LARGE SYSTEM ANALYSIS OF SURE BASED HYPER- PARAMETER OPTIMIZING IN SPARSE BAYESIAN LEARNING

Fangqing Xiao, Jichao Chen, Dirk Slock

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

ABSTRACT

This paper studies stationary points arising from output-domain Stein's unbiased risk estimator (SURE) based hyper-parameter optimization in sparse Bayesian learning from a large-system perspective. By analyzing the coordinatewise stationary conditions and leveraging tools from random matrix theory, we show that the high-dimensional problem admits a deterministic decoupling, leading to a distributional fixed-point characterization of stationary solutions. This framework reveals that multiple stationary points may naturally coexist. We further provide a local stability interpretation, clarifying which stationary solutions are observable under coordinatewise optimization. The analysis focuses on stationary-point structure rather than specific algorithms or finite-dimensional performance, and provides a principled theoretical understanding of the stationary landscape of SURE-based hyper-parameter learning in sparse Bayesian learning.

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

1. INTRODUCTION

Sparse Bayesian learning (SBL) is a widely used framework for sparse signal recovery and linear inverse problems, owing to its flexibility in modeling uncertainty and its strong empirical performance. In its canonical formulation, SBL assigns independent Gaussian priors to signal coefficients with unknown variances and estimates these hyper-parameters from the data, typically via type-II maximum likelihood or evidence maximization [1, 2]. Despite its long history and extensive empirical success, the theoretical properties of hyper-parameter learning in SBL, particularly in high-dimensional settings, remain only partially understood.

A central challenge in SBL lies in the selection and optimization of the hyper-parameters that control sparsity. Stein's unbiased risk estimate (SURE) provides a principled, data-driven criterion for tuning parameters in linear estimators under Gaussian noise [3, 4]. SURE-based approaches have been successfully applied to regularization selection in a variety of inverse problems and sparse recovery settings [5, 6]. Motivated by these advantages, several recent works have explored the use of SURE for hyper-parameter optimization within the SBL framework, aiming to bypass explicit assumptions on the signal prior and noise statistics [7, 8].

However, incorporating SURE into SBL leads to a highly nonconvex objective function over a large number of hyper-parameters. In practice, different optimization algorithms or initializations often converge to distinct stationary solutions, even when applied to the same data. While such behavior has been repeatedly observed, a rigorous theoretical understanding of the stationary landscape of SURE-based SBL objectives—including the structure, multiplicity, and stability of stationary points—is still largely missing. Most existing analyses of SBL focus either on algorithmic convergence properties or on recovery performance under specific assumptions, and do not directly address the geometry of SURE-driven hyper-parameter optimization.

In parallel, large-system analyses based on random matrix theory (RMT) have proved to be powerful tools for understanding high-dimensional estimators. By revealing effective scalar decouplings and self-averaging phenomena, RMT has enabled precise characterizations of ridge regression, approximate message passing, and related linear estimators in the high-dimensional regime [9, 10]. More recently, such techniques have been used to analyze nonconvex objectives and implicit regularization effects in large-scale learning problems. Nevertheless, their potential for elucidating the stationary behavior of SURE-based hyper-parameter learning in SBL has not yet been fully explored.

In this paper, we develop a large-system analysis of stationary points arising from output-domain SURE based hyper-parameter optimization in sparse Bayesian learning based on our prior works [11]. Our analysis focuses on the stationary conditions themselves, rather than on any specific optimization algorithm. By leveraging exact coordinatewise optimality con-

We write $a \triangleq b$ for "a is defined as b". Equality in distribution is denoted by $X \overset{d}{=} Y$, and weak convergence by $X_n \Rightarrow X$. The notation $O_{\mathbb{P}}(\cdot)$ denotes stochastic order. The positive part is $(t)_+ \triangleq \max\{t, 0\}$. The trace of a matrix is denoted by $\mathrm{tr}(\cdot)$, and $\mathrm{diag}(\cdot)$ denotes a diagonal matrix. Expectation and probability are denoted by $\mathbb{E}[\cdot]$ and $\Pr(\cdot)$, respectively.

ditions, leave-one-out techniques, and random matrix theory, we show that the high-dimensional stationary rules admit a deterministic decoupling in the large-system limit. This decoupling leads to a distributional fixed-point characterization of stationary solutions, revealing that multiple stationary points may coexist and that their observability is governed by local stability properties. Together, these results provide a principled theoretical description of the stationary landscape of SURE-based hyper-parameter learning in SBL, without relying on simulations or algorithm-specific assumptions.

2. PROBLEM SETUP AND NOTATION

2.1. System Model

We consider the linear model

$$\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \mathbf{v}, \tag{1}$$

where $\mathbf{A} \in \mathbb{R}^{M\times N}$ has i.i.d. entries

$$A_{mn} \sim \mathcal{N}\left(0, \frac{1}{M}\right), \tag{2}$$

and $\mathbf{v} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}_M)$ is independent of $\mathbf{A}$. We study the large-system regime

$$M, N \to \infty, \quad \frac{M}{N} \to \delta \in (0, 1). \tag{3}$$

2.2. Hyper-parameters and estimators

Let $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_N)^\top \in [0, \infty)^N$ and define

$$\boldsymbol{\Lambda} \triangleq \operatorname{diag}(\lambda_1, \dots, \lambda_N). \tag{4}$$

Define the regularized Gram matrix

$$\mathbf{H}(\boldsymbol{\lambda}) \triangleq \mathbf{A}^\top\mathbf{A} + \boldsymbol{\Lambda}. \tag{5}$$

The associated linear estimator and predicted output are

$$\hat{\mathbf{x}}(\boldsymbol{\lambda}) \triangleq \mathbf{H}(\boldsymbol{\lambda})^{-1}\mathbf{A}^\top\mathbf{y}, \quad \hat{\mathbf{z}}(\boldsymbol{\lambda}) \triangleq \mathbf{A}\hat{\mathbf{x}}(\boldsymbol{\lambda}) = \mathbf{A}\mathbf{H}(\boldsymbol{\lambda})^{-1}\mathbf{A}^\top\mathbf{y}. \tag{6}$$

The output-domain SURE [4] is

$$\mathrm{SURE}_z(\boldsymbol{\lambda}) \triangleq \|\hat{\mathbf{z}}(\boldsymbol{\lambda}) - \mathbf{y}\|_2^2 - M\sigma^2 + 2\sigma^2 \operatorname{Tr}\big(\mathbf{A}\mathbf{H}(\boldsymbol{\lambda})^{-1}\mathbf{A}^\top\big). \tag{7}$$

2.3. Leave-one-out objects and stationarity

For each $i \in \{1, \dots, N\}$ define the leave-one-out matrix

$$\bar{\mathbf{H}}_i(\boldsymbol{\lambda}) \triangleq \mathbf{H}(\boldsymbol{\lambda}) - \lambda_i\mathbf{e}_i\mathbf{e}_i^\top, \tag{8}$$

and the scalar

$$\alpha_i(\boldsymbol{\lambda}) \triangleq \mathbf{e}_i^\top\bar{\mathbf{H}}_i(\boldsymbol{\lambda})^{-1}\mathbf{e}_i. \tag{9}$$

Definition 1 (Coordinatewise stationary point)**.** A vector $\boldsymbol{\lambda}^\star \in [0, \infty]^N$ is a *coordinatewise stationary point* if for every i, holding $\{\lambda_j^\star\}_{j\neq i}$ fixed, $\lambda_i^\star$ minimizes SURE_z along the i-th coordinate.

3. COORDINATEWISE QUADRATIC FORM AND STATIONARY RULE

Fix i and fix $\{\lambda_j\}_{j\neq i}$. Write

$$\mathbf{H} = \bar{\mathbf{H}}_i + \lambda_i\mathbf{e}_i\mathbf{e}_i^\top. \tag{10}$$

By the Sherman–Morrison identity,

$$\mathbf{H}^{-1} = \bar{\mathbf{H}}_i^{-1} - \frac{\lambda_i}{1+\lambda_i\alpha_i}\bar{\mathbf{H}}_i^{-1}\mathbf{e}_i\mathbf{e}_i^\top\bar{\mathbf{H}}_i^{-1}, \tag{11}$$

where α_i is defined in (9). Introduce the scalar reparameterization

$$\beta_i \triangleq \frac{\lambda_i}{1+\lambda_i\alpha_i}, \quad \beta_i \in [0, \alpha_i^{-1}), \quad \lambda_i = \frac{\beta_i}{1-\alpha_i\beta_i}. \tag{12}$$

Define

$$\boldsymbol{\Pi}_{-i} \triangleq \mathbf{A}\bar{\mathbf{H}}_i^{-1}\mathbf{A}^\top, \quad \mathbf{u}_i \triangleq \mathbf{A}\bar{\mathbf{H}}_i^{-1}\mathbf{e}_i, \quad s_i \triangleq \mathbf{u}_i^\top\mathbf{y}, \tag{13}$$

and the residual

$$\mathbf{r}_{-i} \triangleq (\boldsymbol{\Pi}_{-i} - \mathbf{I}_M)\mathbf{y}. \tag{14}$$

Multiplying (11) by $\mathbf{A}$ and $\mathbf{A}^\top$ yields the exact rank-one form

$$\mathbf{A}\mathbf{H}^{-1}\mathbf{A}^\top = \boldsymbol{\Pi}_{-i} - \beta_i\,\mathbf{u}_i\mathbf{u}_i^\top. \tag{15}$$

Hence

$$\hat{\mathbf{z}} - \mathbf{y} = \mathbf{r}_{-i} - \beta_i\,\mathbf{u}_i s_i. \tag{16}$$

Expanding the norm and the trace term gives

$$\|\hat{\mathbf{z}} - \mathbf{y}\|_2^2 = \|\mathbf{r}_{-i}\|_2^2 - 2\beta_i\,\langle\mathbf{r}_{-i}, \mathbf{u}_i\rangle s_i + \beta_i^2\|\mathbf{u}_i\|_2^2 s_i^2, \tag{17}$$

$$\operatorname{Tr}(\mathbf{A}\mathbf{H}^{-1}\mathbf{A}^\top) = \operatorname{Tr}(\boldsymbol{\Pi}_{-i}) - \beta_i\|\mathbf{u}_i\|_2^2. \tag{18}$$

Substituting (17)–(18) into (7) yields

$$\mathrm{SURE}_z(\beta_i) = \mathrm{const} - 2C_{1,i}^{(z)}\beta_i + C_{2,i}^{(z)}\beta_i^2, \tag{19}$$

with

$$C_{1,i}^{(z)} \triangleq \langle\mathbf{r}_{-i}, \mathbf{u}_i\rangle s_i + \sigma^2\|\mathbf{u}_i\|_2^2, \quad C_{2,i}^{(z)} \triangleq \|\mathbf{u}_i\|_2^2 s_i^2. \tag{20}$$

Theorem 1 (Coordinatewise stationary rule)**.** *Any coordinatewise stationary point (Definition 1) satisfies, for every i,*

$$\lambda_i^\star = \begin{cases} 0, & C_{1,i}^{(z)} \le 0, \\ \dfrac{C_{1,i}^{(z)}}{C_{2,i}^{(z)} - \alpha_i C_{1,i}^{(z)}}, & C_{1,i}^{(z)} > 0, \quad C_{2,i}^{(z)} > \alpha_i C_{1,i}^{(z)}, \\ +\infty, & C_{2,i}^{(z)} \le \alpha_i C_{1,i}^{(z)}, \end{cases} \tag{21}$$

where α_i is defined in (9) *and* $(C_{1,i}^{(z)}, C_{2,i}^{(z)})$ *in* (20)*.*

Proof. Since (19) is a scalar quadratic over $\beta_i \in [0, \alpha_i^{-1})$, the minimizer is: (i) $\beta_i^\star = 0$ if $C_{1,i}^{(z)} \le 0$; (ii) the unconstrained minimizer $\beta_i^\star = C_{1,i}^{(z)}/C_{2,i}^{(z)}$ if feasible, i.e., $C_{2,i}^{(z)} > \alpha_i C_{1,i}^{(z)}$; otherwise (iii) the boundary $\beta_i \uparrow \alpha_i^{-1}$, corresponding to $\lambda_i^\star \to +\infty$ via (12). Mapping $\beta_i^\star$ back to $\lambda_i^\star$ yields (21). □

4. LARGE-SYSTEM DECOUPLING VIA RANDOM MATRIX THEORY

In this section, we analyze the large-system behavior of the leave-one-out quantity α_i defined in (9), which plays a central role in the coordinatewise stationary rule derived in Section 3. Our main goal is to show that α_i admits a deterministic equivalent depending only on the local hyper-parameter λ_i and a global scalar that captures the collective effect of all other coordinates.

We assume that finite (non-pruned) hyper-parameters are uniformly bounded: $0 \le \lambda_i \le \lambda_{\max} < \infty$, and that the empirical measures

$$F_{\Lambda^{(N)}} \triangleq \frac{1}{N}\sum_{i=1}^{N} \delta_{\lambda_i} \tag{22}$$

converge weakly to a limit F_Λ supported on $[0, \lambda_{\max}]$. This assumption rules out pathological growth of individual coordinates and ensures that normalized traces of the associated resolvents are well behaved.

Theorem 2 (Deterministic equivalent of α_i)**.** *Under* (1)–(3) *and the boundedness assumption above, there exists a deterministic scalar $\xi > 0$ such that for any deterministic index sequence $i = i(N)$,*

$$\alpha_i = \mathbf{e}_i^\top \bar{\mathbf{H}}_i^{-1}\mathbf{e}_i \xrightarrow{\mathbb{P}} \frac{1}{\lambda_i + \xi}. \tag{23}$$

Moreover, ξ is the unique solution of

$$\xi = \delta \bigg/ \left(1 + \int \frac{1}{\lambda + \xi}\, \mathrm{d}F_\Lambda(\lambda)\right). \tag{24}$$

Theorem 2 shows that, in the large-system limit, the dependence of α_i on the entire hyper-parameter vector $\boldsymbol{\lambda}$ collapses to a simple scalar form. The global parameter ξ acts as an effective regularization induced by the aggregate empirical distribution F_Λ.

Proof sketch. Let $\mathbf{Q} = (\mathbf{A}^\top\mathbf{A} + \boldsymbol{\Lambda})^{-1}$ and $\mathbf{T} = (\mathbf{I}_M + \mathbf{A}\boldsymbol{\Lambda}^{-1}\mathbf{A}^\top)^{-1}$. A standard leave-one-out identity gives

$$(\mathbf{Q})_{ii} = \left(\lambda_i + \mathbf{a}_i^\top \mathbf{T}_{-i}\mathbf{a}_i\right)^{-1},$$

where $\mathbf{a}_i$ denotes the ith column of $\mathbf{A}$ and $\mathbf{T}_{-i}$ removes the contribution of $\mathbf{a}_i$ from the resolvent.

Conditioned on $\mathbf{T}_{-i}$, the vector $\mathbf{a}_i$ is independent of $\mathbf{T}_{-i}$ and distributed as $\mathcal{N}(\mathbf{0}, \frac{1}{M}\mathbf{I})$. Consequently, Gaussian quadratic-form concentration yields

$$\mathbf{a}_i^\top \mathbf{T}_{-i}\mathbf{a}_i - \frac{1}{M}\operatorname{Tr}(\mathbf{T}_{-i}) \xrightarrow{\mathbb{P}} 0.$$

Since $\mathbf{T}$ and $\mathbf{T}_{-i}$ differ only by a rank-one update, their normalized traces are asymptotically equivalent, i.e.,

$$\frac{1}{M}\operatorname{Tr}(\mathbf{T}_{-i}) - \frac{1}{M}\operatorname{Tr}(\mathbf{T}) \xrightarrow{\mathbb{P}} 0.$$

Defining $\xi^{(N)} = \frac{1}{M}\operatorname{Tr}(\mathbf{T})$, standard self-averaging arguments imply $\xi^{(N)} \to \xi$ deterministically, and thus $(\mathbf{Q})_{ii} \to (\lambda_i + \xi)^{-1}$ in probability.

Finally, $\bar{\mathbf{H}}_i^{-1}$ differs from $\mathbf{H}^{-1}$ by a diagonal rank-one perturbation, which implies $\alpha_i - (\mathbf{H}^{-1})_{ii} \to 0$ in probability and yields (23). The trace identity relating ξ and $\frac{1}{N}\operatorname{Tr}(\mathbf{Q})$ leads to the fixed-point equation (24). □

Remark 1. *The decoupling result in Theorem 2 is the key step that enables a distributional characterization of stationary points in the subsequent sections.*

5. DISTRIBUTIONAL FIXED POINTS OF STATIONARY SOLUTIONS

5.1. Empirical stationary distributions

Let $\boldsymbol{\lambda}^{(N)}$ be any sequence of stationary points and define the empirical distribution

$$F_{\Lambda^{(N)}} \triangleq \frac{1}{N}\sum_{i=1}^{N} \delta_{\lambda_i^{(N)}}. \tag{25}$$

Along any convergent subsequence we write $F_{\Lambda^{(N)}} \Rightarrow F_\Lambda^\star$.

5.2. Deterministic equivalents needed for the limiting map

Besides α_i, the coefficients in Theorem 1 depend on $\|\mathbf{u}_i\|_2^2$ and s_i in (13). The next lemma records their large-system deterministic equivalents at the level needed to define a limiting scalar map.

Lemma 1 (Deterministic equivalent of $\|\mathbf{u}_i\|_2^2$)**.** *Under the assumptions of Theorem 2, for any deterministic $i = i(N)$,*

$$\|\mathbf{u}_i\|_2^2 = \|\mathbf{A}\bar{\mathbf{H}}_i^{-1}\mathbf{e}_i\|_2^2 \xrightarrow{\mathbb{P}} \nu(\lambda_i) \triangleq \alpha(\lambda_i) - \lambda_i\alpha(\lambda_i)^2 = \frac{\xi^\star}{(\lambda_i + \xi^\star)^2}, \tag{26}$$

where $\alpha(\lambda) = (\lambda + \xi^\star)^{-1}$ and $\xi^\star$ is the self-consistent solution associated with $F_\Lambda^\star$.

Proof sketch. Using $\mathbf{A}^\top\mathbf{A} = \bar{\mathbf{H}}_i - \sum_{j\neq i}\lambda_j\mathbf{e}_j\mathbf{e}_j^\top$ and the identity $\bar{\mathbf{H}}_i^{-1}\bar{\mathbf{H}}_i\bar{\mathbf{H}}_i^{-1} = \bar{\mathbf{H}}_i^{-1}$, one obtains $\|\mathbf{u}_i\|_2^2 = \alpha_i - \lambda_i\alpha_i^2$. Apply Theorem 2. □

5.3. Limiting scalar stationary map

Let $\xi^\star$ denote the solution of (24) with $F_\Lambda = F_\Lambda^\star$. Define the scalar function

$$\alpha(\lambda) \triangleq \frac{1}{\lambda + \xi^\star}. \tag{27}$$

In the large-system limit, the coordinatewise stationary rule (21) induces a deterministic scalar map $\mathcal{T}(\cdot;\xi^\star)$ obtained by replacing the random quantities in (21) by their deterministic equivalents (e.g., (27), (26), and analogous DEs for the remaining terms in $C_{1,i}^{(z)}$ and $C_{2,i}^{(z)}$).

Theorem 3 (Distributional fixed-point characterization)**.** *Let $F_\Lambda^\star$ be any weak limit of stationary empirical distributions* (25)*. Then a random variable $\Lambda^\star \sim F_\Lambda^\star$ satisfies*

$$\Lambda^\star \stackrel{d}{=} \mathcal{T}(\Lambda^\star; \xi^\star)\,, \tag{28}$$

where $\xi^\star$ is the self-consistent solution of (24) *associated with $F_\Lambda^\star$.*

Proof sketch. By Theorem 2 and Lemma 1, the coordinatewise rule (21) converges (in probability) to a deterministic scalar update depending on λ_i and $\xi^\star$. Since $\boldsymbol{\lambda}^{(N)}$ is stationary for each N, its empirical measure must be invariant under the limiting update, yielding (28). □

Proposition 1 (Existence and potential non-uniqueness)**.** *Equation* (28) *admits at least one solution $F_\Lambda^\star$. In general, uniqueness is not guaranteed due to the piecewise, thresholded nature of $\mathcal{T}$ and the self-consistent dependence on $\xi^\star$.*

Proof sketch. Existence follows from continuity of the induced pushforward map on a compact, convex set of probability measures and Schauder's fixed-point theorem [12]. Non-uniqueness is not ruled out since the map is generally not contractive. □

6. LOCAL STABILITY OF DISTRIBUTIONAL FIXED POINTS

This section provides a local stability interpretation of the distributional fixed points characterized in Section 5. While Section 5 establishes the existence of stationary distributions and highlights their potential non-uniqueness, stability analysis is essential for understanding which of these stationary solutions are actually observable under coordinatewise hyper-parameter optimization.

The key difficulty is that the fixed-point equation (28) is self-consistent: the distribution F_Λ determines the scalar parameter ξ, which in turn governs the induced mapping on distributions. Local stability therefore depends on how small perturbations of the empirical distribution propagate through this coupled mechanism.

Let $F_\Lambda^\star$ satisfy the distributional fixed-point equation (28). To probe its local stability, we consider small perturbations of the empirical distribution of hyper-parameters. Specifically, define

$$F^{(\varepsilon)} \triangleq (1-\varepsilon)F_\Lambda^\star + \varepsilon G, \tag{29}$$

where G is any probability measure supported on $[0, \lambda_{\max}]$ and $\varepsilon \in (0,1)$. Let $\xi^{(\varepsilon)}$ denote the solution of the self-consistent equation (24) associated with $F^{(\varepsilon)}$.

Lemma 2 (Linear response of ξ)**.** *As $\varepsilon \downarrow 0$, the perturbation of the global coupling parameter satisfies*

$$\xi^{(\varepsilon)} - \xi^\star = -\varepsilon\, \frac{\int (\lambda+\xi^\star)^{-1}\,\mathrm{d}G(\lambda) - \int (\lambda+\xi^\star)^{-1}\,\mathrm{d}F_\Lambda^\star(\lambda)}{\frac{1}{\delta} + \int (\lambda+\xi^\star)^{-2}\,\mathrm{d}F_\Lambda^\star(\lambda)} + o(\varepsilon) \tag{30}$$

Proof. Define $\Phi(\xi; F) = \xi - \delta/(1 + \int (\lambda+\xi)^{-1}\mathrm{d}F)$. Since $\partial_\xi \Phi(\xi^\star; F_\Lambda^\star) > 0$, the result follows from the implicit function theorem. □

Lemma 2 quantifies the first-order sensitivity of the global parameter ξ to perturbations of the empirical distribution. Since ξ fully parametrizes the scalar mapping $\mathcal{T}(\cdot;\xi)$ in (28), this linear response characterizes how distributional perturbations propagate through the self-consistent fixed-point mechanism.

Proposition 2 (Local stability)**.** *A distributional fixed point $F_\Lambda^\star$ is locally stable if the induced mapping $\mathcal{T}(\cdot;\xi)$ contracts sufficiently small perturbations of the empirical distribution.*

Proposition 2 formalizes the notion that only stable distributional fixed points can be reached and maintained under coordinatewise hyper-parameter optimization in the large-system limit.

Proposition 3 (Interpretation)**.** *The presence of multiple stationary solutions in SURE$_z$-based sparse Bayesian learning is an intrinsic consequence of the self-consistent, piecewise-defined stationary conditions, rather than an artifact of a particular algorithm or initialization.*

Together, the results of this section link the multiplicity of stationary solutions to stability properties of the underlying distributional fixed-point mapping, thereby clarifying which solutions are observable in practice.

7. CONCLUSION

This paper studied stationary points of output-domain SURE (SURE_z) based hyper-parameter optimization in sparse Bayesian learning from a large-system perspective. By analyzing the coordinatewise stationary conditions and leveraging random matrix theory, we showed that the high-dimensional problem admits a deterministic decoupling, leading to a distributional fixed-point characterization of stationary solutions. This analysis reveals that multiple stationary points may coexist and that their observability is governed by local stability properties. The scope of this work is intentionally limited to stationary-point analysis. We do not study specific optimization algorithms, convergence rates, or finite-dimensional performance. Nevertheless, the results provide a principled theoretical understanding of the stationary landscape of SURE-based hyper-parameter learning, and offer a useful foundation for future studies on algorithmic dynamics and stability mechanisms.

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.

8. 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. Wipf and S. Nagarajan, “A new view of automatic relevance determination,” in *Adv. Neural Inf. Process. Syst.*, 2008, vol. 20.

[3] C. M. Stein, “Estimation of the mean of a multivariate normal distribution,” *Ann. Statist.*, vol. 9, no. 6, pp. 1135–1151, 1981.

[4] Y. C. Eldar, “Generalized SURE for Exponential Families: Applications to Regularization,” *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 471–481, 2009.

[5] D. L. Donoho and I. M. Johnstone, “Adapting to unknown smoothness via wavelet shrinkage,” *J. Amer. Statist. Assoc.*, vol. 90, no. 432, pp. 1200–1224, 1995.

[6] E. J. Candès, M. B. Wakin, and S. P. Boyd, “Enhancing sparsity by reweighted ℓ_1 minimization,” *J. Fourier Anal. Appl.*, vol. 14, no. 5–6, pp. 877–905, 2008.

[7] F. Xiao and D. T. M. Slock, “Towards hyperparameter optimizing of sparse bayesian learning based on Stein’s unbiased risk estimator,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2024.

[8] F. Yu, L. Shen, and G. Song, “Hyperparameter estimation for sparse bayesian learning models,” *arXiv preprint*, vol. arXiv:2401.02544, 2024.

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

[10] Z. Bai and J. W. Silverstein, *Spectral Analysis of Large Dimensional Random Matrices*, Springer, 2010.

[11] F. Xiao and D. T. M. Slock, “Performance analysis of hyperparameter optimization in sparse bayesian learning via Stein’s unbiased risk estimator,” in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, 2025, pp. 2437–2441.

[12] E. Zeidler, *Nonlinear Functional Analysis and Its Applications I: Fixed-Point Theorems*, Springer, 1986.