mathbf{S} = \mathbf{A}.\mathbf{A}$. An asymptotic regime is considered in which the two dimensions M and N of $\mathbf{A}$ tend to infinity at fixed ratio. According to [13, Appendix A], under some conditions the approximate posterior variances $\boldsymbol{\tau}_x$, $\boldsymbol{\tau}_z$ produced by AMP converge to the solution of the following coupled equations

$$\mathbf{1}./\boldsymbol{\tau}_x = \boldsymbol{\sigma}_x^2 + \mathbf{S}^T\,\boldsymbol{\tau}_z\,,\; \mathbf{1}./\boldsymbol{\tau}_z = \boldsymbol{\sigma}_v^2 + \mathbf{S}\,\boldsymbol{\tau}_x\,. \tag{35}$$

In this asymptotic regime, these variances converge to their correct LMMSE values, namely

$$\boldsymbol{\tau}_x = diag(\,(D(\mathbf{1}./\boldsymbol{\sigma}_x^2) + \mathbf{A}^T\,D(\mathbf{1}./\boldsymbol{\sigma}_v^2)\,\mathbf{A})^{-1})\,. \tag{36}$$

V. AMP Mean Convergence to LMMSE

By eliminating the variables $\boldsymbol{p}$, $\mathbf{s}$ and $\widehat{\mathbf{z}}$ in AMBAMP, one can identify that the mean estimate $\widehat{\boldsymbol{x}}$ in ABMAMP satisfies the following second-order recursion

$$\begin{aligned}\widehat{\boldsymbol{x}}^t &= D(\boldsymbol{\sigma}_x^2./(\boldsymbol{\sigma}_x^2+\boldsymbol{\tau}_r^t))\,\widehat{\boldsymbol{x}}^{t-1} - D(\tau_x^t)\,\mathbf{A}^T D(\mathbf{1}./\boldsymbol{\tau}_z^t)\,\mathbf{A}\widehat{\boldsymbol{x}}^{t-1}\\ &+D(\tau_x^t)\,\mathbf{A}^T D(\mathbf{1}./\boldsymbol{\tau}_p^{t-1})\,\mathbf{A}\widehat{\boldsymbol{x}}^{t-2} + D(\tau_x^t)\,\mathbf{A}^T D(\mathbf{1}./\boldsymbol{\sigma}_v^2)\,\boldsymbol{y}\end{aligned} \tag{37}$$

where $D(\boldsymbol{\tau}) = diag(\boldsymbol{\tau})$ denotes a diagonal matrix constructed from a vector. One can note that the dynamics in the AMBAMP algorithm are different from those of the AMP algorithm. At convergence we can solve for the steady-state value $\widehat{\boldsymbol{x}}$ from (37). We get

$$\begin{aligned}&D(\mathbf{1}./\boldsymbol{\tau}_x)\,(\mathbf{I}-D(\boldsymbol{\sigma}_x^2./(\boldsymbol{\sigma}_x^2+\boldsymbol{\tau}_r)))\,\widehat{\boldsymbol{x}}\\ &+\mathbf{A}^T(D(\mathbf{1}./\boldsymbol{\tau}_z)-D(\mathbf{1}./\boldsymbol{\tau}_p))\,\mathbf{A}\widehat{\boldsymbol{x}} = \mathbf{A}^T D(\mathbf{1}./\boldsymbol{\sigma}_v^2)\,\boldsymbol{y}\\ &\Rightarrow\ \widehat{\boldsymbol{x}} = (D(\mathbf{1}./\boldsymbol{\sigma}_x^2)+\mathbf{A}^T\,D(\mathbf{1}./\boldsymbol{\sigma}_v^2)\,\mathbf{A})^{-1}\mathbf{A}^T D(\mathbf{1}./\boldsymbol{\sigma}_v^2)\,\boldsymbol{y}\end{aligned} \tag{38}$$

which is the LMMSE estimate. This LMMSE steady-state value for the mean is also valid for the original AMP algorithm, if it converges.

VI. Concluding Remarks

To arrive at the convergent AMBAMP algoritrhm, a particular formulation of the Bethe Free Energy (BFE) criterion has been considered with a judiciously chosen Method of Moments extension to incorporate equality constraints quadratically. The Lagrangian of the resulting augmented BFE is then introduced. The alternating optimization of the resulting cost function only leads to the desired algorithm of low complexity when alternating optimization is done in one particular order. Other updating orders would also lead to convergent algorithms but not to low complexity updates.

We have drawn attention to an approach to perform alternating constrained optimization. The approach consists of not only partitioning the variables appearing in the cost function, but also to partition the constraints according to this variable partitioning and to identify for each constraint subset the variable subset that can be used to satisfy these constraints. In the alternating optimization of the cost function w.r.t. each variable subset, the possible corresponding constraint subset should be involved in the constrained optimization sub-problem. This same alternating constraint optimization approach is widely applicable. It has for instance been used in beamforming design where normalized beamformers can be updated separately from the signal powers, which get updated while satisfying the power constraints, leading to waterfilling type solutions [16].

Acknowledgements

EURECOM's research is partially supported by its industrial members: ORANGE, BMW, SAP, iABG, Norton LifeLock, and by the French ANR project CellFree6G.

References

[1] M. E. Tipping, "Sparse Bayesian Learning and the Relevance Vector Machine," *J. Mach. Learn. Res.*, vol. 1, 2001.

[2] D. P. Wipf and B. D. Rao, "Sparse Bayesian Learning for Basis Selection ," *IEEE Trans. on Sig. Proc.*, vol. 52, no. 8, Aug. 2004.

[3] R. Giri and Bhaskar D. Rao, "Type I and type II bayesian methods for sparse signal recovery using scale mixtures," *IEEE Trans. on Sig Process.*, vol. 64, no. 13, 2018.

[4] X. Tan and J. Li, "Computationally Efficient Sparse Bayesian Learning via Belief Propagation ," *IEEE Trans. on Sig. Proc.*, vol. 58, no. 4, Apr. 2013.

[5] C. K. Thomas and D. Slock, "Low Complexity Static and Dynamic Sparse Bayesian Learning Combining BP, VB and EP Message Passing," in *Asilomar Conf. on Sig., Sys., and Comp.*, CA, USA, 2019.

[6] J. Du et al., "Convergence Analysis of Distributed Inference with Vector-Valued Gaussian Belief Propagation," *Jrnl. of Mach. Learn. Res.*, April 2018.

[7] J. Du et al., "Convergence Analysis of the Information Matrix in Gaussian Belief Propagation ," in *IEEE Intl. Conf. on Acoustics, Speech, and Sig. Process.*, New Orleans, LA, USA, 2017.

[8] Q. Su and Y. Wu, "Convergence Analysis of the Variance in Gaussian Belief Propagation," *IEEE Trans. on Sig. Process.*, vol. 62, no. 19, Oct. 2014.

[9] B. Cseke and T. Heskes, "Properties of Bethe Free Energies and Message Passing in Gaussian Models ," *Jrnl. of Art. Intell. Res.*, May 2011.

[10] K. P. Murphy et al., "Loopy belief propagation for approximate inference: an empirical study. ," in *In 15th Conf. Uncert. in Art. Intell. (UAI)*, Stockholm, Sweden, 1999.

[11] D. M. Malioutov, J. K. Johnson, and A. S. Willsky, "Walk-Sums and Belief Propagation in Gaussian Graphical Models ," *Jrnl. of Mach. Learn. Res.*, Oct. 2006.

[12] S. Rangan, "Generalized Approximate Message Passing for Estimation with Random Linear Mixing," in *Proc. IEEE Int. Symp. Inf. Theory*, Saint Petersburg, Russia, 2011, extended version: arxiv1010.5141.

[13] S. Rangan, P. Schniter, A. Fletcher, and V. Cevher, "Fixed Points of Generalized Approximate Message Passing with Arbitrary Matrices," *IEEE Trans. Info. Theory*, Dec. 2016.

[14] S. Rangan, A. Fletcher, P. Schniter, and U.S. Kamilov, "Inference for Generalized Linear Models via Alternating Directions and Bethe Free Energy Minimization," *IEEE Trans. Info. Theory*, Jan. 2017.

[15] C.K. Thomas and D. Slock, "Posterior Variance Predictions in Sparse Bayesian Learning under Approximate Inference Techniques," in *Asilomar Conf. on Sig., Sys., and Comp.*, 2020.

[16] C.K. Thomas and D. Slock, "Hybrid Beamforming Design in Multi-Cell MU-MIMO Systems with per-RF or per-Antenna Power Constraints," in *IEEE Vehic. Tech. Conf. (VTCfall)*, Chicago, IL, USA, Aug. 2018.

Dirk T.M. Slock received an EE degree from Ghent University, Belgium in 1982. In 1984 he was awarded a Fulbright scholarship for Stanford University, USA, where he received the MSEE, MS in Statistics, and PhD in EE in 1986, 1989 and 1989 resp. While at Stanford, he developed new fast recursive least-squares algorithms for adaptive filtering. In 1989-91, he was a member of the research staff at the Philips Research Laboratory Belgium. In 1991, he joined EURECOM where he is now professor. At EURECOM, he teaches statistical signal processing (SSP) and signal processing techniques for wireless communications. He invented semi-blind channel estimation, the chip equalizer-correlator receiver used by 3G HSDPA mobile terminals, spatial multiplexing cyclic delay diversity (MIMO-CDD) now part of LTE, and his work led to the Single Antenna Interference Cancellation (SAIC) integrated in the GSM standard in 2006. Recent keywords are multi-cell multi-user (Massive) MIMO, imperfect CSIT, distributed resource allocation, variational and empirical Bayesian learning techniques, large random matrix analysis, audio source separation, location estimation and exploitation. He graduated about 40

PhD students, leading to an edited book and 500+ papers. In 1992 he received one best journal paper award from IEEE-SP and one from EURASIP. He is the coauthor of two IEEE Globecom'98, one IEEE SIU'04, one IEEE SPAWC'05, one IEEE WPNC16 and one IEEE SPAWC18 best student paper award, and a honorary mention (finalist in best student paper contest) at IEEE SSP'05, IWAENC'06, IEEE Asilomar'06 and IEEE ICASSP17. He was an associate editor for the IEEE-SP Transactions in 1994-96 and the IEEE Signal Processing Letters in 2009-10. He was the General Chair of the IEEE-SP SPAWC'06 and IWAENC14 workshops, and EUSIPCO15. He cofounded the start-ups SigTone in 2000 (music signal processing products) and Nestwave in 2014 (Ultra Low-Power Indoor and Outdoor Mobile Positioning). He is a Fellow of IEEE and EURASIP. In 2018 he received the URSI France medal.