LARGE-SYSTEM FIXED-POINT LAW AND DETERMINISTIC CLOSURE FOR SPARSE BAYESIAN LEARNING

Fangqing Xiao, Dirk Slock

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

ABSTRACT

Sparse Bayesian Learning is widely used for high-dimensional inverse problems, yet fixed-point behavior in underdetermined systems is poorly understood. We give a distributional account of the hyperparameters at convergence under independent Gaussian designs with deterministic heterogeneous signals. Each coordinate follows a spike-and-tail law with a translated noncentral chi-square tail. A quenched self-averaging principle turns empirical averages into deterministic quantities and yields a compact closure based on three resolvent-derived scalars, predicting activation, error, and detection without a signal prior. With learned noise variance, a simple balance fixes the effective activation threshold at one. Simulations validate the coordinate law and the closure.

***Index Terms*—** Sparse Bayesian Learning, deterministic equivalents, self-averaging, large-system limit

1. INTRODUCTION

Sparse Bayesian Learning, also known as Automatic Relevance Determination, is widely used for high-dimensional linear inverse problems and is valued for automatic support pruning and strong empirical performance [1, 2, 3]. It is typically implemented via an expectation-maximization routine or Type-II evidence maximization. Yet the behavior of individual coordinates at fixed points in underdetermined systems remains unclear: practitioners observe large fluctuations across coordinates without a distributional explanation or a concise link to macroscopic performance. Our focus is precisely on fixed points, so the results apply to both implementation families.

Prior work on approximate message passing offers sharp predictions for iterative dynamics [4, 5, 6], but does not target SBL fixed points and usually assumes random signals. Classical SBL and ARD papers develop evidence maximization, efficient updates, and hierarchical models [1, 2, 3, 7, 8, 9, 10], while related analyses via Bethe free energy study convergence of approximate variants [11]. None provides a coordinate-wise distribution at fixed points together with a variance-mismatch-aware macroscopic closure. This paper delivers three elements under independent Gaussian designs with deterministic heterogeneous signals and Gaussian noise: (i) a coordinate-wise limiting law with a spike at zero and a translated, scaled noncentral chi-square tail; (ii) a quenched self-averaging principle that yields a compact closure based on three resolvent-derived summary quantities predicting activation, error, and detection without a signal prior; and (iii) a clean separation between algorithmic and data-generating noise, including a simple balance under Type-II noise learning that fixes the effective activation threshold at one.

2. SYSTEM MODEL AND FIXED-POINT SETUP

We consider sparse linear estimation with an i.i.d. Gaussian design in the underdetermined, high-dimensional regime. The observation model is

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

where $A_{ij} \sim \mathcal{N}(0, 1/m)$, $\mathbf{v}_g \sim \mathcal{N}(\mathbf{0}, \sigma_g^2 \mathbf{I}_m)$, and the signal $\mathbf{x} \in \mathbb{R}^n$ is deterministic (heterogeneous coordinates allowed). We work throughout with $m, n \to \infty$ and $m/n \to \delta \in (0, 1)$, following the SBL/ARD setup [1, 2, 3].

A key modeling choice is to separate the *algorithmic* variance σ^2 used inside EM from the *true* variance σ_g^2 that generates the data (matched and mismatched cases are both allowed). Let $\mathbf{\Gamma} = \mathrm{diag}(\gamma_1, \ldots, \gamma_n)$ and define the resolvent

$$\mathbf{Q} \triangleq \left(\sigma^2 \mathbf{I}_m + \mathbf{A}\mathbf{\Gamma}\mathbf{A}^\top\right)^{-1}. \tag{2}$$

At any EM *fixed point*, the posterior mean and covariance of $\mathbf{x}$ and the hyperparameters obey

$$\boldsymbol{\mu} = \mathbf{\Gamma}\mathbf{A}^\top\mathbf{Q}\mathbf{y}, \qquad \mathbf{\Sigma} = \mathbf{\Gamma} - \mathbf{\Gamma}\mathbf{A}^\top\mathbf{Q}\mathbf{A}\mathbf{\Gamma}, \tag{3a}$$

$$\gamma_i = \mu_i^2 + [\mathbf{\Sigma}]_{ii}, \quad i = 1, \ldots, n. \tag{3b}$$

If the variance is learned (Type-II MLE), we use the standard evidence update

$$\sigma^2 = \frac{1}{m}\Big(\|\mathbf{y} - \mathbf{A}\boldsymbol{\mu}\|_2^2 + \mathrm{tr}(\mathbf{A}\mathbf{\Sigma}\mathbf{A}^\top)\Big). \tag{4}$$

We write $a \triangleq b$ for "a is defined as b"; $X \overset{d}{=} Y$ for equality in distribution; and $X_n \Rightarrow X$ for weak convergence. $O_{\mathbb{P}}(\cdot)$ denotes stochastic order. The positive part is $(t)_+ = \max\{t, 0\}$. $\mathrm{tr}(\cdot)$ denotes the trace and $\mathrm{diag}(\cdot)$ forms a diagonal matrix. $\mathbb{E}[\cdot]$, $\mathrm{Var}(\cdot)$, and $\Pr(\cdot)$ denote expectation, variance, and probability.

Our statements and proofs concern *fixed points* of (2)-(3) (and (4) when used), not the transient iterates; identities follow from standard matrix inversion lemmas [12].

2.1. Three global scalars and a roadmap

We repeatedly use three resolvent-derived scalars:

$$\eta = \tfrac{1}{m}\mathrm{tr}(\mathbf{Q}), \qquad \tau = \tfrac{1}{m}\mathrm{tr}(\mathbf{Q}^2), \qquad B = \tfrac{1}{m}\,\mathbf{y}^\top \mathbf{Q}^2 \mathbf{y}. \tag{5}$$

They play complementary roles: η/B acts as the effective activation threshold in the coordinate-wise law; (η, τ) summarize first- and second-order resolvent information used by our closure; and B carries the data-generating noise power through σ_g^2. In Sec. 3 we characterize the coordinate-wise limiting distribution of the γ_i at fixed points. Sec. 4 then replaces sample averages by expectations under this law (self-averaging in the regime $\delta \in (0,1)$) to obtain a low-dimensional deterministic closure for (η, τ, B), and to quantify how variance mismatch shifts the threshold via η/B.

2.2. Asymptotic regime and scope

All results are stated in the high-dimensional limit $m, n \to \infty$ with $m/n \to \delta \in (0,1)$. Questions of global uniqueness or stability of the coupled system, as well as universality beyond Gaussian designs, are left to an extended version.

3. COORDINATE-WISE LIMIT LAW AND SELF-AVERAGING

We develop a distributional description of SBL fixed points at the level of individual coordinates. In the underdetermined regime, we (i) isolate an exact one-coordinate identity, (ii) show a local Gaussian limit for the leave-one-out statistics that feed it, and (iii) obtain the coordinate-wise limiting law for the hyperparameters. Concentration inputs follow standard high-dimensional probability [13, 14]; weak convergence uses the continuous mapping theorem and Slutsky's lemma [15]; self-averaging relies on bounded-difference/ Efron-Stein arguments [16, 17]. Noncentral chi-square facts are classical [18].

3.1. An exact single-coordinate identity

Let $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_n]$. For the ith coordinate, define the leave-one-out resolvent and statistics

$$\mathbf{Q}_{-i} = \Big(\sigma^2 \mathbf{I}_m + \sum_{j \neq i} \gamma_j\, \mathbf{a}_j \mathbf{a}_j^\top\Big)^{-1}, \tag{6a}$$

$$s_i = \mathbf{a}_i^\top \mathbf{Q}_{-i} \mathbf{a}_i, \qquad r_i = \mathbf{a}_i^\top \mathbf{Q}_{-i} \mathbf{y}. \tag{6b}$$

Lemma 1 (Exact fixed-point identity)**.** *At any EM fixed point,*

$$\gamma_i = \Big(\frac{r_i^2 - s_i}{s_i^2}\Big)_+, \quad \textit{equivalently} \quad \gamma_i s_i^2 + s_i - r_i^2 = 0,\ \gamma_i \geq 0. \tag{7}$$

Proof sketch. By the matrix inversion lemma, $\mathbf{a}_i^\top \mathbf{Q} \mathbf{a}_i = s_i/(1+\gamma_i s_i)$ and $\mathbf{a}_i^\top \mathbf{Q} \mathbf{y} = r_i/(1+\gamma_i s_i)$. Using (3a)-(3b), $\mu_i = \gamma_i r_i/(1+\gamma_i s_i)$ and $[\mathbf{\Sigma}]_{ii} = \gamma_i/(1+\gamma_i s_i)$. Substituting into $\gamma_i = \mu_i^2 + [\mathbf{\Sigma}]_{ii}$ and clearing denominators yields the claim. □

3.2. Local Gaussian limit for leave-one-out statistics

Write $\mathbf{y} = \mathbf{a}_i x_i + \mathbf{y}_{-i}$ with $\mathbf{y}_{-i} = \mathbf{A}_{-i}\mathbf{x}_{-i} + \mathbf{v}_g$.

Lemma 2 (Concentration and local CLT)**.** *In the regime $\delta \in (0,1)$,*

$$s_i = \eta + O_{\mathbb{P}}\Big(\sqrt{\tfrac{\log n}{m}}\Big), \qquad r_i = \eta x_i + \zeta_i + O_{\mathbb{P}}\Big(\sqrt{\tfrac{\log n}{m}}\Big), \tag{8}$$

where, conditional on $(\mathbf{Q}_{-i}, \mathbf{y}_{-i})$,

$$\zeta_i \sim \mathcal{N}\Big(0,\ \tfrac{1}{m}\|\mathbf{Q}_{-i}\mathbf{y}_{-i}\|_2^2\Big), \tag{9a}$$

$$\frac{1}{m}\|\mathbf{Q}_{-i}\mathbf{y}_{-i}\|_2^2 = B + O_{\mathbb{P}}\Big(\sqrt{\tfrac{\log n}{m}}\Big). \tag{9b}$$

Hence $(r_i, s_i) \Rightarrow (\eta x_i + \zeta, \eta)$ with $\zeta \sim \mathcal{N}(0, B)$.

Proof sketch. Hanson-Wright and rank-one resolvent stability imply $s_i = \eta + O_{\mathbb{P}}(\sqrt{\log n/m})$ [13, 14]. Decompose $r_i = x_i s_i + \mathbf{a}_i^\top \mathbf{Q}_{-i}\mathbf{y}_{-i}$; the second term is conditionally Gaussian with variance $m^{-1}\|\mathbf{Q}_{-i}\mathbf{y}_{-i}\|_2^2$, which matches $B = m^{-1}\mathbf{y}^\top\mathbf{Q}^2\mathbf{y}$ up to $O_{\mathbb{P}}(\sqrt{\log n/m})$ by matrix Bernstein and a rank-one comparison. Slutsky's lemma gives the joint limit [15]. □

3.3. Limiting law of the hyperparameters

To account for coordinate heterogeneity, associate to each i an auxiliary variable

$$S^{(i)} \mid x_i \ \sim\ \chi_1^2\Big(\frac{\eta^2 x_i^2}{B}\Big), \tag{10}$$

and define $h(r,s) = ((r^2 - s)_+)/s^2$ for $s > 0$.

Theorem 1 (Coordinate-wise limiting distribution)**.** *With probability tending to one over $(\mathbf{A}, \mathbf{v}_g)$, at any EM fixed point,*

$$\gamma_i \ \overset{d}{=}\ \frac{B}{\eta^2}\Big(S^{(i)} - \frac{\eta}{B}\Big)_+, \qquad S^{(i)} \mid x_i \sim \chi_1^2\Big(\frac{\eta^2 x_i^2}{B}\Big). \tag{11}$$

Moreover,

$$\gamma_i = \frac{B}{\eta^2}\Big(S^{(i)} - \frac{\eta}{B}\Big)_+ \ +\ O_{\mathbb{P}}\Big(\sqrt{\tfrac{\log n}{m}}\Big). \tag{12}$$

Proof sketch. By Lemma 1, $\gamma_i = h(r_i, s_i)$. Lemma 2 gives $(r_i, s_i) \Rightarrow (\eta x_i + \zeta, \eta)$ with $\zeta \sim \mathcal{N}(0, B)$. Continuity of h on $\{s > 0\}$ and the 1-Lipschitz property of the positive part yield (11) via the continuous mapping theorem; a local Lipschitz bound plus Lemma 2 gives (12). □

The law combines a point mass at zero with a translated-scaled noncentral χ^2 tail. The activation threshold is η/B, while the noncentrality $\eta^2 x_i^2/B$ sets how frequently and how strongly coordinate i turns active; both are macroscopic and will be identified by the closure in Sec. 4.

3.4. Quenched self-averaging of sample means

Let $\psi : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ be pseudo-Lipschitz of order 2 and define

$$\mathcal{E}_i[\psi] \triangleq \mathbb{E}\big[\psi(\Gamma_i, x_i) \,\big|\, x_i\big], \qquad \Gamma_i \overset{d}{=} \frac{B}{\eta^2}\Big(S^{(i)} - \frac{\eta}{B}\Big)_+. \tag{13}$$

Proposition 1 (Self-averaging)**.** *For*

$$\begin{aligned} &\psi_1(\gamma, x) = \tfrac{\gamma}{1+\eta\gamma}, \quad \psi_2(\gamma, x) = \Big(\tfrac{\gamma}{1+\eta\gamma}\Big)^2, \\ &\psi_w(\gamma, x) = \tfrac{x^2}{(1+\eta\gamma)^2}, \end{aligned} \tag{14}$$

one has

$$\frac{1}{n}\sum_{i=1}^{n} \psi(\gamma_i, x_i) = \frac{1}{n}\sum_{i=1}^{n} \mathcal{E}_i[\psi] + O_{\mathbb{P}}\Big(\sqrt{\tfrac{\log n}{m}}\Big) + O_{\mathbb{P}}(n^{-1/2}). \tag{15}$$

Proof sketch. View $n^{-1}\sum_i \psi(\gamma_i, x_i)$ as a function of the columns of $\mathbf{A}$. Replacing one column by an i.i.d. copy changes the average by $O(1/n)$ (rank-one resolvent stability and pseudo-Lipschitzness), so Efron-Stein/McDiarmid gives $O_{\mathbb{P}}(n^{-1/2})$ fluctuations. Replacing γ_i by its limit in Theorem 1 contributes $O_{\mathbb{P}}(\sqrt{\log n/m})$ via Lemma 2; conditioning on x_i and integrating over $S^{(i)}$ yields the right-hand side. □

(11)-(15) are the only ingredients needed for the closure: the ratio η/B and the coordinate-wise noncentralities $\{\eta^2 x_i^2/B\}$ enter through $S^{(i)}$, and self-averaging replaces empirical sums by conditional expectations. Section 4 turns these into a deterministic, low-dimensional system for (η, τ, B).

4. DETERMINISTIC CLOSURE IN THE UNDERDETERMINED REGIME

Using the coordinate-wise law from Sec. 3, the random EM fixed point collapses to three deterministic scalars (η, τ, B). The key is to replace empirical averages by conditional expectations under $S^{(i)} \mid x_i \sim \chi_1^2(\eta^2 x_i^2/B)$ from (10). We record the induced functionals, state the closure, and note how variance mismatch shifts the threshold η/B. Throughout we work with $\delta \in (0, 1)$.

4.1. Distribution-induced functionals

Let $r \triangleq B/\eta$ and $u \triangleq \eta^2/B$. For $s > 0$, define

$$\begin{aligned} &\phi_1(s; r) = \frac{(s - 1/r)_+}{1 + r(s - 1/r)_+}, \phi_2(s; r) = \frac{(s - 1/r)_+^2}{(1 + r(s - 1/r)_+)^2}, \\ &\phi_w(s; r; x) = \frac{x^2}{(1 + r(s - 1/r)_+)^2}. \end{aligned} \tag{16}$$

For each coordinate i, draw $S^{(i)} \mid x_i \sim \chi_1^2(u x_i^2)$ [18] and set

$$\begin{aligned} S_1(\eta, B) &= \frac{r}{\eta} \cdot \frac{1}{n}\textstyle\sum_{i=1}^n \mathbb{E}\big[\phi_1(S^{(i)}; r) \mid x_i\big], \\ S_2(\eta, B) &= \frac{r^2}{\eta^2} \cdot \frac{1}{n}\textstyle\sum_{i=1}^n \mathbb{E}\big[\phi_2(S^{(i)}; r) \mid x_i\big], \\ \overline{X}_w(\eta, B) &= \frac{1}{n}\textstyle\sum_{i=1}^n \mathbb{E}\big[\phi_w(S^{(i)}; r; x_i) \mid x_i\big]. \end{aligned} \tag{17}$$

Proposition 2 (Sample-mean replacement)**.** *At any EM fixed point with $\delta \in (0, 1)$, with probability $\to 1$,*

$$\begin{aligned} &\frac{1}{n}\textstyle\sum_{i=1}^n \frac{\gamma_i}{1 + \eta\gamma_i} = S_1(\eta, B) + O_{\mathbb{P}}\Big(\sqrt{\tfrac{\log n}{m}}\Big) + O_{\mathbb{P}}(n^{-1/2}), \\ &\frac{1}{n}\textstyle\sum_{i=1}^n \Big(\frac{\gamma_i}{1 + \eta\gamma_i}\Big)^2 = S_2(\eta, B) + O_{\mathbb{P}}\Big(\sqrt{\tfrac{\log n}{m}}\Big) + O_{\mathbb{P}}(n^{-1/2}), \\ &\frac{1}{n}\textstyle\sum_{i=1}^n \frac{x_i^2}{(1 + \eta\gamma_i)^2} = \overline{X}_w(\eta, B) + O_{\mathbb{P}}\Big(\sqrt{\tfrac{\log n}{m}}\Big) + O_{\mathbb{P}}(n^{-1/2}). \end{aligned} \tag{18}$$

Proof sketch. Apply Proposition 1 with ψ_1, ψ_2, ψ_w in (14) and substitute $\Gamma_i \overset{d}{=} \frac{r}{\eta}(S^{(i)} - 1/r)_+$ from Theorem 1, conditioning on x_i to get (17). The error terms come from Proposition 1. □

Each inner expectation in (17) is a one-dimensional integral under a noncentral χ_1^2 whose noncentrality depends on x_i; the outer average is merely across coordinates (no prior on $\{x_i\}$ is assumed).

4.2. Closure and the cross term

We now state the deterministic relations satisfied by (η, τ, B); the only subtlety is the mixed signal-noise term in B, which vanishes at rate $m^{-1/2}$.

Theorem 2 (Deterministic closure)**.** *With probability $\to 1$ in the underdetermined regime, any EM fixed point satisfies*

$$\begin{aligned} &\eta^{-1} = \sigma^2 + \frac{1}{\delta} S_1(\eta, B), \quad \tau = \frac{\eta^2}{1 - \dfrac{\eta^2}{\delta} S_2(\eta, B)}, \\ &B = \tau\Big(\sigma_g^2 + \frac{1}{\delta}\,\overline{X}_w(\eta, B)\Big). \end{aligned} \tag{19}$$

In the matched case $\sigma^2 = \sigma_g^2$, the same variance parameter appears in all three equations.

Proof sketch. From $(\sigma^2 \mathbf{I} + \mathbf{A}\boldsymbol{\Gamma}\mathbf{A}^\top)\mathbf{Q} = \mathbf{I}$ and its $\mathbf{Q}$-right-multiplied form, use $\mathbf{a}_i^\top \mathbf{Q} \mathbf{a}_i = \frac{s_i}{1+\gamma_i s_i}$ and $\mathbf{a}_i^\top \mathbf{Q}^2 \mathbf{a}_i = \frac{s_i^2}{(1+\gamma_i s_i)^2}$, then replace empirical sums by Proposition 2 to obtain the first two relations. For $B = \frac{1}{m}\mathbf{y}^\top \mathbf{Q}^2 \mathbf{y}$, expand $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{v}_g$; the cross term is mean zero with variance $O(m^{-1})$ (Lemma 3), while the signal and noise terms yield $\tau\,\delta^{-1}\overline{X}_w$ and $\sigma_g^2 \tau$, respectively. □

Lemma 3 (Vanishing cross term)**.** *Let $C_m \triangleq \frac{2}{m}\mathbf{x}^\top \mathbf{A}^\top \mathbf{Q}^2 \mathbf{v}_g$ with $\mathbf{Q} = (\sigma^2 \mathbf{I}_m + \mathbf{A}\boldsymbol{\Gamma}\mathbf{A}^\top)^{-1}$. If $\sigma^2 \geq \sigma_{\min} > 0$ and $n^{-1}\sum_i x_i^2 \leq \bar{s} < \infty$, then $\mathbb{E}[C_m \mid \mathbf{A}, \mathbf{x}] = 0$ and $\mathrm{Var}(C_m \mid \mathbf{A}, \mathbf{x}) = O_{\mathbb{P}}(1/m)$; hence $C_m = O_{\mathbb{P}}(m^{-1/2}) \to 0$.*

Proof sketch. Condition on $(\mathbf{A}, \mathbf{x})$. Since $\mathbf{v}_g \sim \mathcal{N}(\mathbf{0}, \sigma_g^2 \mathbf{I})$, C_m is Gaussian with variance:

$$\tfrac{4\sigma_g^2}{m^2}\mathbf{x}^\top \mathbf{A}^\top \mathbf{Q}^4 \mathbf{A}\mathbf{x} \le \tfrac{4\sigma_g^2}{m^2}\|\mathbf{A}\|^2\|\mathbf{Q}\|^4\|\mathbf{x}\|^2 = O_{\mathbb{P}}(1/m),$$

using $\|\mathbf{A}\| = O_{\mathbb{P}}(1)$, $\|\mathbf{Q}\| \le \sigma_{\min}^{-2}$, and $\|\mathbf{x}\|^2 \le m\bar{s}/\delta$. □

The functionals $S_1, S_2, \overline{X}_w$ in (17) reduce to one-dimensional expectations under noncentral χ_1^2 and can be computed by standard quadrature, Marcum-Q representations, or the Poisson-mixture with incomplete gamma functions [18].

4.3. Practical evaluation with noise learning

At a fixed point of the evidence update (4), using $\mathbf{y} - \mathbf{A}\boldsymbol{\mu} = \sigma^2 \mathbf{Q}\mathbf{y}$ and $\mathrm{tr}(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top) = \sigma^2 m - \sigma^4 \mathrm{tr}(\mathbf{Q})$, (4) simplifies to $\sigma^2 = \sigma^2 + \sigma^4(\mathbf{y}^\top \mathbf{Q}^2 \mathbf{y} - \mathrm{tr}(\mathbf{Q}))/m$, so $\eta = B$ and $r = B/\eta = 1$ at convergence. Then

$$\phi_1(s;1) = \frac{(s-1)_+}{1+(s-1)_+}, \quad \phi_2(s;1) = \frac{(s-1)_+^2}{(1+(s-1)_+)^2}, \tag{20}$$

$$\phi_w(s;1;x) = \frac{x^2}{(1+(s-1)_+)^2}, \tag{21}$$

with $S^{(i)} \mid x_i \sim \chi_1^2(\eta x_i^2)$. Define

$$\begin{aligned} \Psi_1(\eta) &= \tfrac{1}{n}\textstyle\sum_{i=1}^n \mathbb{E}\big[\phi_1(S^{(i)};1) \mid x_i\big], \\ \Psi_2(\eta) &= \tfrac{1}{n}\textstyle\sum_{i=1}^n \mathbb{E}\big[\phi_2(S^{(i)};1) \mid x_i\big]. \end{aligned} \tag{22}$$

The closure becomes one-dimensional:

$$\eta^{-1} = \sigma^2 + \frac{1}{\delta}\cdot\frac{1}{\eta}\Psi_1(\eta), \qquad \tau = \frac{\eta^2}{1 - \frac{1}{\delta}\Psi_2(\eta)}, \tag{23}$$
$$\eta = \tau\Big(\sigma_g^2 + \frac{1}{\delta}\overline{X}_w(\eta,\eta)\Big),$$

where $\overline{X}_w(\eta,\eta) = \frac{1}{n}\sum_i \mathbb{E}[\phi_w(S^{(i)};1;x_i) \mid x_i]$. One may solve (23) for η alone and then recover $\sigma^2 = \eta^{-1} - \delta^{-1}\eta^{-1}\Psi_1(\eta)$ from the first equation.

5. NUMERICAL SIMULATION

We validate two claims without running SBL: (i) the coordinate-wise law, and (ii) the scalar closure. We use $A_{ij} \sim \mathcal{N}(0, 1/m)$, $n = 200$, $x_i = 0.95^{\,i-1}$, and $\sigma^2 = 0.05$.

(i) Coordinate-wise law. Given (η, x_i) and a view of B (for learned noise we set $B = \eta$), draw $S_i \sim \chi_1^2(\lambda_i)$ with $\lambda_i = \eta^2 x_i^2/B$ and map $\Gamma_i = (S_i - 1)_+/\eta$. Fig. 1 overlays the histogram of Γ_i on $\{\Gamma_i > 0\}$ with the conditional density $\eta f_{\chi_1^2(\lambda)}(1+\eta y)/(1 - F_{\chi_1^2(\lambda)}(1))$, and reports the spike mass at zero $F_{\chi_1^2(\lambda)}(1)$ (empirical vs. theory). The QQ plot of $\Gamma_i \mid \Gamma_i > 0$ is near-linear across the bulk and upper tail. This jointly validates the spike-at-zero and the translated noncentral chi-square tail, showing that per-coordinate fluctuations at fixed points are fully governed by macroscopic quantities rather than idiosyncratic coordinates. **(ii) Scalar closure.** In the fixed-noise case we solve

$$\eta^{-1} = \sigma^2 + \delta^{-1} S_1(\eta), \quad S_1(\eta) = \frac{1}{n}\sum_{i=1}^n \mathbb{E}\Big[\frac{\Gamma_i}{1+\eta\Gamma_i}\Big],$$

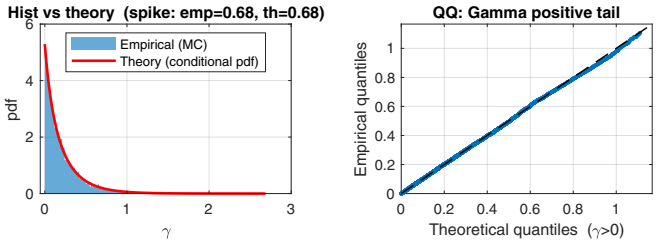

Fig. 1. **Per-coordinate law.** Left: histogram on $\{\Gamma_i > 0\}$ with the conditional density; the title reports $\Pr(\Gamma_i = 0)$ (empirical vs. theory). Right: QQ-plot of $\Gamma_i \mid \Gamma_i > 0$ shows quantile-level agreement, including the upper tail.

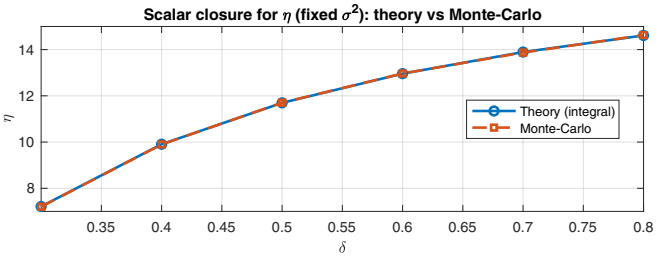

Fig. 2. **Scalar closure for η.** Theory: solution to $\eta^{-1} = \sigma^2 + \delta^{-1}S_1(\eta)$ using the closed-form S_1; MC: replacing S_1 by a Monte-Carlo estimate. Agreement across δ supports self-averaging and the deterministic closure.

with $S_i \sim \chi_1^2(\eta x_i^2)$ and $\Gamma_i = (S_i - 1)_+/\eta$. We evaluate S_1 via the Poisson-mixture for the noncentral chi-square and incomplete gamma functions, and find η by `fzero`; as a check we replace S_1 by a Monte-Carlo estimate (4k samples/coordinate). Fig. 2 shows close overlap between the deterministic curve and MC markers across δ.

6. CONCLUSION

We studied fixed points of Sparse Bayesian Learning under independent Gaussian designs and gave a coordinate-wise distribution for the learned hyperparameters. A quenched self-averaging principle led to a compact closure that predicts activation, error, and detection without a signal prior. When the noise variance is learned, the effective activation threshold is fixed at one. Two small simulations, run without the learning loop, confirmed both the law and the closure. Future work includes universality beyond Gaussian designs and finite-sample refinements.

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 H2030 project CONVERGE, and by a Huawei France funded Chair towards Future Wireless Networks.

7. REFERENCES

[1] M. E. Tipping, "Sparse bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, 2001.

[2] D. P. Wipf and B. D. Rao, "Sparse bayesian learning for basis selection," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2153–2164, 2004.

[3] M. E. Tipping and A. C. Faul, "Fast marginal likelihood maximisation for sparse bayesian models," in *Proc. AISTATS*, 2003.

[4] M. Bayati and A. Montanari, "The dynamics of message passing on dense graphs, with applications to compressed sensing," *IEEE Trans. Inf. Theory*, vol. 57, no. 2, pp. 764–785, 2011.

[5] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2011, pp. 2168–2172.

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

[7] D. P. Wipf and S. Nagarajan, "A new view of automatic relevance determination," *NeuroImage*, vol. 37, no. 4, pp. 1102–1120, 2008.

[8] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2346–2356, 2008.

[9] S. D. Babacan, R. Molina, and A. K. Katsaggelos, "Bayesian compressive sensing using laplace priors," *IEEE Trans. Image Process.*, vol. 19, no. 1, pp. 53–63, 2010.

[10] F. Xiao and D. Slock, "Towards Hyperparameter Optimizing of Sparse Bayesian Learning Based on Stein's Unbiased Risk Estimator," in *Proc. IEEE ISIT-W*, 2024, pp. 1–5.

[11] C. K. Thomas and D. T. M. Slock, "Convergence analysis of sparse bayesian learning under approximate inference techniques," in *Proc. Asilomar Conf. Signals, Systems, and Computers*, Pacific Grove, CA, USA, 2019.

[12] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins Univ. Press, Baltimore, MD, USA, 4th edition, 2013.

[13] R. Vershynin, *High-Dimensional Probability: An Introduction with Applications in Data Science*, Cambridge Ser. Stat. Probab. Math. Cambridge Univ. Press, Cambridge, UK, 2018.

[14] J. A. Tropp, "An introduction to matrix concentration inequalities," *Found. Trends Mach. Learn.*, vol. 8, no. 1-2, pp. 1–230, 2015.

[15] P. Billingsley, *Convergence of Probability Measures*, Wiley, New York, NY, USA, 2nd edition, 1999.

[16] S. Boucheron, G. Lugosi, and P. Massart, *Concentration Inequalities: A Nonasymptotic Theory of Independence*, Oxford Univ. Press, Oxford, UK, 2013.

[17] C. McDiarmid, "On the method of bounded differences," in *Surveys in Combinatorics*, pp. 148–188. Cambridge Univ. Press, Cambridge, UK, 1989.

[18] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions, Vol. 1*, Wiley, New York, NY, USA, 2nd edition, 1994.