When Bayes meets Kullback-Leibler: a Tale of Message Passing and Alternating Optimization

Dirk Slock[1]

[1] Communication Systems Department, EURECOM, France

TU Berlin, 01/12/2023

Outline

1. Bethe Free Energy (BFE) Minimization and Expectation Propagation (EP) - min KLD

2. reVAMP: revisited VAMP
   - EP-like Derivation
   - Relation to CWCU MMSE Estimation

3 Kullback-Leibler Divergence (KLD) Optimization Criteria

- true posterior $p(\boldsymbol{x}|\boldsymbol{y})$, approximating posterior $q(\boldsymbol{x})$
  KLD $D(q||p) = \int q(\boldsymbol{x}) \ln \frac{q(\boldsymbol{x})}{p(\boldsymbol{x})}$.

1. Variational Bayes (VB)

$$\min_q D(q||p) \text{ with } q(\boldsymbol{x}) = \prod_i q_i(x_i),\ p = p(\boldsymbol{x},\boldsymbol{y}) = \prod_a p_a(\boldsymbol{x}_a)$$

   Correlations according to posterior model: zero. Affects variances.
   Mean field: case of scalar $\{x_i\}$.

2. better: Bethe Free Energy (BFE), Belief Propagation (BP), Expectation Propagation (EP)

$$\min_q D(q||p) \text{ with } q(\boldsymbol{x}) = \frac{\prod_a q_a(\boldsymbol{x}_a)}{\prod_i (q_i(x_i))^{N_i-1}},\ p = p(\boldsymbol{x},\boldsymbol{y})$$

3. more desirable:

$$\min_q D(p||q) \text{ with } q(\boldsymbol{x}) = \prod_i q_i(x_i)$$

   Correlations/variances captured by true posterior. Optimized approximately (asymptotically) by EP!

Posterior variance prediction suboptimality and KLD formulations

- Consider $p \sim \mathcal{N}(\boldsymbol{\mu}_p, \boldsymbol{\Sigma}_p)$ as the true posterior of $\boldsymbol{x}$, with $q \sim \mathcal{N}(\boldsymbol{\mu}_q, \boldsymbol{\Sigma}_q)$ is the approximate Gaussian posterior with a vector mean and a diagonal covariance.
- Consider the case when we optimize $KLD(q||p)$:

$$KLD(q||p) = \int \left[\frac{1}{2}\log\frac{|\boldsymbol{\Sigma}_q|}{|\boldsymbol{\Sigma}_p|} - \frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_p)^T\boldsymbol{\Sigma}_p^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_p) + \frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_q)^T\boldsymbol{\Sigma}_q^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_q)\right] q(\boldsymbol{x})d\boldsymbol{x}$$
$$= \frac{1}{2}\left[\log\frac{|\boldsymbol{\Sigma}_q|}{|\boldsymbol{\Sigma}_p|} + \mathrm{tr}\{\boldsymbol{\Sigma}_q^{-1}\boldsymbol{\Sigma}_p - \mathbf{I}\} + (\boldsymbol{\mu}_q - \boldsymbol{\mu}_p)^T\boldsymbol{\Sigma}_q^{-1}(\boldsymbol{\mu}_q - \boldsymbol{\mu}_p)\right]. \quad (1)$$

computing gradient w.r.t $\boldsymbol{\mu}_q, \boldsymbol{\Sigma}_q$ leads to:
$\boldsymbol{\mu}_q = \boldsymbol{\mu}_p$ and $\boldsymbol{\Sigma}_q = (Diag(\boldsymbol{\Sigma}_p^{-1}))^{-1}$, incorrect posterior variances (correct diagonal precisions)!

- Consider the case when we optimize $KLD(p||q)$:

$$KLD(p||q) = \frac{1}{2}\left[\log\frac{|\boldsymbol{\Sigma}_p|}{|\boldsymbol{\Sigma}_q|} + \mathrm{tr}\{\boldsymbol{\Sigma}_p^{-1}\boldsymbol{\Sigma}_q - \mathbf{I}\} + (\boldsymbol{\mu}_q - \boldsymbol{\mu}_p)^T\boldsymbol{\Sigma}_p^{-1}(\boldsymbol{\mu}_q - \boldsymbol{\mu}_p)\right]. \quad (2)$$

computing gradient w.r.t $\boldsymbol{\mu}_q, \boldsymbol{\Sigma}_q$ leads to:
$\boldsymbol{\mu}_q = \boldsymbol{\mu}_p$ and $\boldsymbol{\Sigma}_q = Diag(\boldsymbol{\Sigma}_p)$. Exact posterior variances!

Generalized Linear Model (GLM)

- Example: GLM, which is essentially a linear mixing model

$$\mathbf{z} = \mathbf{A}\,\boldsymbol{x}\;,\;\; p_{\boldsymbol{x}}(\boldsymbol{x}) = \prod_{i=1}^{N} p_{x_i}(x_i)\,,\;\; p_{\boldsymbol{y}|\mathbf{z}}(\boldsymbol{y}|\mathbf{z}) = \prod_{k=1}^{M} p_{y_k|z_k}(y_k|z_k) \qquad (3)$$

  with (possibly) non identically independently distributed (n.i.i.d.) prior $p_{\boldsymbol{x}}(\boldsymbol{x})$ and n.i.i.d. measurements $p_{\boldsymbol{y}|\mathbf{z}}(\boldsymbol{y}|\mathbf{z})$. (Markov chain $\boldsymbol{x} \to \mathbf{z} \to \boldsymbol{y}$)
- Bayesian estimation: interested in the posterior

$$\begin{aligned} p_{\boldsymbol{x},\mathbf{z}|\boldsymbol{y}}(\boldsymbol{x},\mathbf{z}|\boldsymbol{y}) &= \frac{p_{\boldsymbol{x},\mathbf{z},\boldsymbol{y}}(\boldsymbol{x},\mathbf{z},\boldsymbol{y})}{p_{\boldsymbol{y}}(\boldsymbol{y})} = \frac{p_{\boldsymbol{x}}(\boldsymbol{x})\,p_{\mathbf{z}|\boldsymbol{x}}(\mathbf{z}|\boldsymbol{x})\,p_{\boldsymbol{y}|\mathbf{z}}(\boldsymbol{y}|\mathbf{z})}{p_{\boldsymbol{y}}(\boldsymbol{y})} \\ &= \frac{1}{Z(\boldsymbol{y})} e^{-\sum_{i=1}^{N} f_{x_i}(x_i) - \sum_{k=1}^{M} f_{z_k}(z_k)}\,\delta(\mathbf{z}-\mathbf{A}\boldsymbol{x}) \end{aligned} \qquad (4)$$

  where we have the negative loglikelihoods for prior and measurements

$$f_{x_i}(x_i) = -\ln p_{x_i}(x_i)\,,\;\; f_{z_k}(z_k) = -\ln p_{y_k|z_k}(y_k|z_k) \qquad (5)$$

  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 $Z(\boldsymbol{y})$).
- The problem in Bayesian estimation is the computation of this constant $Z(\boldsymbol{y})$ and of the posterior means and variances.

Variational Free Energy (VFE) Minimization

- Here $\boldsymbol{y}$ = data, $\boldsymbol{x}$ = all variables (both $\boldsymbol{x}$ and $\mathbf{z}$ in GLM).
- Bayes posterior

$$p(\boldsymbol{x}|\boldsymbol{y}) = \frac{p(\boldsymbol{x},\boldsymbol{y})}{Z}\,,\quad Z = p(\boldsymbol{y}) = \int p(\boldsymbol{x},\boldsymbol{y})\,d\boldsymbol{x} \tag{6}$$

- Consider an approximate posterior $q(\boldsymbol{x})$. The Variational (or Gibbs) Free Energy (VFE) is

$$\begin{aligned} F(q) &= D(q(\boldsymbol{x})||p(\boldsymbol{x},\boldsymbol{y})) = -\int q(\boldsymbol{x})\,\ln p(\boldsymbol{x},\boldsymbol{y}\,d\boldsymbol{x}) + \int q(\boldsymbol{x})\,\ln q(\boldsymbol{x})\,d\boldsymbol{x} \\ &= \text{energy - entropy} = -\ln Z + D(q(\boldsymbol{x})||p(\boldsymbol{x}|\boldsymbol{y})) \end{aligned} \tag{7}$$

  where $-\ln Z$ = Helmhotz Free Energy.
  Alternatively

$$\ln Z = -F(q) + D(q(\boldsymbol{x})||p(\boldsymbol{x}|\boldsymbol{y})) = \text{ evidence.} \tag{8}$$

  Minimizing $F(q)$ is equivalent to maximizing $-F(q)$, the Evidence Lower Bound (ELBO).
  Does not require to know Z, but allows to find or approximate it.
- Minimizing $F(q)$ over unrestricted $q(\boldsymbol{x})$ yields $F(q) = -\ln Z$ for $q(\boldsymbol{x}) = p(\boldsymbol{x}|\boldsymbol{y})$, hence allows to find/approximate $p(\boldsymbol{x}|\boldsymbol{y})$.
- In practice, restrict $q(\boldsymbol{x})$ to some feasible set.

Variational Bayes - Mean Field

- Restrict

$$q(\boldsymbol{x}) = \prod_i q_i(\boldsymbol{x}_i) \tag{9}$$

for a certain partition $\{\boldsymbol{x}_i\}$ of $\boldsymbol{x}$.

- If the x_i are scalars, Variational Bayes (VB) is called Mean Field (MF).
- Alternating minimization yields

$$\begin{aligned} q_i(\boldsymbol{x}_i) &= \arg\min_{q'_i(\boldsymbol{x}_i)} F(q'_i(\boldsymbol{x}_i)\, q(\boldsymbol{x}_{\bar{i}})\ , \quad q(\boldsymbol{x}_{\bar{i}}) = \prod_{j \neq i} q_j(\boldsymbol{x}_j) \\ &\sim e^{\int q(\boldsymbol{x}_{\bar{i}})\, \ln p(\boldsymbol{x}, \boldsymbol{y})\, d\boldsymbol{x}_{\bar{i}}} \end{aligned} \tag{10}$$

Or $\ln q_i(\boldsymbol{x}_i) = \mathbb{E}_{q(\boldsymbol{x}_{\bar{i}})} \ln p(\boldsymbol{x}, \boldsymbol{y}) + c^t$.

$F(q)$ being convex in each $q_i(\boldsymbol{x}_i)$ separately, alternating minimization leads to a local minimum.

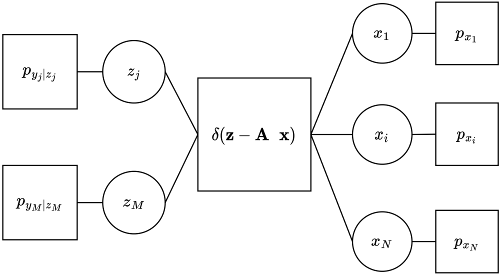
pdf Factorizations and Factor Graphs

- Joint pdf factorization into $M + N + 1$ factors

$$p(\boldsymbol{x}, \mathbf{z}, \boldsymbol{y}) = \delta(\mathbf{z} - \mathbf{A}\boldsymbol{x}) \prod_{i=1}^{N} p_{x_i}(x_i) \prod_{k=1}^{M} p_{y_k|z_k}(y_k|z_k) \qquad (11)$$

where $\delta(\mathbf{z} - \mathbf{A}\boldsymbol{x}) = \prod_{k=1}^{M} \delta(z_k - \boldsymbol{a}_k^T \boldsymbol{x})$, $\mathbf{A}^T = [\boldsymbol{a}_1 \cdots \boldsymbol{a}_M]$.

- Factor graph without cycles.



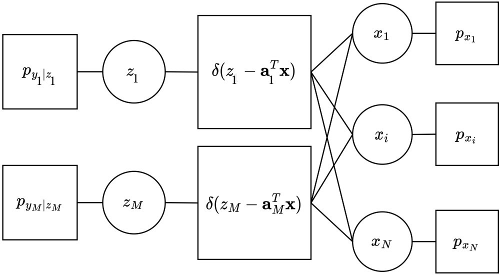
Circles: variable nodes, squares: factor nodes.

Another Factorization and Corresponding Factor Graph

- Joint pdf factorization into $2M + N$ factors

$$p(\boldsymbol{x}, \mathbf{z}, \boldsymbol{y}) = \prod_{i=1}^{N} p_{x_i}(x_i) \prod_{k=1}^{M} p_{y_k|z_k}(y_k|z_k)\, \delta(z_k - \boldsymbol{a}_k^T \boldsymbol{x})\,. \qquad (12)$$

- Factor graph with cycles.



- Factorizations not unique! $\Rightarrow$ BFE, EP not unique.

Expectation Propagation (EP) (Minka style)

- Factorization of joint pdf

$$p(\boldsymbol{x}, \boldsymbol{y}) = \prod_a p_a(\boldsymbol{x}_a) \tag{13}$$

where the $\boldsymbol{x}_a$ are (possibly overlapping) subsets of $\boldsymbol{x}$.
We're not interested in how $p_a(\boldsymbol{x}_a)$ depends on $\boldsymbol{y}$.

- EP posterior approximation [1]

$$q(\boldsymbol{x}) = \frac{1}{Z_q} \prod_a q_a(\boldsymbol{x}_a) \tag{14}$$

similar to p but the $q_s(\boldsymbol{x}_a)$ are in an exponential family $\mathcal{F}$ with sufficient statistic $\phi(\boldsymbol{x})$. Then $q(\boldsymbol{x})$ is also in this exponential family (closure under pdf multiplication/division).

- Alternating updating: for any q_a, with $q_{\overline{a}}(\boldsymbol{x}) = \prod_{b \neq a} q_b(\boldsymbol{x}_b) \sim q(\boldsymbol{x})/q_a(\boldsymbol{x}_a)$,

$$\begin{aligned}
&\widetilde{p}_a(\boldsymbol{x}) = \tfrac{1}{Z_a}\, p_a(\boldsymbol{x}_a)\, q_{\overline{a}}(\boldsymbol{x}), \quad \text{tilted posterior approximation} \\
&\widetilde{q}_a(\boldsymbol{x}) = \arg\min_{\widetilde{q}'_a \in \mathcal{F}} D(\widetilde{p}_a || \widetilde{q}'_a) = Proj_{\mathcal{F}}\{\widetilde{p}_a\} : \ \mathbb{E}_{\widetilde{q}_a}\, \phi(\boldsymbol{x}) = \mathbb{E}_{\widetilde{p}_a}\, \phi(\boldsymbol{x}) \\
&q_a(\boldsymbol{x}_a) = \widetilde{q}_a(\boldsymbol{x})/q_{\overline{a}}(\boldsymbol{x}) \quad \text{local KLD} \qquad \text{moment matching}
\end{aligned} \tag{15}$$

- Extremes: $q_a(\boldsymbol{x}_a)$ fully factorized, $q_a(\boldsymbol{x}_a) = \prod_i q_{ai}(x_i)$, or not at all, $q_a(\boldsymbol{x})$.
- Usually overlooked: the tilted posteriors $\widetilde{p}_a(\boldsymbol{x})$, which are outside the exponential family, could be better approximations than $q(\boldsymbol{x})$.

Bethe Free Energy (BFE) Minimization

- Introduce two sets of approximating factors, $q_a(\boldsymbol{x}_a)$ at factor level and $q_i(x_i)$ at variable level.

$$\min_q \; D(q||p) \text{ with } q(\boldsymbol{x}) = \frac{\prod_a q_a(\boldsymbol{x}_a)}{\prod_i (q_i(x_i))^{N_i-1}}, \; p = p(\boldsymbol{x}, \boldsymbol{y})$$

under consistency requirements: $q_a(x_i) = q_i(x_i)\,, \; \forall i, \forall a \in \mathcal{N}_i$

where $\mathcal{N}_i = \{a : x_i \in \boldsymbol{x}_a\}$, $N_i = |\mathcal{N}_i|$, $\mathcal{N}_a = \{i : x_i \in \boldsymbol{x}_a\}$.

- BFE

$$D(q||p) = F_B(\{q_a\}, \{q_i\}) = \sum_a D(q_a||p_a) + \sum_i (N_i - 1)\, H(q_i)$$

with entropies $H(q) = -\int q(x) \ln q(x)\, dx$.

- Lagrangian with consistency and normalization constraints

$$L(q) = \quad F_B(q) + \textstyle\sum_a \lambda_a(\int q_a(\boldsymbol{x}_a)\, d\boldsymbol{x}_a - 1) + \displaystyle\sum_{i:N_i>1} \lambda_i(\int q_i(x_i)\, dx_i - 1)$$
$$+ \sum_{i:N_i>1} \sum_{a \in \mathcal{N}_i} \int \lambda_{ai}(x_i)\, (q_i(x_i) - \int q_a(\boldsymbol{x}_a)\, d\boldsymbol{x}_{a\setminus i})\, dx_i$$

- Solving for extrema:

$$q_a(\boldsymbol{x}_a) = p_a(\boldsymbol{x}_a)\, \exp[\lambda_a - 1 + \sum_{i \in \mathcal{N}_a} \lambda_{ai}(x_i)]$$
$$q_i(x_i) = \exp[\tfrac{1}{N_i - 1}(1 - \lambda_i + \sum_{a \in \mathcal{N}_i} \lambda_{ai}(x_i))]$$

BFE Minimization: Belief Propagation (BP)

- Introduce $\lambda_{ai}(x_i) = \ln m_{i\to a}(x_i)\,,$
- then Belief Propagation cycles through the updates

$$m_{a\to i}(x_i) = \int q_a(\boldsymbol{x}_a)/m_{i\to a}(x_i)\, d\boldsymbol{x}_{a\setminus i} = \int p_a(\boldsymbol{x}_a) \prod_{j\in\mathcal{N}_a\setminus i} m_{j\to a}(x_j)\, d\boldsymbol{x}_{a\setminus i}$$

$$m_{i\to a}(x_i) = \prod\nolimits_{c\in\mathcal{N}_i\setminus a} m_{c\to i}(x_i)$$

- with

$$q_a(\boldsymbol{x}_a) \sim p_a(\boldsymbol{x}_a) \prod_{i\in\mathcal{N}_a} m_{i\to a}(x_i) = p_a(\boldsymbol{x}_a) \prod_{i\in\mathcal{N}_a} \prod_{c\in\mathcal{N}_i\setminus a} m_{c\to i}(x_i)$$

$$q_i(x_i) \sim \prod_{a\in\mathcal{N}_i} m_{a\to i}(x_i) \;\; (= m_{a\to i}(x_i)\, m_{i\to a}(x_i)\,,\ \forall a \in \mathcal{N}_i)$$

- At the level of the messages, everything is at variable level. The multivariate factors p_a only appear as multivariate in their approx's q_a.
- The BFE entropy terms are non-convex $\Rightarrow$ convex majorizer:

$$\begin{aligned} F_B(q) \leq F_B^m(q) \;\; &= \sum\nolimits_a D(q_a||p_a) + \sum\nolimits_i (N_i-1)\,(H(q_i) + D(q_i||q_i^{t-1})) \\ &= \sum\nolimits_a D(q_a||p_a) - \sum\nolimits_i (N_i-1) \int dx_i q_i(x_i)\, \ln q_i^{t-1}(x_i) \end{aligned}$$

where the q_i^{t-1} are the q_i from the previous iteration $t-1$. Majorization does not require a double loop, unlike [2].

BFE Minimization with Moment Constraints: Expectation Propagation (EP)

- BP can be untractable due to products of pdfs.
- Relax consistency constraints $q_a(x_i) = q_i(x_i)\,,\ \forall i, \forall a \in \mathcal{N}_i$ to moment constraints for some sufficient statistics $\phi(\boldsymbol{x})$ for exponential family of pdfs $\mathcal{F}$

$$\mathbb{E}_{q_a(x_i)}\,\phi(x_i) = \mathbb{E}_{q_i(x_i)}\,\phi(x_i)\,,\ \ \forall i, \forall a \in \mathcal{N}_i$$

  leads to messages in $\mathcal{F}$, which is closed under pdf multiplication.
- The only change in BP to get EP:

$$m_{a\to i}(x_i) = \frac{Proj_{\mathcal{F}}\{\int q_a(\boldsymbol{x}_a)\, d\boldsymbol{x}_{a\setminus i}\}}{m_{i\to a}(x_i)}$$

- If one removes the projection operation, EP falls back on BP.
- In EP only exponential family messages propagate. At convergence one gets also the $q_i(x_i)$ in the exponential family, but the $q_a(\boldsymbol{x}_a)$ are more general due to the presence of the original factor $p_a(\boldsymbol{x}_a)$.
- BFE perspective: EP also involves defining $\{q_a\}$, $\{q_i\}$, resulting BFE.
- BP and EP can be extended to mix with VB, by adding a factorized portion to the BFE posterior model to be plugged into the VFE, leading to e.g. mixed EP-VB algorithms, see [3], [4].

Outline

1. Bethe Free Energy (BFE) Minimization and Expectation Propagation (EP) - min KLD

2. reVAMP: revisited VAMP
   - EP-like Derivation
   - Relation to CWCU MMSE Estimation

Introduction

- The recovery of signal vectors is a fundamental problem in signal processing.
- Even in lower dimensions, the application of Bayesian estimation (e.g., Minimum Mean Squared Error (MMSE)) becomes challenging in a non-Gaussian scenario due to the intractability.
- Approximate Message Passing (AMP) demonstrated effectiveness in recovering high-dimensional signals. However, poor convergence properties when dealing with ill-conditioned measurement matrix.
- Vector AMP (VAMP) is robust for ill-conditioned measurement matrix. However, it cannot predict element-wise posterior variances.

Contribution

- This work adopts a similar Expectation-Propagation (EP)-like derivation, as described in [5] to derive revisited VAMP (reVAMP).
- This algorithm provides individual MSE and posterior covariance matrix as a byproduct. As a trade-off, it has a complexity of $O(N^3)$ per iteration.
- This work explores the relationship between the CWCU estimator and the derivation of extrinsic (denoiser input) in reVAMP, and extends the CWCU estimator by considering non-zero prior mean.

System Model (Gaussian Noise case)

- Consider the linear mixing data model:

$$\boldsymbol{y} = \mathbf{A}\boldsymbol{x} + \boldsymbol{v}, \quad p_{\boldsymbol{x}}(\boldsymbol{x}), \quad p_{\boldsymbol{v}}(\boldsymbol{v}), \tag{16}$$

with N n.i.i.d. inputs

$$p_{\boldsymbol{x}}(\boldsymbol{x}) = \prod_{i=1}^{N} p_{x_i}(x_i), \tag{17}$$

and Gaussian noise of size M

$$p_{\boldsymbol{v}}(\boldsymbol{v}) = \mathcal{N}(\boldsymbol{v}; \mathbf{0}, \mathbf{C}_{\boldsymbol{v}\boldsymbol{v}}). \tag{18}$$

Outline

1. Bethe Free Energy (BFE) Minimization and Expectation Propagation (EP) - min KLD

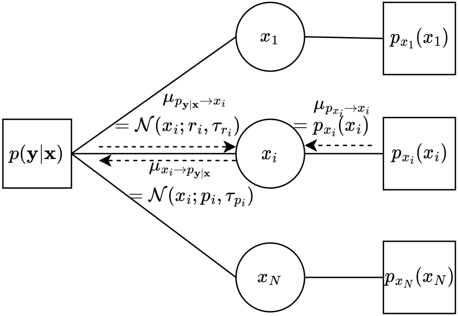
2. reVAMP: revisited VAMP
   - EP-like Derivation
   - Relation to CWCU MMSE Estimation

Factorization

- Factorization scheme:

$$p(\boldsymbol{x}, \boldsymbol{y}) = p(\boldsymbol{y}|\boldsymbol{x}) \prod_{i=1}^{N} p_{x_i}(x_i). \tag{19}$$

- Factor graph:



reVAMP Motivation : Gaussian extrinsics approximation

- reVAMP is motivated by only a single asymptotic approximation: the asymptotic Gaussianity of extrinsics.
  The extrinsic pdf of a variable x_i is the conditional pdf $p(\boldsymbol{y}|x_i)$, in which x_i is treated as a deterministic variable (no prior information), but the other variables $\boldsymbol{x}_{\bar{i}}$ remain random and their prior pdf is exploited to eliminate them from the joint pdf. The randomness of $\boldsymbol{x}$ and $\mathbf{A}$ will quickly lead to Gaussianity of $p(\boldsymbol{y}|x_i)$ by the CLT (think of asymptotic Gaussianity of Maximum Likelihood estimates).
- reVAMP introduces both Gaussian and non-Gaussian marginal posteriors from Gaussian extrinsics and the true prior. This involves also the introduction of Gaussian approximations for the priors. Which in turn also leads to a multivariate Gaussian posterior approximation, which exhibits the posterior correlations between the variables.
- reVAMP postulates a factored posterior approximation of the form $q(\boldsymbol{x}) = \prod_i q(x_i) = \prod_i m_i(x_i)\, q_i(x_i)$ where the $m_i(x_i)$ are the Gaussian extrinsics and the $q_i(x_i)$ are Gaussian approximations to the priors $p(x_i)$.
- A byproduct are non-Gaussian posterior marginals of the form $m_i(x_i)\, p(x_i)$ where $p(x_i)$ is the true prior for x_i. Note that involving the true priors is something that could also be considered in VB. But unlike VB, reVAMP attemps to optimize the better KLD, KLD(p, q).

reVAMP Motivation (2)

- Consider optimizing $q(\boldsymbol{x}) = \prod_i m_i(x_i)\, q_i(x_i)$ by minimizing

$$\mathsf{KLD}(p(\boldsymbol{x}|\boldsymbol{y})||q(\boldsymbol{x})) = \sum_{i=1}^{N} \mathsf{KLD}(p(\boldsymbol{x}|\boldsymbol{y})||q(x_i)) + (N-1)\, H(p(\boldsymbol{x}|\boldsymbol{y}))\ . \quad (20)$$

  We can minimize alternatingly w.r.t. the factors $q(x_i)$. We get apart from an additive constant

$$\mathsf{KLD}(p(\boldsymbol{x}|\boldsymbol{y})||q(x_i)) = \mathsf{KLD}(p(x_i|\boldsymbol{y})||q(x_i)) - H(p(\boldsymbol{x}|\boldsymbol{y})) + H(p(x_i|\boldsymbol{y})) \quad (21)$$

  The true posterior for x_i can be written as

$$p(x_i|\boldsymbol{y}) = \underbrace{p_{x_i}(x_i)}_{\text{prior}} \underbrace{\left(\int p(\boldsymbol{y}|\boldsymbol{x}) \prod_{j\neq i}^{N} p_{x_j}(x_j) dx_j \right)}_{\text{extrinsic } p(\boldsymbol{y}|x_i)} / Z_i(\boldsymbol{y}), \quad (22)$$

- For a very large class of models for $\mathbf{A}$ and $\boldsymbol{x}$, it is clear that the CLT will allow to approximate the extrinsic $p(\boldsymbol{y}|x_i)$ by a Gaussian distribution $m_i(x_i)$

$$p(\boldsymbol{y}|x_i) = \int p(\boldsymbol{y}|\boldsymbol{x}) \prod_{j\neq i}^{N} p_{x_j}(x_j) dx_j \approx m_i(x_i) = \mathcal{N}(x_i; r_i, \tau_{r_i})\ . \quad (23)$$

reVAMP Motivation (3)

- Hence we need to consider the minimization of
$$\text{KLD}(p(x_i|\boldsymbol{y})||q(x_i)) = \text{KLD}(p(x_i)p(\boldsymbol{y}|x_i)/Z_i||q(x_i))$$
$$\approx \text{KLD}(p(x_i)m_i(x_i)/Z_i||q(x_i)) = \text{KLD}(p(x_i)m_i(x_i)/Z_i||q_i(x_i)m_i(x_i)/Z_i')\,. \tag{24}$$
- The reVAMP algorithm [6] approximates the posterior to Gaussian with the approximated Gaussian extrinsic:
$$p(x_i|\boldsymbol{y}) \approx p(x_i)\mathcal{N}\left(x_i; r_i, \tau_{r_i}\right)/Z_i(\boldsymbol{y}) \approx \mathcal{N}\left(x_i; \widehat{x}_i, \tau_{x_i}\right) = q(x_i)\,. \tag{25}$$
The approximate Gaussian posterior $q(x_i)$ is obtained by moment matching with the better posterior approximation $p_{x_i}(x_i)\, m_i(x_i)/Z_i$.
- We interpret the quotient of the approximated posterior and the approximate extrinsic as the approximated Gaussian prior.
$$p_{x_i}(x_i) \approx q_i(x_i) = \mathcal{N}\left(x_i; a_i, \sigma^2_{x_i}\right) \propto \mathcal{N}\left(x_i; \widehat{x}_i, \tau_{x_i}\right)/\mathcal{N}\left(x_i; r_i, \tau_{r_i}\right),$$
$$1/\sigma^2_{x_i} = 1/\tau_{x_i} - 1/\tau_{r_i}\ ,\ a_i = \sigma^2_{x_i}(\widehat{x}_i/\tau_{x_i} - r_i/\tau_{r_i})\,. \tag{26}$$
This Gaussian approximation $q_i(x_i)$ does not correspond to direct moment matching of the true prior $p_{x_i}(x_i)$.

reVAMP Motivation (4)

- Hence reVAMP does alternating minimization of KLD(p, q) which becomes iterative because an extrinsic $m_i(x_i)$ depends on the approximate Gaussian priors $\prod_{j \neq i} q_j(x_j)$. Since alternating minimization of a convex cost function converges, reVAMP can be expected to converge.
- Apart from the improved marginal posteriors $m_i(x_i) p_{x_i}(x_i)/Z_i'$, reVAMP also produces the joint Gaussian posterior approximation $q'(\boldsymbol{x}) = p(\boldsymbol{y}|\boldsymbol{x}) \prod_i q_i(x_i)/Z'$.
- The Gaussian extrinsics approximations $p(x_i|\boldsymbol{y}) \approx m_i(x_i)$ are asymptotically tight. The Gaussian approximations that are not tight and that constitute the variational approximations are approximating marginal posteriors by Gaussian $q(x_i)$ or what follows from that, approximating priors $p_{x_i}(x_i)$ by Gaussian $q_i(x_i)$. Or the overall multivariate Gaussian posterior approximation is not tight also, but at least captures full second-order moments.
- re(G)VAMP can be derived using EP, see GLM BFE considered later.

reGVAMP

KLD optimization $\arg\min_{q_{\boldsymbol{x},\mathbf{z}|\boldsymbol{y}}} KLD[p(\boldsymbol{x},\mathbf{z}|\boldsymbol{y})\|q_{\boldsymbol{x},\mathbf{z}|\boldsymbol{y}}(\boldsymbol{x},\mathbf{z})]$, with approximate posterior

$$\begin{aligned} q_{\boldsymbol{x},\mathbf{z}}(\boldsymbol{x},\mathbf{z}) &= \prod_i q_{x_i|\boldsymbol{y}}(x_i)\prod_j q_{z_j|\boldsymbol{y}}(z_j) \\ &= \prod_i q_{x_i}(x_i)m_{x_i}(x_i)\prod_j q_{z_j}(z_j)m_{z_j}(z_j), \end{aligned} \tag{27}$$

where q_{x_i} and q_{z_j} are the approximated prior and likelihood while m_{x_i} and m_{z_j} are the extrinsic for x_i and z_j. The KLD becomes

$$\begin{aligned} &KLD[p(\boldsymbol{x},\mathbf{z}|\boldsymbol{y})\|q_{\boldsymbol{x}|\boldsymbol{y}}(\boldsymbol{x})]+KLD[p(\boldsymbol{x},\mathbf{z}|\boldsymbol{y})\|q_{\mathbf{z}|\boldsymbol{y}}(\mathbf{z})]+\text{const} \\ &= \sum_i KLD[p(\boldsymbol{x},\mathbf{z}|\boldsymbol{y})\|q_{x_i|\boldsymbol{y}}(x_i)] + \sum_j KLD[p(\boldsymbol{x},\mathbf{z}|\boldsymbol{y})\|q_{z_j|\boldsymbol{y}}(z_j)] + c^t \\ &= \sum_i KLD[p(x_i|\boldsymbol{y})\|q_{x_i|\boldsymbol{y}}(x_i)] + \sum_j KLD[p(z_j|\boldsymbol{y})\|q_{z_j|\boldsymbol{y}}(z_j)] + c^t \end{aligned}$$

In the last equality, we marginalize out the irrelevant variables. The marginalized posteriors $p(x_i|\boldsymbol{y})$ and $p(z_j|\boldsymbol{y})$ are

$$p(x_i|\boldsymbol{y}) \propto \underbrace{p_{x_i}(x_i)}_{\text{prior}} \underbrace{\int p(\boldsymbol{y}|\mathbf{z})p(\mathbf{z}|\boldsymbol{x})\prod_{k\neq i} p_{x_k}(x_k)d\mathbf{z}d\boldsymbol{x}_{\bar{i}}}_{\text{extrinsic } p(\boldsymbol{y}|x_i)},$$

$$p(z_j|\boldsymbol{y}) \propto p(\boldsymbol{y},z_j) = \int p(\boldsymbol{y},\mathbf{z})d\mathbf{z}_{\bar{j}} = \underbrace{p_{y_j|z_j}(z_j)}_{\text{prior}} \underbrace{\int \prod_{k\neq j} p_{y_k|z_k}(z_k)\delta(\mathbf{z}-\mathbf{A}\boldsymbol{x})p(\boldsymbol{x})d\boldsymbol{x}d\mathbf{z}_{\bar{j}}}_{\text{extrinsic } p(\boldsymbol{y}_{\bar{j}},z_j)}$$

reGVAMP (2)

In order to see which probability the extrinsic for z corresponds to, consider short hand notation

$$p(\mathbf{z}) = \int \delta(\mathbf{z} - \mathbf{A}\boldsymbol{x})p(\boldsymbol{x})d\boldsymbol{x} = p(\mathbf{z}_{\bar{j}}|z_j)p(z_j), \tag{28}$$

which depend only on the prior for $\boldsymbol{x}$. Therefore,

$$\begin{aligned} &\int \prod_{k \neq j} p_{y_k|z_k}(z_k)\delta(\mathbf{z} - \mathbf{A}\boldsymbol{x})p(\boldsymbol{x})d\boldsymbol{x} \\ &= p_{\boldsymbol{y}_{\bar{j}}|\mathbf{z}_{\bar{j}}}(\mathbf{z}_{\bar{j}})p(\mathbf{z}_{\bar{j}}|z_j)p(z_j) = p(\boldsymbol{y}_{\bar{j}}, \mathbf{z}_{\bar{j}}, z_j), \end{aligned} \tag{29}$$

Thus, we have

$$p(x_i|\boldsymbol{y}) \simeq p_{x_i}(x_i)m_{x_i}(x_i),\ p(z_j|\boldsymbol{y}) \simeq p_{y_j|z_j}(z_j)m_{z_j}(z_j).$$

Due to CLT, extrinsics can be approximated as Gaussian. The marginal KLDs become

$$\arg\min_{q_{x_i|\boldsymbol{y}}} KLD[p(x_i|\boldsymbol{y})\|q_{x_i|\boldsymbol{y}}(x_i)] \simeq \arg\min_{q_{x_i}} KLD[p_{x_i}(x_i)m_{x_i}(x_i)\|q_{x_i}(x_i)m_{x_i}(x_i)] \tag{30}$$

$$\arg\min_{q_{z_j|\boldsymbol{y}}} KLD[p(z_j|\boldsymbol{y})\|q_{z_j|\boldsymbol{y}}(z_j)] \simeq \arg\min_{q_{z_j}} KLD[p_{y_j|z_j}(z_j)m_{z_j}(z_j)\|q_{z_j}(z_j)m_{z_j}(z_j)]. \tag{31}$$

(re)GVAMP and Bayesian Cramer-Rao Bound (CRB)

- Bayesian CRBs are notoriously loose when distributions are non-Gaussian.
- In Bayesian estimation the tight MMSE lower bound is the MSE achieved by MMSE estimation.
- reGVAMP provides local MMSE estimates and associated MSE in which the only approximation is the Gaussian approximation of extrinsics.
- Note that a joint vector MMSE estimate is a vector of scalar MMSE estimates. MMSE estimation is local.
- To the extent that extrinsics can be approximated by Gaussians (or more generally by the exponential family used), Expectation Propagation performs (approximate) alternating minimization of $D(p||q)$.
  A similar motivation was mentioned by Minka [1], but the motivation is only justifiable for (marginalized) extrinsics, not for joint pdfs as mentioned in [1].

Outline

1. Bethe Free Energy (BFE) Minimization and Expectation Propagation (EP) - min KLD

2. reVAMP: revisited VAMP
   - EP-like Derivation
   - Relation to CWCU MMSE Estimation

Extrinsics and Component-Wise Conditionally Unbiased (CWCU) MMSE Estimation

- CWCU MMSE estimation was introduced in [7], [8] and further elaborated in [9], where a detailed derivation can be found and conditions are provided for the existence of such estimators.
- Derivation from extrinsics: consider jointly Gaussian $\boldsymbol{y}$ and x (scalar)
$\left[\begin{array}{c}\boldsymbol{y}\\ x\end{array}\right] \sim \mathcal{N}\left(\left[\begin{array}{c}\mathbf{m}_{\boldsymbol{y}}\\ \mathbf{m}_x\end{array}\right], \left[\begin{array}{cc}\mathbf{C}_{\boldsymbol{y}\boldsymbol{y}} & \mathbf{C}_{\boldsymbol{y}x}\\ \mathbf{C}_{x\boldsymbol{y}} & \mathbf{C}_{xx}\end{array}\right]\right)$ (so, $\mathbf{m}_x$ and $\mathbf{C}_{xx}$ are scalar).
Then the extrinsic $p(\boldsymbol{y}|x)$ is Gaussian and
$$\begin{aligned}&-2\ln p(\boldsymbol{y}|x) = c^t + (\boldsymbol{y}-\mathbf{m}_{\boldsymbol{y}|x})^T \mathbf{C}_{\boldsymbol{y}|x}^{-1}(\boldsymbol{y}-\mathbf{m}_{\boldsymbol{y}|x}),\ \text{ with}\\ &\mathbf{m}_{\boldsymbol{y}|x} = \mathbf{m}_y + \mathbf{C}_{\boldsymbol{y}x}\mathbf{C}_{xx}^{-1}(\boldsymbol{x}-\mathbf{m}_{\boldsymbol{x}}),\ \mathbf{C}_{\boldsymbol{y}|x} = \mathbf{C}_{\boldsymbol{y}\boldsymbol{y}} - \mathbf{C}_{\boldsymbol{y}x}\mathbf{C}_{xx}^{-1}\mathbf{C}_{x\boldsymbol{y}}\end{aligned} \tag{32}$$
Reinterpreting as a pdf in x, we can rewrite this quadratic exponent as
$$\begin{aligned}&-2\ln p(\boldsymbol{y}|x) = c(\boldsymbol{y}) + (x-\widehat{x}_{CL})^2/\mathbf{C}_{\widetilde{x}_{CL}\widetilde{x}_{CL}},\ \ d = \frac{\mathbf{C}_{xx}}{\mathbf{C}_{x\boldsymbol{y}}\mathbf{C}_{\boldsymbol{y}\boldsymbol{y}}^{-1}\mathbf{C}_{\boldsymbol{y}x}} \geq 1,\\ &\widehat{x}_{CL} = \mathbf{m}_x + d\,\mathbf{C}_{x\boldsymbol{y}}\mathbf{C}_{\boldsymbol{y}\boldsymbol{y}}^{-1}(\boldsymbol{y}-\mathbf{m}_{\boldsymbol{y}}) = d\,\widehat{x}_L + (1-d)\,\mathbf{m}_x\\ &\qquad \text{with } \widehat{x}_L = \mathbf{m}_x + \mathbf{C}_{x\boldsymbol{y}}\mathbf{C}_{\boldsymbol{y}\boldsymbol{y}}^{-1}(\boldsymbol{y}-\mathbf{m}_{\boldsymbol{y}}),\\ &\mathbf{C}_{\widetilde{x}_{CL}\widetilde{x}_{CL}} = d\mathbf{C}_{\widetilde{x}_L\widetilde{x}_L},\ \mathbf{C}_{\widetilde{x}_L\widetilde{x}_L} = \mathbf{C}_{xx} - \mathbf{C}_{x\boldsymbol{y}}\mathbf{C}_{\boldsymbol{y}\boldsymbol{y}}^{-1}\mathbf{C}_{\boldsymbol{y}x}\end{aligned} \tag{33}$$
where $\widehat{x}_{CL}$, $\mathbf{C}_{\widetilde{x}_{CL}\widetilde{x}_{CL}}$ are the CWCU LMMSE estimate and error variance, and $\widehat{x}_L$, $\mathbf{C}_{\widetilde{x}_L\widetilde{x}_L}$ are the LMMSE (and hence MMSE since Gaussian) estimate and error variance.

Extrinsics and CWCU MMSE Estimation (2)

- Now interpreting the previous x as a component x_i of a vector $\boldsymbol{x}$, we can write

$$
\begin{aligned}
&\widehat{\boldsymbol{x}}_{CL} = \mathbf{m}_{\boldsymbol{x}} + \mathbf{D}\,\mathbf{C}_{\boldsymbol{x}\boldsymbol{y}}\mathbf{C}_{\boldsymbol{y}\boldsymbol{y}}^{-1}(\boldsymbol{y}-\mathbf{m}_{\boldsymbol{y}}) = \mathbf{D}\,\widehat{\boldsymbol{x}}_L + (\mathbf{I}-\mathbf{D})\,\mathbf{m}_x \\
&\quad \text{with } \mathbf{D} = \text{diag}(\mathbf{C}_{\boldsymbol{x}\boldsymbol{x}})(\text{diag}(\mathbf{C}_{\widehat{\boldsymbol{x}}_L\widehat{\boldsymbol{x}}_L}))^{-1},\ \mathbf{C}_{\widehat{\boldsymbol{x}}_L\widehat{\boldsymbol{x}}_L} = \mathbf{C}_{\boldsymbol{x}\boldsymbol{y}}\mathbf{C}_{\boldsymbol{y}\boldsymbol{y}}^{-1}\mathbf{C}_{\boldsymbol{y}\boldsymbol{x}} \\
&\mathbf{C}_{\widetilde{\boldsymbol{x}}_{CL}\widetilde{\boldsymbol{x}}_{CL}} = \mathbf{C}_{\widetilde{\boldsymbol{x}}_L\widetilde{\boldsymbol{x}}_L} + (\mathbf{D}-\mathbf{I})\mathbf{C}_{\widehat{\boldsymbol{x}}_L\widehat{\boldsymbol{x}}_L}(\mathbf{D}-\mathbf{I})
\end{aligned}
\tag{34}
$$

  where the last identity follows from $\widetilde{\boldsymbol{x}}_{CL} = \boldsymbol{x} - \widehat{\boldsymbol{x}}_{CL} = \widetilde{\boldsymbol{x}}_L - (\mathbf{D}-\mathbf{I})\,\mathbf{C}_{\boldsymbol{x}\boldsymbol{y}}\mathbf{C}_{\boldsymbol{y}\boldsymbol{y}}^{-1}(\boldsymbol{y}-\mathbf{m}_{\boldsymbol{y}})$ and the two terms in this difference are decorrelated by the orthogonality property of LMMSE.
- Recall the definition of CWCU LMMSE:

$$
\widehat{x}_{i,CL} = \mathbf{f}_i^T\boldsymbol{y} + g_i,\quad \min_{\mathbf{f}_i, g_i:\, \mathbb{E}_{\boldsymbol{y}|x_i}\widehat{x}_{i,CL}=x_i} \mathbb{E}(x_i - \widehat{x}_{i,CL})^2\,. \tag{35}
$$

- The assumption of jointly Gaussian $\boldsymbol{y}, \boldsymbol{x}$ can be extended to a linear model with pairwise Gaussian $\boldsymbol{x}$ components and arbitrary noise, or decorrelated Gaussian noise and arbitrary independent priors.

Extrinsics and CWCU MMSE Estimation (3)

- We'll show: $\mathbf{D} = D(\boldsymbol{\tau}_{CL}./\boldsymbol{\tau}_L)$,
  $\boldsymbol{\tau}_L = \text{diag}_M(\mathbf{C}_{\widetilde{\boldsymbol{x}}_L\widetilde{\boldsymbol{x}}_L})$, $\boldsymbol{\tau}_{CL} = \text{diag}_M(\mathbf{C}_{\widetilde{\boldsymbol{x}}_{CL}\widetilde{\boldsymbol{x}}_{CL}})$, $D(\boldsymbol{\tau}) = \text{diag}_M(\boldsymbol{\tau})$
-
$$\begin{aligned}
\mathbf{C}_{\widetilde{\boldsymbol{x}}_{CL}\widetilde{\boldsymbol{x}}_{CL}} &= \mathbf{C}_{\widetilde{\boldsymbol{x}}_L\widetilde{\boldsymbol{x}}_L} + (\mathbf{D}-\mathbf{I})\mathbf{C}_{\widehat{\boldsymbol{x}}_L\widehat{\boldsymbol{x}}_L}(\mathbf{D}-\mathbf{I}) \\
&= \mathbf{C}_{\boldsymbol{x}\boldsymbol{x}} - \mathbf{C}_{\widehat{\boldsymbol{x}}_L\widehat{\boldsymbol{x}}_L}\mathbf{D} - \mathbf{D}\,\mathbf{C}_{\widehat{\boldsymbol{x}}_L\widehat{\boldsymbol{x}}_L} + \mathbf{D}\,\mathbf{C}_{\widehat{\boldsymbol{x}}_L\widehat{\boldsymbol{x}}_L}\mathbf{D}
\end{aligned}$$
$$\begin{aligned}
\Rightarrow\quad D(\boldsymbol{\tau}_{CL}) &= \text{diag}(\mathbf{C}_{\widetilde{\boldsymbol{x}}_{CL}\widetilde{\boldsymbol{x}}_{CL}}) \\
&= \text{diag}(\mathbf{C}_{\boldsymbol{x}\boldsymbol{x}}) - \text{diag}(\mathbf{C}_{\widehat{\boldsymbol{x}}_L\widehat{\boldsymbol{x}}_L})\,\mathbf{D} - \mathbf{D}\,\text{diag}(\mathbf{C}_{\widehat{\boldsymbol{x}}_L\widehat{\boldsymbol{x}}_L}) + \mathbf{D}\,\text{diag}(\mathbf{C}_{\widehat{\boldsymbol{x}}_L\widehat{\boldsymbol{x}}_L})\mathbf{D} \\
&= \text{diag}(\mathbf{C}_{\boldsymbol{x}\boldsymbol{x}})\,(\text{diag}(\mathbf{C}_{\widehat{\boldsymbol{x}}_L\widehat{\boldsymbol{x}}_L}))^{-1}\,\text{diag}(\mathbf{C}_{\boldsymbol{x}\boldsymbol{x}}) - \text{diag}(\mathbf{C}_{\boldsymbol{x}\boldsymbol{x}})
\end{aligned}$$
  where we used $\mathbf{D} = \text{diag}(\mathbf{C}_{\boldsymbol{x}\boldsymbol{x}})\,(\text{diag}(\mathbf{C}_{\widehat{\boldsymbol{x}}_L\widehat{\boldsymbol{x}}_L}))^{-1}$.
- Now we want to show $\mathbf{D}\,D(\boldsymbol{\tau}_L) = D(\boldsymbol{\tau}_{CL})$:
$$\begin{aligned}
\mathbf{D}\,D(\boldsymbol{\tau}_L) &= \mathbf{D}\,\text{diag}(\mathbf{C}_{\widetilde{\boldsymbol{x}}_L\widetilde{\boldsymbol{x}}_L}) \\
&= \text{diag}(\mathbf{C}_{\boldsymbol{x}\boldsymbol{x}})\,(\text{diag}(\mathbf{C}_{\widehat{\boldsymbol{x}}_L\widehat{\boldsymbol{x}}_L}))^{-1}\,(\text{diag}(\mathbf{C}_{\boldsymbol{x}\boldsymbol{x}}) - \text{diag}(\mathbf{C}_{\widehat{\boldsymbol{x}}_L\widehat{\boldsymbol{x}}_L})) \\
&= D(\boldsymbol{\tau}_{CL})
\end{aligned}$$
- In the Generalized Linear Model of GAMP or reGVAMP, we have extrinsics
$$\begin{aligned}
\boldsymbol{x} &: \ \widehat{\boldsymbol{x}}_{CL} = \mathbf{r},\ \mathbf{D} = D(\boldsymbol{\tau}_r./\boldsymbol{\tau}_x) \\
\mathbf{z} &: \ \widehat{\mathbf{z}}_{CL} = \boldsymbol{p},\ \mathbf{D} = D(\boldsymbol{\tau}_p./\boldsymbol{\tau}_z)
\end{aligned} \tag{36}$$

Conclusions

- Present an iterative method to calculate the posteriors in a linear mixing model under the condition
  - Independent prior
  - Gaussian noise
- The complexity is $O(N^3)$ per iteration regardless of parallel update or sequential update due to the matrix inverse in LMMSE step.
- We are currently extending this method to the generalized linear model.
- Further research is needed to perform a convergence analysis.

A Few References I

[1] T.P. Minka,
"Expectation Propagation for Approximate Bayesian Inference ,"
in *Proc. Conf. on Uncert. in Art. Intell. (UAI)*, San Francisco, CA, USA, 2001.

[2] T. Heskes, M. Opper, W. Wiegerinck, O. Winther, and O. Zoeter,
"Approximate Inference Techniques with Expectation Constraints,"
J. Stat. Mech: Theory Exp., Nov. 2005.

[3] 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.

[4] D. Zhang, X. Song, W. Wang, G. Fettweis, and X. Gao,
"Unifying Message Passing Algorithms Under the Framework of Constrained Bethe Free Energy Minimization,"
IEEE Trans. Wireless Comm's, Jul. 2021.

[5] S. Rangan, P. Schniter, and A. K. Fletcher,
"Vector Approximate Message Passing,"
IEEE Trans. On Info. Theo., Oct. 2019.

[6] Zilu Zhao, Fangqing Xiao, and Dirk Slock,
"Approximate Message Passing for Not So Large niid Generalized Linear Models,"
in *Proc. Int'l Workshop Signal Processing Advances in Wireless Comm's (SPAWC)*, Sept. 2023.

[7] M. Triki and D. Slock,
"Component-Wise Conditionally Unbiased Bayesian Parameter Estimation: General Concept and Applications to Kalman Filtering and LMMSE Channel Estimation,"
in *IEEE Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, USA, 2005.

[8] M. Triki and D. Slock,
"Investigation of Some Bias and MSE Issues in Block-Component-Wise Conditionally Unbiased LMMSE,"
in *IEEE Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, USA, 2006.

A Few References II

[9] M. Huemer and O. Lang,
"CWCU LMMSE Estimation: Prerequisites and Properties,"
arXiv:1412.1567, Dec. 2014.