FIXED-POINT EQUALIZATION IN DIAGONAL EXPECTATION PROPAGATION: SCALAR DECOUPLING AND BAYES-MMSE OPTIMALITY

Fangqing Xiao, Jichao Chen, Dirk Slock

Communication Systems Department
Eurecom, France
Email: {fangqing.xiao, jichao.chen, dirk.slock}@eurecom.fr

ABSTRACT

We analyze diagonal Expectation Propagation (EP) at fixed points for linear inverse problems with separable priors and right-orthogonally invariant (ROI) sensing. Starting from exact finite-dimensional extrinsic identities, a simple leave-one-out (LOO) concentration establishes equalization of the likelihood-site precisions. A Haar-based central limit theorem then shows that extrinsic residuals are Gaussian and white, yielding an effective scalar AWGN channel. With Bayes denoisers, the prior site equalizes and a single-scalar consistency links the effective noise level to the Bayes MMSE, so that the fixed-point mean-squared error attains the Bayes-optimal value. Our analysis is deliberately simple: it requires asymptotics only for the LOO step and avoids site uniformity and free-probability averaging. We characterize fixed points rather than convergence; algorithmic issues such as damping or projection are left for future work.

Index Terms— Expectation Propagation, scalar decoupling, Bayes-optimal MMSE, leave-one-out analysis, right-orthogonally invariant matrices

1. INTRODUCTION

Expectation Propagation (EP) is a flexible framework for approximate Bayesian inference with strong empirical performance across signal processing and machine learning [1, 2]. Foundational works clarified EP's fixed-point conditions and their relation to expectation-consistent (EC) approximations [3], while extensions such as Power EP and damping explored stability–accuracy tradeoffs [4].

For high-dimensional linear inverse problems, rigorous asymptotics are available for AMP-type algorithms via state evolution, both under i.i.d. designs and, more generally, under right–orthogonally invariant (ROI) matrices through VAMP/OAMP [5–8]. Closer to EP itself, existing analyses rely either on conditioning arguments for Haar matrices [9] or on free-probability tools that provide deterministic equivalents for covariance updates [10].

In this work we analyze *diagonal EP* at fixed points for linear models with separable priors and ROI sensing. Starting from the exact finite-dimensional EP equations, we show how scalar decoupling emerges endogenously at fixed points without assuming uniform sites or a priori averaging. The key ingredients are: (i) exact extrinsic identities combined with leave-one-out concentration, which yield equalization of likelihood-site precisions; and (ii) Haar-measure concentration together with compact SVD arguments, which establish a central limit theorem for extrinsic residuals and thus Gaussianity/whiteness [11, 12]. With Bayes denoisers, prior-site equalization follows and leads to a single-scalar consistency relation whose mean-squared error matches the Bayes–MMSE prediction of the induced scalar channel.

Compared with dynamics-based approaches [9] and free-probability frameworks [10], our contribution is a concise fixed-point theory that (a) avoids Gaussian surrogates, (b) identifies precisely where asymptotics enter (only through concentration), and (c) explains why scalar decoupling is an intrinsic fixed-point property of diagonal EP under ROI. We thus recover the same Bayes–MMSE scalar closure, but via exact EP fixed-point identities plus LOO/Haar concentration, rather than state evolution or free-probability averaging.

2. SYSTEM MODEL AND EXPECTATION PROPAGATION

We consider the linear observation model

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{v}, \tag{1}$$

where $\mathbf{A} \in \mathbb{R}^{m\times n}$ is the sensing matrix, $\mathbf{x} \in \mathbb{R}^n$ is the unknown signal, and $\mathbf{v} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}_m)$ is Gaussian noise. The prior distribution is assumed to be separable, i.e., $p_0(\mathbf{x}) = \prod_{i=1}^{n} p_0(x_i)$.

Notations: For sequences of random variables, $\xrightarrow{p}$ denotes convergence in probability and $\Rightarrow$ denotes convergence in distribution. We write $o(1)$ for deterministic sequences that vanish as $n \to \infty$, and $o_p(1)$ for random quantities that vanish in probability. For a sequence $(a_i)_{i=1}^m$, $\mathrm{diag}(a_i)$ denotes the $m \times m$ diagonal matrix with entries $a_1, \dots, a_m$ on its main diagonal. Here, $\mathrm{diag}(\mathbf{M}_{ii})$ denotes the diagonal matrix with diagonal entries $(M_{11}, \dots, M_{nn})$.

Acknowledgement: EURECOM's research is partially supported by its industrial members: ORANGE, BMW, SAP, iABG, Norton LifeLock, by the Franco-German project 5G-OPERA (BPI), the French project YACARI (PEPR-5G), the EU INFRA project CONVERGE, and by a Huawei France funded Chair towards Future Wireless Networks.

Diagonal expectation propagation (EP) approximates the posterior using two Gaussian site factors [1,2]

$$\tilde{f}_1(\mathbf{x}) \propto \exp\Big(-\tfrac{1}{2}(\mathbf{x}-\mathbf{r}_1)^\top \boldsymbol{\Gamma}_1(\mathbf{x}-\mathbf{r}_1)\Big), \tag{2}$$

$$\tilde{f}_2(\mathbf{x}) \propto \exp\Big(-\tfrac{1}{2}(\mathbf{x}-\mathbf{r}_2)^\top \boldsymbol{\Gamma}_2(\mathbf{x}-\mathbf{r}_2)\Big), \tag{3}$$

where $\mathbf{r}_1, \mathbf{r}_2 \in \mathbb{R}^n$ are mean parameters, and $\boldsymbol{\Gamma}_1 = \operatorname{diag}(\gamma_{1,i})$, $\boldsymbol{\Gamma}_2 = \operatorname{diag}(\gamma_{2,i})$ are diagonal precision matrices. Combining these sites yields the global Gaussian approximation with mean and covariance

$$\mathbf{m} = (\boldsymbol{\Gamma}_1 + \boldsymbol{\Gamma}_2)^{-1}(\boldsymbol{\Gamma}_1\mathbf{r}_1 + \boldsymbol{\Gamma}_2\mathbf{r}_2), \qquad \mathbf{C} = (\boldsymbol{\Gamma}_1 + \boldsymbol{\Gamma}_2)^{-1}. \tag{4}$$

On the likelihood side, given $(\mathbf{r}_1, \boldsymbol{\Gamma}_1)$, the Gaussian posterior is

$$\boldsymbol{\Sigma} = (\boldsymbol{\Gamma}_1 + \sigma^{-2}\mathbf{A}^\top\mathbf{A})^{-1}, \quad \mathbf{x}_2 = \boldsymbol{\Sigma}\big(\boldsymbol{\Gamma}_1\mathbf{r}_1 + \sigma^{-2}\mathbf{A}^\top\mathbf{y}\big). \tag{5}$$

Writing $v_{2,i} = (\boldsymbol{\Sigma})_{ii}$ and $\alpha_{2,i} = (\boldsymbol{\Sigma}\boldsymbol{\Gamma}_1)_{ii}$, the corresponding extrinsic variance is

$$\tau_{1,i} = \frac{v_{2,i}}{(1-\alpha_{2,i})^2}. \tag{6}$$

On the prior side, given $(r_{2,i}, \gamma_{2,i})$, the scalar posterior is

$$p(x_i \mid r_{2,i}, \gamma_{2,i}) \ \propto \ p_0(x_i)\, \exp\Big(-\tfrac{\gamma_{2,i}}{2}(x_i - r_{2,i})^2\Big). \tag{7}$$

Its posterior mean defines the denoiser

$$x_{1,i} = f(r_{2,i}; \gamma_{2,i}) := \mathbb{E}[X_i \mid r_{2,i}, \gamma_{2,i}], \tag{8}$$

with variance $v_{1,i}$ and divergence $\alpha_{1,i}$. The corresponding extrinsic variance is

$$\tau_{2,i} = \frac{v_{1,i}}{(1-\alpha_{1,i})^2}. \tag{9}$$

Definition 1 (EP fixed point)**.** *A tuple* $(\mathbf{r}_1, \mathbf{r}_2, \boldsymbol{\Gamma}_1, \boldsymbol{\Gamma}_2)$ *is said to be a* diagonal-EP fixed point *if, for every coordinate i, the marginal mean and variance obtained from the likelihood-side posterior and the prior-side posterior agree with those of the global Gaussian approximation* (4)*, i.e.,*

$$x_{1,i} = x_{2,i} = m_i, \qquad v_{1,i} = v_{2,i} = C_{ii}, \quad \forall i. \tag{10}$$

Equivalently, the site precisions $(\gamma_{1,i}, \gamma_{2,i})$ *satisfy the consistency relations*

$$\gamma_{2,i} = v_{1,i}^{-1} - \gamma_{1,i}, \qquad \gamma_{1,i} = v_{2,i}^{-1} - \gamma_{2,i}, \quad \forall i. \tag{11}$$

These equations characterize the fixed point of the EP iteration. [2, 3, 13]

Equivalently, the iteration of diagonal EP can be summarized as follows.

Algorithm 1 Diagonal Expectation Propagation (EP)

```
Input: sensing matrix $\mathbf{A}$, observation $\mathbf{y}$, prior $p_0(x_i)$, noise variance $\sigma^2$.
Initialize: $(\mathbf{r}_1, \boldsymbol{\Gamma}_1), (\mathbf{r}_2, \boldsymbol{\Gamma}_2)$ with $\boldsymbol{\Gamma}_1, \boldsymbol{\Gamma}_2 \succ 0$.
repeat
    Likelihood update:
        $\boldsymbol{\Sigma} = (\boldsymbol{\Gamma}_1 + \sigma^{-2}\mathbf{A}^\top\mathbf{A})^{-1}, \quad \mathbf{x}_2 = \boldsymbol{\Sigma}(\boldsymbol{\Gamma}_1\mathbf{r}_1 + \sigma^{-2}\mathbf{A}^\top\mathbf{y}).$   (12)
    Extract $v_{2,i}, \alpha_{2,i}$ and form extrinsic $(r_{1,i}, \tau_{1,i})$.
    Prior update: For each $i$, compute $x_{1,i}, v_{1,i}, \alpha_{1,i}$ via the scalar denoiser and extrinsic variance $\tau_{2,i}$.
    Site update:
        $\gamma_{2,i} \leftarrow v_{1,i}^{-1} - \gamma_{1,i}, \qquad \gamma_{1,i} \leftarrow v_{2,i}^{-1} - \gamma_{2,i}.$   (13)
until convergence of $(\mathbf{r}_1, \mathbf{r}_2, \boldsymbol{\Gamma}_1, \boldsymbol{\Gamma}_2)$.
Output: posterior mean $\mathbf{m}$, covariance $\mathbf{C}$ from (4).
```

For later use, define the auxiliary matrix

$$\mathbf{Q} = (\sigma^2\mathbf{I} + \mathbf{A}\boldsymbol{\Gamma}_1^{-1}\mathbf{A}^\top)^{-1}. \tag{14}$$

Applying the Woodbury identity gives [14]

$$\boldsymbol{\Sigma}\boldsymbol{\Gamma}_1 = \mathbf{I} - \boldsymbol{\Gamma}_1^{-1}\mathbf{A}^\top\mathbf{Q}\mathbf{A}, \qquad \boldsymbol{\Sigma}\sigma^{-2}\mathbf{A}^\top = \boldsymbol{\Gamma}_1^{-1}\mathbf{A}^\top\mathbf{Q}. \tag{15}$$

At any fixed point, setting $\mathbf{D}_2 = \operatorname{diag}((\boldsymbol{\Sigma}\boldsymbol{\Gamma}_1)_{ii})$ yields

$$(\mathbf{I} - \mathbf{D}_2)(\mathbf{r}_1 - \mathbf{x}) = (\boldsymbol{\Sigma}\boldsymbol{\Gamma}_1 - \mathbf{D}_2)(\mathbf{r}_2 - \mathbf{x}) + \boldsymbol{\Sigma}\sigma^{-2}\mathbf{A}^\top\mathbf{v}, \tag{16}$$

an exact finite-dimensional identity.

3. ASSUMPTIONS AND LOO CONCENTRATION

In order to analyze diagonal EP under random sensing, we work with an orthogonal-invariant design model. This setting encompasses the i.i.d. Gaussian ensemble as a special case, while allowing more general spectra through a deterministic limit law. We also assume bounded site precisions on the likelihood side, which ensures stability of the resolvent. These conditions enable concentration tools based on Haar measure and leave-one-out (LOO) resolvents.

Assumption 1 (ROI sensing and bounded site-1)**.** *Let* $m/n \to \delta \in (0,1)$ *and consider the compact SVD [15, 16]*

$$\mathbf{A} = \mathbf{U}\left[\operatorname{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_m}) \quad \mathbf{0}\right]\mathbf{V}^\top, \tag{17}$$

where $\mathbf{U} \in O(m)$ *and* $\mathbf{V} \in O(n)$ *are independent Haar orthogonal matrices and independent of the singular values* $(\lambda_k)_{k=1}^m$*. Assume (i)* spectral regularity*:* $\sup_k \lambda_k \leq \Lambda < \infty$*, the empirical distribution of* (λ_k) *converges, and* $\frac{1}{m}\operatorname{tr}(\mathbf{A}\mathbf{A}^\top) \to 1$*; and (ii)* bounded site-1 precisions*:* $0 < \underline{\gamma} \leq \gamma_{1,i} \leq \overline{\gamma} < \infty$*. Define* $\mathbf{Q} = (\sigma^2\mathbf{I} + \mathbf{A}\boldsymbol{\Gamma}_1^{-1}\mathbf{A}^\top)^{-1}$ *and, for each i, the leave-one-out (LOO) resolvent*

$$\mathbf{Q}_{-i} = \Big(\sigma^2\mathbf{I} + \sum_{j \neq i} \gamma_{1,j}^{-1}\mathbf{a}_j\mathbf{a}_j^\top\Big)^{-1}. \tag{18}$$

The next result shows that the quadratic form $\mathbf{a}_i^\top \mathbf{Q}_{-i}\mathbf{a}_i$ concentrates around the normalized trace of $\mathbf{Q}$. This LOO concentration property is the key step behind the equalization phenomenon of site precisions.

Proposition 1 (LOO quadratic-form concentration). *Let* $u = \frac{1}{m}\mathrm{tr}\,\mathbf{Q}$. *Under Assumption 1, for every* i,

$$\mathbf{a}_i^\top \mathbf{Q}_{-i}\mathbf{a}_i \;=\; u + o_p(1). \tag{19}$$

Proof. Conditioning on $\mathbf{A}_{-i}$, the remaining right singular vector $\boldsymbol{v}_i$ is uniform on the unit sphere. Writing $\mathbf{a}_i = \mathbf{U}\mathrm{diag}(\sqrt{\lambda})\boldsymbol{z}_i$ with $\mathbb{E}[\boldsymbol{z}_i\boldsymbol{z}_i^\top] = \frac{1}{n}\mathbf{I}_m$, one obtains

$$\mathbb{E}[\mathbf{a}_i^\top \mathbf{Q}_{-i}\mathbf{a}_i \mid \mathbf{A}_{-i}] = \tfrac{1}{n}\mathrm{tr}(\mathbf{Q}_{-i}\mathbf{A}\mathbf{A}^\top). \tag{20}$$

By Lévy's lemma [11], the quadratic form $\boldsymbol{z}_i^\top\widetilde{\mathbf{B}}\boldsymbol{z}_i$ (with $\widetilde{\mathbf{B}} = \mathrm{diag}(\sqrt{\lambda})\,\mathbf{U}^\top\mathbf{Q}_{-i}\mathbf{U}\,\mathrm{diag}(\sqrt{\lambda})$) concentrates around its mean, since it is $2\|\widetilde{\mathbf{B}}\|$-Lipschitz and $\|\mathbf{Q}_{-i}\| \le \sigma^{-2}$. Hence

$$\mathbf{a}_i^\top \mathbf{Q}_{-i}\mathbf{a}_i = \tfrac{1}{n}\mathrm{tr}(\mathbf{Q}_{-i}\mathbf{A}\mathbf{A}^\top) + o_p(1). \tag{21}$$

To relate $\mathbf{Q}_{-i}$ to $\mathbf{Q}$, apply Sherman–Morrison–Woodbury:

$$\mathbf{Q} = \mathbf{Q}_{-i} - \frac{\mathbf{Q}_{-i}(\gamma_{1,i}^{-1}\mathbf{a}_i\mathbf{a}_i^\top)\mathbf{Q}_{-i}}{1+\gamma_{1,i}^{-1}\mathbf{a}_i^\top\mathbf{Q}_{-i}\mathbf{a}_i}. \tag{22}$$

Taking traces and using bounded $\gamma_{1,i}$ together with $\|\mathbf{Q}_{-i}\|_F^2 \le m/\sigma^4$ and $\|\mathbf{a}_i\|^2 = O_p(1)$ yields

$$\frac{1}{m}\mathrm{tr}\,\mathbf{Q}_{-i} = \frac{1}{m}\mathrm{tr}\,\mathbf{Q} + o(1). \tag{23}$$

Finally, premultiplying the definition of $\mathbf{Q}_{-i}$ by $(\sigma^2\mathbf{I} + \mathbf{A}\boldsymbol{\Gamma}_1^{-1}\mathbf{A}^\top - \gamma_{1,i}^{-1}\mathbf{a}_i\mathbf{a}_i^\top)$ and taking traces gives

$$\frac{1}{m}\mathrm{tr}\,\mathbf{Q}_{-i} = \frac{1}{\sigma^2}\Big(1-\frac{1}{m}\mathrm{tr}(\mathbf{Q}_{-i}\mathbf{A}\boldsymbol{\Gamma}_1^{-1}\mathbf{A}^\top)+\frac{\gamma_{1,i}^{-1}}{m}\mathbf{a}_i^\top\mathbf{Q}_{-i}\mathbf{a}_i\Big). \tag{24}$$

Combining (20)–(21) with the above identity and Assumption 1 shows that

$$\mathbf{a}_i^\top \mathbf{Q}_{-i}\mathbf{a}_i - \tfrac{1}{m}\mathrm{tr}\,\mathbf{Q} = o_p(1). \tag{25}$$

□

4. LIKELIHOOD-SITE EQUALIZATION

The leave-one-out concentration result in Proposition 1 implies that the quadratic forms $\mathbf{a}_i^\top\mathbf{Q}_{-i}\mathbf{a}_i$ become asymptotically deterministic. We now use this fact to show that the likelihood-site precisions $\gamma_{2,i}$ equalize across all coordinates at any EP fixed point.

Proposition 2 (Equalization of $\boldsymbol{\Gamma}_2$). *Let* $u = \frac{1}{m}\mathrm{tr}\,\mathbf{Q}$. *At any diagonal-EP fixed point,*

$$\gamma_{2,i} \xrightarrow{p} u, \quad \forall i, \qquad \textit{hence} \qquad \boldsymbol{\Gamma}_2 \xrightarrow{p} u\,\mathbf{I}_n. \tag{26}$$

Proof. At a fixed point, the variance identity gives

$$v_{2,i} = \frac{1}{\gamma_{1,i}+\gamma_{2,i}}. \tag{27}$$

On the other hand, by the leave-one-out representation of EP,

$$v_{2,i} = \frac{1}{\gamma_{1,i}+\mathbf{a}_i^\top\mathbf{Q}_{-i}\mathbf{a}_i}. \tag{28}$$

Equating the two expressions yields $\gamma_{2,i} = \mathbf{a}_i^\top\mathbf{Q}_{-i}\mathbf{a}_i$. By Proposition 1, $\mathbf{a}_i^\top\mathbf{Q}_{-i}\mathbf{a}_i = u + o_p(1)$, which establishes $\gamma_{2,i} = u + o_p(1)$ for each i. □

5. EXTRINSIC DECOMPOSITION

Having established that the likelihood-site precisions equalize (Proposition 2), we next analyze the structure of the prior-site residual $\mathbf{r}_1 - \mathbf{x}$. The following lemma shows that it admits a clean decomposition into a "mixing" term depending on $\mathbf{r}_2 - \mathbf{x}$ and a Gaussian noise term.

Lemma 1 (Extrinsic Decomposition). *Let* $\mathbf{D}_2 = \mathrm{diag}((\boldsymbol{\Sigma}\boldsymbol{\Gamma}_1)_{ii})$ *and define* $\alpha_2 = \frac{1}{n}\mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{\Gamma}_1)$. *At any diagonal-EP fixed point, the following exact identity holds:*

$$(\mathbf{I}-\mathbf{D}_2)(\mathbf{r}_1-\mathbf{x}) = (\boldsymbol{\Sigma}\boldsymbol{\Gamma}_1-\mathbf{D}_2)(\mathbf{r}_2-\mathbf{x})+\boldsymbol{\Sigma}\sigma^{-2}\mathbf{A}^\top\mathbf{v}. \tag{29}$$

Moreover, by orthogonal-group concentration, $\|\mathbf{D}_2-\alpha_2\mathbf{I}\|_{\max} \xrightarrow{p} 0$. *Consequently,*

$$\mathbf{r}_1 - \mathbf{x} = \frac{1}{1-\alpha_2}\Big[(\mathbf{I}-\boldsymbol{\Gamma}_1^{-1}\mathbf{A}^\top\mathbf{Q}\mathbf{A}-\alpha_2\mathbf{I})(\mathbf{r}_2-\mathbf{x}) + \boldsymbol{\Gamma}_1^{-1}\mathbf{A}^\top\mathbf{Q}\,\mathbf{v}\Big] + o_p(1), \tag{30}$$

where the Woodbury relations

$$\boldsymbol{\Sigma}\boldsymbol{\Gamma}_1 = \mathbf{I}-\boldsymbol{\Gamma}_1^{-1}\mathbf{A}^\top\mathbf{Q}\mathbf{A}, \qquad \boldsymbol{\Sigma}\sigma^{-2}\mathbf{A}^\top = \boldsymbol{\Gamma}_1^{-1}\mathbf{A}^\top\mathbf{Q} \tag{31}$$

have been used.

Proof. The exact decomposition (29) follows directly from (16). For the concentration step, note that each diagonal element of $\boldsymbol{\Sigma}\boldsymbol{\Gamma}_1$ is a Lipschitz function of the random orthogonal matrix $\mathbf{V}$ with Lipschitz constant $2\|\boldsymbol{\Sigma}\boldsymbol{\Gamma}_1\| = O(1)$ under Assumption 1. By Lévy's lemma, $\mathbf{D}_2$ concentrates around its mean $\alpha_2\mathbf{I}$. Therefore, $(\mathbf{I}-\mathbf{D}_2)^{-1} = (1-\alpha_2)^{-1}\mathbf{I} + o_p(1)$, and substituting the Woodbury identities (15) yields (30). □

6. ANNEALED AWGN

With the equalization of site precisions (Proposition 2) and the extrinsic decomposition (Lemma 1), we are now in position to analyze the distributional behavior of the residuals. The next theorem establishes that the EP extrinsic errors behave as effective additive white Gaussian noise (AWGN) channels.

Theorem 1 (Annealed AWGN). *Under Assumption 1 and Proposition 2, there exist deterministic constants* $\tau_1, \tau_2 > 0$ *such that, for any fixed* k *and distinct indices* $i_1, \ldots, i_k$,

$$\big((\mathbf{r}_1 - \mathbf{x})_{i_1}, \ldots, (\mathbf{r}_1 - \mathbf{x})_{i_k}\big) \Rightarrow \mathcal{N}(\mathbf{0}, \tau_1 \mathbf{I}_k), \tag{32}$$

$$\big((\mathbf{r}_2 - \mathbf{x})_{i_1}, \ldots, (\mathbf{r}_2 - \mathbf{x})_{i_k}\big) \Rightarrow \mathcal{N}(\mathbf{0}, \tau_2 \mathbf{I}_k), \tag{33}$$

with vanishing cross-covariances across distinct indices and annealed asymptotic independence from $\{X_{i_\ell}\}$.

Proof. From lemma 1,

$$\mathbf{r}_1 - \mathbf{x} = \frac{1}{1-\alpha_2}\Big(\mathbf{T}_{\text{mix}} + \mathbf{T}_{\text{noise}}\Big) + o_p(1), \tag{34}$$

where $\mathbf{T}_{\text{mix}} = \mathbf{C}(\mathbf{r}_2 - \mathbf{x})$ with $\mathbf{C} = \mathbf{I} - \mathbf{\Gamma}_1^{-1}\mathbf{A}^\top\mathbf{Q}\mathbf{A} - \alpha_2\mathbf{I}$, and $\mathbf{T}_{\text{noise}} = \mathbf{\Gamma}_1^{-1}\mathbf{A}^\top\mathbf{Q}\,\mathbf{v}$.

The rows of $\mathbf{C}$ have bounded ℓ_2-norm and vanishing entrywise ℓ_∞ norm by spectral regularity and orthogonal-group concentration. A leave-one-out replacement of a single column of $\mathbf{V}$ shows that, up to $o_p(1)$, the coefficients of each row are asymptotically independent of the coordinates of $(\mathbf{r}_2 - \mathbf{x})$. This permits treating $(\mathbf{T}_{\text{mix}})_i$ as a weighted triangular array with weights $\{c_{ij}\}$ satisfying $\sum_j c_{ij}^2 = o_p(1)$ and $\max_j |c_{ij}| = o_p(1)$. By Lindeberg's condition, the central limit theorem applies, and the Cramér–Wold device yields [17]

$$(\mathbf{T}_{\text{mix}})_i \Rightarrow \mathcal{N}(0, v_{\text{m}}), \tag{35}$$

with variance $v_{\text{m}} = (\lim_n \frac{1}{n}\text{tr}(\mathbf{C}\mathbf{C}^\top)) \cdot \tau_2$, where $\tau_2 = \lim_n \frac{1}{n}\sum_j \text{Var}((\mathbf{r}_2 - \mathbf{x})_j)$. For distinct $i \neq \ell$, cross-covariances vanish since rows of $\mathbf{C}$ are asymptotically orthogonal under Haar mixing.

The noise term $\mathbf{T}_{\text{noise}}$ is Gaussian with covariance $\sigma^2(\mathbf{Q}\,\mathbf{A}\mathbf{\Gamma}_1^{-1})^2$. By the concentration of Lipschitz spectral statistics under ROI [15, 16], its normalized trace converges to a deterministic limit v_{g}. Moreover, $\mathbf{T}_{\text{noise}}$ is independent of $\mathbf{T}_{\text{mix}}$ in the LOO construction. Thus $(\mathbf{r}_1 - \mathbf{x})_i$ converges in distribution to $\mathcal{N}(0, \tau_1)$ with

$$\tau_1 = \frac{v_{\text{m}} + v_{\text{g}}}{(1-\alpha_2)^2}, \tag{36}$$

and whiteness follows from the vanishing cross-covariances. Independence from X_i is ensured by the LOO decoupling argument, since the dependence on $\mathbf{x}$ in the i-th coordinate is erased by replacement up to $o_p(1)$. The same reasoning, with the roles of sites 1 and 2 exchanged, yields an analogous Gaussian limit for $\mathbf{r}_2 - \mathbf{x}$ with variance $\tau_2 > 0$. ☐

7. MAIN THEOREM

The annealed AWGN characterization (Theorem 1) enables us to connect diagonal EP to Bayes-optimal estimation. In particular, we can now establish that any fixed point of diagonal EP with a Bayes prior-site denoiser achieves the Bayes MMSE performance.

Theorem 2 (Bayes–MMSE optimality of diagonal-EP fixed points). *Let* $\text{mmse}(\tau)$ *denote the Bayes mean-squared error in the scalar AWGN channel*

$$R = X + \sqrt{\tau}\,Z, \quad Z \sim \mathcal{N}(0,1),\ X \sim p_0, \tag{37}$$

i.e.,

$$\text{mmse}(\tau) = \mathbb{E}\big[(X - \mathbb{E}[X \mid R])^2\big]. \tag{38}$$

Under Assumption 1 with separable prior p_0 *of finite second moment and Gaussian noise, any fixed point of diagonal EP with a Bayes denoiser satisfies*

$$\mathbf{\Gamma}_2 = \gamma_2^\star \mathbf{I}, \tag{39}$$

and the empirical mean-squared error equals $\text{mmse}(\tau_1)$, *i.e.,*

$$\frac{1}{n}\sum_{i=1}^{n}(x_{1,i} - x_i)^2 \xrightarrow{p} \text{mmse}(\tau_1). \tag{40}$$

Proof. By Proposition 2, the likelihood-site precision equalizes, i.e.,

$$\mathbf{\Gamma}_2 \to \gamma_2^\star \mathbf{I}, \qquad \gamma_2^\star = \tfrac{1}{m}\text{tr}\,\mathbf{Q}. \tag{41}$$

By Theorem 1, the extrinsic channel decouples as

$$\mathbf{r}_1 = \mathbf{x} + \sqrt{\tau_1}\mathbf{z}_1, \tag{42}$$

with $\mathbf{z}_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$ (annealed). For the Bayes denoiser matched to p_0, the resulting empirical MSE in the decoupled scalar AWGN channel equals $\text{mmse}(\tau_1)$ [18], i.e.,

$$\frac{1}{n}\sum_{i=1}^{n}(x_{1,i} - x_i)^2 \xrightarrow{p} \text{mmse}(\tau_1). \tag{43}$$

This proves that any fixed point of diagonal EP with a Bayes denoiser attains the Bayes-optimal value. ☐

Theorem 2 shows that diagonal EP fixed points are Bayes-optimal in terms of the empirical MSE whenever the prior-site denoiser is matched to the true separable prior. Importantly, this conclusion does not require the site precisions to be scalar during the iterations: the diagonal-site structure can accommodate coordinate-dependent posterior variances through moment matching, while the global performance is still characterized by a single effective noise level τ_1 via the decoupled AWGN channel. Therefore, the annealed AWGN characterization (Theorem 1) provides a complete and tractable description of the asymptotic performance of diagonal EP at its fixed points.

8. CONCLUSION

We showed that, under ROI sensing and LOO stability, diagonal EP exhibits endogenous equalization at fixed points. This leads to scalar decoupling and Bayes–MMSE performance, established through exact identities, LOO concentration, and a Haar-type CLT. Our analysis characterizes fixed points rather than algorithmic convergence; practical damping and projection strategies, as well as extensions to non-ROI designs and non-Gaussian noise, are left for future work.

9. REFERENCES

[1] T. P. Minka, "Expectation Propagation for Approximate Bayesian Inference," Tech. Rep., Microsoft Research, 2001.

[2] M. Seeger, "Expectation Propagation for Approximate Bayesian Inference," Tech. Rep., EPFL, 2005.

[3] M. Opper and O. Winther, "Expectation Consistent Approximate Inference," *J. Mach. Learn. Res.*, 2005.

[4] T. P. Minka, "Power EP," Tech. Rep., Microsoft Research, 2004.

[5] D. L. Donoho, A. Maleki, and A. Montanari, "Message Passing Algorithms for Compressed Sensing," *Proc. Natl. Acad. Sci. USA*, 2009.

[6] M. Bayati and A. Montanari, "The Dynamics of Message Passing on Dense Graphs, with Applications to Compressed Sensing," *IEEE Trans. Inf. Theory*, 2011.

[7] S. Rangan, P. Schniter, and A. K. Fletcher, "Vector Approximate Message Passing," *IEEE Trans. Inf. Theory*, 2019.

[8] J. Ma and L. Ping, "Orthogonal AMP," *IEEE Access*, 2017.

[9] K. Takeuchi, "Rigorous Dynamics of Expectation-Propagation-Based Signal Recovery from Unitarily Invariant Measurements," *arXiv:1701.05284*, 2017.

[10] B. Çakmak and M. Opper, "Expectation Propagation for Approximate Inference: Free Probability Framework," *arXiv:1801.05411*, 2018.

[11] M. Ledoux, *The Concentration of Measure Phenomenon*, vol. 89 of *Math. Surveys and Monographs*, Amer. Math. Soc., 2001.

[12] R. Vershynin, *High-Dimensional Probability: An Introduction with Applications in Data Science*, Cambridge Univ. Press, 2018.

[13] F. Xiao and D. Slock, "Parameter Estimation via Expectation Maximization–Expectation Consistent Algorithm," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, 2024, pp. 9506–9510.

[14] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins Univ. Press, 2013.

[15] R. Couillet and M. Debbah, *Random Matrix Methods for Wireless Communications*, Cambridge Univ. Press, 2011.

[16] A. M. Tulino and S. Verdú, *Random Matrix Theory and Wireless Communications*, Found. Trends Commun. Inf. Theory, 2004.

[17] P. Billingsley, *Probability and Measure*, Wiley, 3rd edition, 1995.

[18] D. Guo, S. Shamai, and S. Verdú, "Mutual Information and Minimum Mean-Square Error in Gaussian Channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, 2005.