Bernoulli and Gauss Take a Look at the MAP

Pierre Barbault, Matthieu Kowalski, Charles Soussen

Bernoulli and Gauss Take a Look at the MAP

Pierre Barbault[a], Matthieu Kowalski[b], Charles Soussen[a]

[a]*L2S, CentraleSupelec, Université Paris-Saclay, CNRS, Gif-sur-Yvette, France*
[b]*Inria, Université Paris-Saclay, CNRS, Laboratoire Interdisciplinaire des Sciences du Numériques, Gif-sur-Yvette, France*

Abstract

This article discusses two Maximum *a Posteriori* (MAP) interpretations for state-of-the-art methods used in sparse inverse problems: the joint-MAP and the Marginal-MAP. Canonically rooted in a Bayesian framework, sparsity is modeled by a general spike and slab distribution. The focus is on the recovery of the solution support rather than on signal amplitudes. We study the prominent Bernoulli-Gaussian model leading to NP-hard optimization problems. We show that a judicious re-parametrization of the joint-MAP may indeed be a nice surrogate of the marginal-MAP. Additionally, we explore common continuous relaxations of the support and encompass them under the scope of a parametrized distribution. Upon describing the behavior of a few relaxations, strong links are established between the Bernoulli-Gaussian joint-MAP, marginal-MAP, and well-studied methods such as the Lasso and Sparse Bayesian Learning. Finally, the utilization of randomized rounding for both joint-MAP and marginal-MAP problems yields valuable insights into obtaining sparse solutions with an emphasis on support recovery.

Keywords:
Bernoulli-Gaussian prior, spike-and-slab prior, maximum *a posteriori* estimation, randomized rounding
2020 MSC: 62J07,
2020 MSC: 62F15

1. Introduction

Inverse problems aim to extract relevant information from degraded acquisitions. The following general linear model summarizes this framework

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{e}\,, \tag{1}$$

where $\mathbf{y} \in \mathbb{R}^M$ is the measurement vector given by the M sensors and $\mathbf{x} \in \mathbb{R}^N$ is the unknown source signal. The forward operator $\mathbf{H} \in \mathbb{R}^{M\times N}$ is supposed to be known. Then, the measurements are degraded by a noise $\mathbf{e} \in \mathbb{R}^M$, which is supposed to be white and Gaussian of variance σ_e^2, i.e., $\mathbf{e} \sim \mathcal{N}(0, \sigma_e^2\mathbf{I}_M)$. In many applications, the structure of $\mathbf{H}$ makes retrieving the source signal an ill-posed problem. Additional information must thus be artificially introduced, taking the form of knowledge of the prior source distribution. Here, the interest is laid upon sparse signals, in which most entries in $\mathbf{x}$ are equal to 0.

Sparsity has become one of the most successful state-of-the-art approaches for solving ill-posed inverse problems. Among optimization-based approaches for sparse reconstruction, the Lasso proposed by Tibshirani (1996) and variants (Argyriou et al., 2012; McDonald et al., 2016) are probably the most popular because they can be formulated as convex optimization problems. Although the Lasso is a powerful method to identify the support of $\mathbf{x}$, the nonzero amplitudes are underestimated due to the ℓ_1 regularization. In practice, debiasing has to be performed in a post-processing step (Bruce and Gao, 1996; Wright et al., 2009).

On the other hand, the Bayesian framework makes it possible to model uncertainties in the source signal within a probabilistic setting, based on likelihood and prior distributions. Bayesian inference then consists of optimizing or sampling the posterior density of our variable of interest $\mathbf{x}$ using Bayes'rule. The Bayesian approach to decision theory, as described for example by (Robert, 1997), aims at minimizing the Bayes' risk, that is, the posterior expected loss

$$\mathbb{E}_{\mathbf{x}|\mathbf{y}}\left\{\mathcal{L}(\mathbf{x}, \hat{\mathbf{x}})\right\}\,, \tag{2}$$

where $\hat{\mathbf{x}}$ is some estimator of $\mathbf{x}$ and $\mathcal{L}$ a given loss function. The Bayes estimator is then the minimizer of the Bayes risk. Two point estimators are of particular interest:

1. The posterior expectation is the Bayes estimator for the squared ℓ_2 loss. This estimator is often called the MMSE (Minimum Mean Squared Error) estimator.
2. The Maximum A Posteriori (MAP) is the Bayes estimator for the $1-0$ loss and discrete data.

At first sight, a suitable solution of the inverse problem defined by Equation (1) should minimize the mean square error with respect to $\mathbf{x}$. Then, the natural Bayes estimator should be the posterior expectation. In this case, the MAP is not a suitable Bayes estimator because the coefficients of $\mathbf{x}$ are continuous valued. Moreover, the number of parameters increases with the number of observations, so the equivalence between the MAP and the maximum likelihood estimators does not apply. This consideration of the definition of a Bayes estimator corroborates the analysis made by Gribonval (2011) on the interpretation of a chosen prior in an optimization framework.

While the Lasso and its variants can be interpreted as MAP estimators with suitable priors, this Bayesian interpretation is one possible, but is not the unique as debated by Gribonval and Machart (2013). In his talk, Tibshirani (2010) even says humorously, but provocatively, in the "Top 7 reasons why this Lasso/L1 stuff may have gone too far": *The Bayesians are getting really pissed off!*. (Gribonval and Nikolova, 2021) extends the discussion on the Bayesian interpretation of proximity operators and shows that MMSE estimation can be formulated as a penalized least squares optimization problem beyond the Gaussian noise. One of the take-home message of this discussion is that the MAP interpretation of penalized least-square optimization problems is not always the most relevant, since the MMSE estimator can also be written as a penalized least-square optimization problem. In this article, we take a complementary point of view by showing that the usual MAP approaches consisting of minimizing a $\ell_2 + \ell_\alpha$ functionnal, with $\alpha < 2$, are good surrogates for the so-called Marginal-MAP estimator of the support of the sparse signal for which the MAP is the Bayes estimator.

The starting point of this work is the statistical description of sparse signals in the form $x_n = q_n r_n$, where $q_n = \mathbb{1}_{x_n \neq 0}$ is a binary process modeling the support of $\mathbf{x}$, and r_n refers to the nonzero coefficients of $\mathbf{x}$ (Mendel, 1983). Let us stress that the main bottleneck in sparse estimation is the reconstruction of the support $\mathbf{q}$. Two classical MAP estimators can usually be considered to recover $\mathbf{q}$, namely the joint-MAP ($\mathbf{q}$ and $\mathbf{r}$ are jointly estimated from data $\mathbf{y}$) and marginal-MAP ($\mathbf{q}$ is estimated alone, while variables $\mathbf{r}$ are marginalized out) as discussed by Champagnat et al. (1996).

Contributions. The first purpose of this article is to bring a unified view on classical approaches used for sparse inverse problems, including the Lasso, Sparse Bayesian Learning (SBL) and ℓ_0-minimization. The links between marginal and joint-MAP estimators lead to a novel interpretation of these methods. These links mainly come from the study of continuous relaxations of the spike-and-slab model, which are less classical in the inverse problem literature.

In Section 2, we study the joint-MAP and marginal-MAP estimators for the support of a Bernoulli-Gaussian (BG) model. The main results are stated in Theorem 2.1 and Theorem 2.4, which show that optimizing the joint-MAP criterion with an appropriate choice of hyper-parameter may help to optimize the marginal-MAP criterion. Since the BG model leads to challenging combinatorial optimization problems, we study a family of continuous priors for the support $\mathbf{q}$ in Section 3. Using an appropriate choice of the continuous prior, the joint-MAP reduces to the usual $\ell_2 + \ell_\alpha$ optimization problem. Specifically, we show that the Laplacian prior is not the only interpretation of the Lasso. An alternative prior for $\mathbf{x}$ is proposed, which leads to the same optimization problem. Moreover, the marginal-MAP approach sheds new light on Sparse Bayesian Learning (Wipf and Rao, 2004) which can be viewed as a continuous relaxation of the Bernoulli-Gaussian model. Finally, we study in Section 4 the randomized rounding approach Raghavan and Tompson (1987), that is rounding a fractional solution in $[0, 1]$ to an integer solution with appropriate randomization, to estimate a good Boolean solution for support the original joint-MAP and marginal-MAP problems.

2. The Bernoulli-Gaussian model and the Maxima A Posteriori

The Bernoulli-Gaussian model falls into the category of spike and slab models (Ishwaran and Rao, 2005; Yen, 2011) and is defined as follows. Let $\mathbf{q} \in \{0; 1\}^N$ be the support of $\mathbf{x}$, *i.e.*, $q_n = 1$ if $x_n \neq 0$ and $q_n = 0$ otherwise. In

addition, let $\mathbf{r} \in \mathbb{R}^N$ such that $x_n = q_n r_n$. When $q_n = 1$, r_n aims to model the amplitude of x_n. The probabilistic model for source signals reads

$$\mathbf{x} = \mathbf{Q}\mathbf{r} \quad \text{with} \quad \mathbf{Q} = \mathrm{Diag}\,[\mathbf{q}] \in \mathbb{R}^{N\times N}\,, \tag{3}$$

where q_n are independent and identically distributed (i.i.d) variables following the Bernoulli distribution of parameter $p \in (0,1)$ (with $p(q_n = 1) = p$), and r_n are i.i.d. centered Gaussian random variables:

$$\forall n,\;\; q_n \sim \mathcal{B}(p) \text{ and } r_n \sim \mathcal{N}(0, \sigma_r^2)\,. \tag{4}$$

To get sparsity, the parameter p of the Bernoulli distribution is assumed to satisfy $0 < p \leq 1/2$. In the following, we focus on estimating the support $\mathbf{q}$.

2.1. Alternative MAP formulations

Champagnat et al. (1996) define two MAP estimators by maximizing the joint posterior $\log p(\mathbf{r}, \mathbf{q}|\mathbf{y})$ in both $\mathbf{q}$ and $\mathbf{r}$, and the marginal posterior $\log p(\mathbf{q}|\mathbf{y})$ with respect to the support $\mathbf{q}$ only. The latter is Bayes' estimator and maximizes the accuracy of the estimated support.

Joint-MAP. Up to direct calculation, one can see that maximizing the joint posterior distribution $\log p(\mathbf{r}, \mathbf{q}|\mathbf{y})$ is equivalent to minimizing:

$$\mathcal{J}(\mathbf{r}, \mathbf{q}; \lambda) \triangleq \frac{1}{2\sigma_e^2}\|\mathbf{y} - \mathbf{H}\mathbf{Q}\mathbf{r}\|_2^2 + \frac{1}{2\sigma_r^2}\|\mathbf{r}\|_2^2 + \lambda\|\mathbf{q}\|_0\,, \tag{5}$$

where $\|\cdot\|_0$ stands for the ℓ_0-"norm" or number of non-zero entries, and (see, e.g., (Soussen et al., 2011))

$$\lambda = \log\left(\frac{1-p}{p}\right)\,. \tag{6}$$

Hereafter, $(\hat{\mathbf{r}}, \hat{\mathbf{q}})$ will refer to the MAP estimator, that is, the minimizer of (5). Furthermore, because of the sparsity assumption, $p < \frac{1}{2}$, thus λ will be treated as a non-negative hyperparameter.

One can notice from Equation (5) that the dependency of $\mathcal{J}$ upon $\mathbf{r}$ is quadratic. Hence, $(\hat{\mathbf{r}}, \hat{\mathbf{q}})$ needs to satisfy the consistency condition:

$$\hat{\mathbf{r}}(\hat{\mathbf{q}}) = \sigma_r^2\Big(\sigma_e^2\mathbf{I}_N + \sigma_r^2\mathbf{Q}^T\mathbf{H}^T\mathbf{H}\mathbf{Q}\Big)^{-1}\hat{\mathbf{Q}}^T\mathbf{H}^T\mathbf{y}\,, \tag{7}$$

which is derived by annealing the gradient of $\mathcal{J}$ with respect to $\mathbf{r}$. Using the Woodbury matrix identity, the latter equation can be rewritten as:

$$\hat{\mathbf{r}}(\hat{\mathbf{q}}) = \sigma_r^2\hat{\mathbf{Q}}^T\mathbf{H}^T\Gamma_y(\hat{\mathbf{q}})^{-1}\mathbf{y}\,, \tag{8}$$

where matrix $\Gamma_y(\mathbf{q})$ is defined for all $\mathbf{q} \in \{0;1\}^N$, by

$$\Gamma_y(\mathbf{q}) = \sigma_e^2\mathbf{I}_M + \sigma_r^2\mathbf{H}\mathbf{Q}\mathbf{Q}^T\mathbf{H}^T. \tag{9}$$

One can notice from (8) that $\hat{r}_n(\hat{\mathbf{q}}) = 0$ whenever $\hat{q}_n = 0$.

As shown by Soussen et al. (2011), minimizing Equation (5) jointly in $(\mathbf{q}, \mathbf{r}) \in \{0,1\}^N \times \mathbb{R}^N$ is equivalent to addressing the ℓ_0-penalized least squares problem:

$$\min_{\mathbf{x}\in\mathbb{R}^N}\left\{\frac{1}{2\sigma_e^2}\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \frac{1}{2\sigma_r^2}\|\mathbf{x}\|^2 + \lambda\|\mathbf{x}\|_0\right\}. \tag{10}$$

Another reformulation is obtained by plugging Equation (8) into Equation (5). The joint-MAP estimate $\mathbf{q}$ thus minimizes the criterion:

$$\mathcal{J}_0(\mathbf{q}; \lambda) \triangleq \mathcal{J}(\hat{\mathbf{r}}(\mathbf{q}), \mathbf{q}; \lambda)\,, \tag{11}$$

which will be referred to as the joint-MAP criterion and writes (Goussard et al., 1990):

$$\mathcal{J}_0\left(\mathbf{q};\lambda\right) = \frac{1}{2}\mathbf{y}^t\Gamma_y(\mathbf{q})^{-1}\mathbf{y} + \lambda\|\mathbf{q}\|_0\ , \tag{12}$$

with $\Gamma_y(\mathbf{q})$ defined in (9).

The global optimization of $\mathcal{J}_0(\mathbf{q};\lambda)$ is computationally expensive, the problem being NP-hard. Recently, Hazimeh et al. (2021) proposed a branch and bound based mixed integer programming method to solve exactly Equation (10). Alternatively, local minimizers can be obtained using a proximal descent similar to the iterative (hard) thresholding studied by Blumensath and Davies (2009); Kowalski (2014) or the majorize-minimization procedure proposed by Yen (2011) to optimize the BG spike and slab model. Greedy methods inspired from the single most likely replacement of Goussard et al. (1990) may also be employed, see (Soussen et al., 2011).

Marginal-MAP. Marginal-MAP estimation relies on the maximization of the posterior

$$p(\mathbf{q}|\mathbf{y}) = \int p(\mathbf{q},\mathbf{r}|\mathbf{y})\,\mathrm{d}\mathbf{r} \tag{13}$$

after marginalizing out $\mathbf{r}$. According to Champagnat et al. (1996), the marginal-MAP estimate can be found by minimizing

$$\mathcal{M}_0(\mathbf{q};\lambda) \triangleq \frac{1}{2}\mathbf{y}^t\Gamma_y^{-1}(\mathbf{q})\mathbf{y} + \frac{1}{2}\log|\Gamma_y(\mathbf{q})| + \lambda\|\mathbf{q}\|_0\ . \tag{14}$$

Hence, using Equation (12), we have:

$$\mathcal{M}_0(\mathbf{q};\lambda) = \mathcal{J}_0\left(\mathbf{q};\lambda\right) + \frac{1}{2}\log\left|\Gamma_y(\mathbf{q})\right|. \tag{15}$$

The marginal-MAP aims to provide a Bayesian estimator of the support of the solution. Once this support is known, one can still estimate the non-zero coefficients using Equation (8). (Protter et al., 2010) compare this estimate with the MMSE estimate of the BG signal for matrices $\mathbf{H}$ having orthogonal columns.

Hereafter, we study the relation between the minimizers of the marginal-MAP criterion $\mathcal{M}_0(\mathbf{q},\lambda)$ defined in Equation (14) and those of the joint-MAP $\mathcal{J}_0(\mathbf{q},\lambda)$ defined in Equation (12).

2.2. Joint-MAP vs Marginal-MAP minimizers

In the case of orthogonal matrices $\mathbf{H}$, the MAP estimators have a closed-form expression involving a thresholding operation on the amplitudes y_n; see Appendix B. The estimated supports $\mathbf{q}$ related to $\mathcal{J}_0$ and $\mathcal{M}_0$ differ only by the thresholding value. Furthermore, Turek et al. (2011) performed a comparative theoretical and experimental analysis of the posterior mean estimator of $\mathbf{x}$ and its "marginal-MAP" estimator $\mathbf{x} = \mathbf{Q}^M\hat{\mathbf{r}}(\mathbf{q}^M)$, where $\mathbf{q}^M$ is obtained by minimizing Equation (14) and the $\hat{\mathbf{r}}$ operator was defined in (8). For arbitrary (non-orthogonal) matrices $\mathbf{H}$ of rank $\min(M,N)$, these thresholds appear naturally in the following bounds of the marginal-MAP functional.

Theorem 2.1. *Suppose that the rank of* $\mathbf{H}$ *is* $\min(M,N)$. *Let* $\mathcal{J}_0(\mathbf{q};\lambda)$ *and* $\mathcal{M}_0(\mathbf{q};\lambda)$ *be the joint and marginal-MAP criteria given by* (12) *and* (14)*, respectively. Let* $L = \|\mathbf{H}\mathbf{H}^T\|$ *be the largest eigenvalue of* $\mathbf{H}\mathbf{H}^T$*, then for all* $\lambda > 0$*,*

$$\mathcal{J}_0(\mathbf{q};\lambda) + \frac{M}{2}\log(\sigma_e^2) \le \mathcal{M}_0(\mathbf{q};\lambda) \le \mathcal{J}_0(\mathbf{q};\kappa(\lambda)) + \frac{M}{2}\log(\sigma_e^2) \tag{16}$$

with

$$\kappa(\lambda) = \lambda + \log\left(\sqrt{\frac{L\sigma_r^2+\sigma_e^2}{\sigma_e^2}}\right). \tag{17}$$

Moreover

$$\mathcal{M}_0(\mathbf{q};\lambda) = \mathcal{J}_0(\mathbf{q};\kappa(\lambda)) + \frac{M}{2}\log(\sigma_e^2) \tag{18}$$

iff the nonzero columns of $\mathbf{HQ}$ *are orthogonal with norm* $\sqrt{L}$.

All proofs are deferred to the supplementary materials. In the following, the notation $\kappa(\lambda)$ will refer to the definition given in Theorem 2.1. Moreover, we will always assume that $\text{rank}(\mathbf{H}) = \min(M, N)$.

Unfortunately, this theorem does not allows one to conclude anything about the computation of the minimizers of $\mathcal{M}_0(\mathbf{q};\lambda)$ from those of $\mathcal{J}_0(\mathbf{q};\kappa_{\mathbf{q}})$ when the equality (18) is not achieved. In Appendix B.1, we provide a closed form expression of the minimizers of the both criteria in the orthogonal case.

Hereafter, we study how some global minimizers of $\mathcal{J}_0(\mathbf{q};\kappa(\lambda))$ may be related to the minimizers of $\mathcal{M}_0(\mathbf{q};\lambda)$. Let ($\mathbf{q}^{\mathcal{J}}$) (resp. $\mathbf{q}^{\mathcal{M}}$) be a global minimizer of $\mathcal{J}_0(\mathbf{q};\kappa(\lambda))$ (resp. $\mathcal{M}_0(\mathbf{q},\lambda)$). Using the singular value decomposition of $\mathbf{HQ}^{\mathcal{J}}$ (resp. $\mathbf{HQ}^{\mathcal{M}}$), we can derive necessary and sufficient conditions that must be satisfied by $\mathbf{q}^{\mathcal{J}}$ (resp. $\mathbf{q}^{\mathcal{M}}$). Next Lemma states these conditions for the JMAP.

Lemma 2.2 (joint-MAP necessary and sufficient conditions)**.** *Suppose that the rank of* $\mathbf{H}$ *is* $\min(M,N)$*. Let* $\mathbf{q}^{\mathcal{J}}$ *be a global minimizer of* $\mathcal{M}_0(\mathbf{q};\lambda)$*. Let* $z_k^{\mathcal{J}}$ *be the singular values of* $\mathbf{HQ}^{\mathcal{J}}$*, such that* $\mathbf{HQ}^{\mathcal{J}} = \mathbf{U}^{\mathcal{J}}\text{Diag}\left[z_k^{\mathcal{J}}\right](\mathbf{V}^{\mathcal{J}})^T$ *with orthogonal (unitary) matrices* $\mathbf{U}^{\mathcal{J}}$ *and* $\mathbf{V}$*. Let* $\mathbf{q}$ *be any support and let* z_k *be the singular values of* $\mathbf{HQ}$*, such that* $\mathbf{HQ} = \mathbf{U}\text{Diag}\left[z_k\right]\mathbf{V}^T$ *with orthogonal (unitary) matrices* $\mathbf{U}$ *and* $\mathbf{V}$*. We denote by* $\mathbf{u}_k^{\mathcal{J}}$ *(resp.* $\mathbf{u}_k$*) the columns of* $\mathbf{U}^{\mathcal{J}}$ *(resp.* $\mathbf{U}$*).*

Then $\mathbf{q}^{\mathcal{J}}$ *is a global minimizer of* $\mathcal{J}(\mathbf{q};\lambda)$ *iff the following conditions are satisfied for all* $\mathbf{q}$

JMAP Upper Singular Conditions: *If* $\|\mathbf{q}\|_0 \leq \|\mathbf{q}^{\mathcal{J}}\|_0$

$$\left(\|\mathbf{q}^{\mathcal{J}}\|_0 - \|\mathbf{q}\|_0\right)2\sigma_e^2\lambda \leq \sum_{k=\|\mathbf{q}\|_0+1}^{\|\mathbf{q}^{\mathcal{J}}\|_0}\left(\frac{(z_k^{\mathcal{J}})^2\sigma_r^2}{(z_k^{\mathcal{J}})^2\sigma_r^2+\sigma_e^2}\right)\langle\mathbf{u}_k{}^{\mathcal{J}},\mathbf{y}\rangle^2 + \sum_{k=1}^{\|\mathbf{q}\|_0}\left\{\left(\frac{(z_k^{\mathcal{J}})^2\sigma_r^2}{(z_k^{\mathcal{J}})^2\sigma_r^2+\sigma_e^2}\right)\langle\mathbf{u}_k^{\mathcal{J}},\mathbf{y}\rangle^2 - \left(\frac{z_k^2\sigma_r^2}{z_k^2\sigma_r^2+\sigma_e^2}\right)\langle\mathbf{u}_k,\mathbf{y}\rangle^2\right\}$$

JMAP Lower Singular Conditions: *If* $\|\mathbf{q}^{\mathcal{J}}\|_0 \leq \|\mathbf{q}\|_0$

$$\left(\|\mathbf{q}\|_0 - \|\mathbf{q}^{\mathcal{J}}\|_0\right)2\sigma_e^2\lambda \geq \sum_{k=\|\mathbf{q}^{\mathcal{J}}\|_0+1}^{\|\mathbf{q}\|_0}\left(\frac{z_k^2\sigma_r^2}{z_k^2\sigma_r^2+\sigma_e^2}\right)\langle\mathbf{u}_k^{\mathcal{J}},\mathbf{y}\rangle^2 + \sum_{k=1}^{\|\mathbf{q}\|_0}\left\{\left(\frac{(z_k^{\mathcal{J}})^2\sigma_r^2}{(z_k^{\mathcal{J}})^2\sigma_r^2+\sigma_e^2}\right)\langle\mathbf{u}_k^{\mathcal{J}},\mathbf{y}\rangle^2 - \left(\frac{z_k^2\sigma_r^2}{z_k^2\sigma_r^2+\sigma_e^2}\right)\langle\mathbf{u}_k,\mathbf{y}\rangle^2\right\}$$

For the MMAP, similar conditions are derived in the next Lemma.

Lemma 2.3 (marginal-MAP necessary and sufficient conditions)**.** *Suppose that the rank of* $\mathbf{H}$ *is* $\min(M,N)$*. Let* $\mathbf{q}^{\mathcal{M}}$ *be a global minimizer of* $\mathcal{M}_0(\mathbf{q};\lambda)$*. Let* $z_k^{\mathcal{M}}$ *be the singular values of* $\mathbf{HQ}^{M}$*, such that* $\mathbf{HQ}^{\mathcal{M}} = \mathbf{U}^{\mathcal{M}}\text{Diag}\left[z_k^{\mathcal{M}}\right](\mathbf{V}^{\mathcal{M}})^T$ *with orthogonal (unitary) matrices* $\mathbf{U}^{\mathcal{M}} \in \mathbb{R}^{MM}$ *and* $\mathbf{V}^{\mathcal{M}} \in \mathbb{R}^{NN}$*. Let* $\mathbf{q}$ *be any support and let* z_k *be the singular values of* $\mathbf{HQ}$*, such that* $\mathbf{HQ} = \mathbf{U}\text{Diag}\left[z_k\right]\mathbf{V}^T$ *with orthogonal (unitary) matrices* $\mathbf{U}$ *and* $\mathbf{V}$*. We denote by* $\mathbf{u}_k^{\mathcal{M}}$ *(resp.* $\mathbf{u}_k$*) the columns of* $\mathbf{U}^{\mathcal{M}}$ *(resp.* $\mathbf{U}$*).*

Then $\mathbf{q}^{\mathcal{M}}$ *is a global minimizer of* $\mathcal{M}(\mathbf{q};\lambda)$ *iff the following conditions are satisfied for all* $\mathbf{q}$

MMAP Upper Singular Conditions: *If* $\|\mathbf{q}\|_0 \leq \|\mathbf{q}^{\mathcal{M}}\|_0$

$$\left(\|\mathbf{q}^{\mathcal{M}}\|_0 - \|\mathbf{q}\|_0\right)2\sigma_e^2\left(\lambda + \log\left(\sqrt{\frac{\left(z_k^{\mathcal{M}}\sigma_r\right)^2+\sigma_e^2}{\sigma_e^2}}\right)\right) \leq \sum_{k=\|\mathbf{q}\|_0+1}^{\|\mathbf{q}^{\mathcal{M}}\|_0}\left(\frac{(z_k^{\mathcal{M}})^2\sigma_r^2}{(z_k^{\mathcal{M}})^2\sigma_r^2+\sigma_e^2}\right)\langle\mathbf{u}_k^{\mathcal{M}},\mathbf{y}\rangle^2 +$$
$$\sum_{k=1}^{\|\mathbf{q}\|_0}\left\{\left(\frac{(z_k^{\mathcal{M}})^2\sigma_r^2}{(z_k^{\mathcal{M}})^2\sigma_r^2+\sigma_e^2}\right)\langle\mathbf{u}_k^{\mathcal{M}},\mathbf{y}\rangle^2 - \left(\frac{z_k^2\sigma_r^2}{z_k^2\sigma_r^2+\sigma_e^2}\right)\langle\mathbf{u}_k,\mathbf{y}\rangle^2 + \log\left(\sqrt{\frac{(z_k\sigma_r)^2+\sigma_e^2}{\sigma_e^2}}\right)\right\}$$

MMAP Lower Singular Conditions: *If* $\|\mathbf{q}^{\mathcal{M}}\|_0 \leq \|\mathbf{q}\|_0$

$$\left(\|\mathbf{q}\|_0 - \|\mathbf{q}^{\mathcal{M}}\|_0\right)2\sigma_e^2\left(\lambda + \log\left(\sqrt{\frac{\left(z_k^{\mathcal{M}}\sigma_r\right)^2+\sigma_e^2}{\sigma_e^2}}\right)\right) \geq \sum_{k=\|\mathbf{q}^{\mathcal{M}}\|_0+1}^{\|\mathbf{q}\|_0}\left(\frac{z_k^2\sigma_r^2}{z_k^2\sigma_r^2+\sigma_e^2}\right)\langle\mathbf{u}_k,\mathbf{y}\rangle^2 +$$
$$\sum_{k=1}^{\|\mathbf{q}\|_0}\left\{\left(\frac{(z_k^{\mathcal{M}})^2\sigma_r^2}{(z_k^{\mathcal{M}})^2\sigma_r^2+\sigma_e^2}\right)\langle\mathbf{u}_k^{\mathcal{M}},\mathbf{y}\rangle^2 - \left(\frac{z_k^2\sigma_r^2}{z_k^2\sigma_r^2+\sigma_e^2}\right)\langle\mathbf{u}_k,\mathbf{y}\rangle^2 + \log\left(\sqrt{\frac{(z_k\sigma_r)^2+\sigma_e^2}{\sigma_e^2}}\right)\right\}$$

Because for all k we have

$$\lambda + \log\left(\sqrt{\frac{\left(z_k^{\mathcal{M}}\sigma_r\right)^2 + \sigma_e^2}{\sigma_e^2}}\right) \leq \lambda + \log\left(\sqrt{\frac{L\sigma_r^2 + \sigma_e^2}{\sigma_e^2}}\right) = \kappa(\lambda) \quad \text{and} \quad \log\left(\sqrt{\frac{\left(z_k\sigma_r\right)^2 + \sigma_e^2}{\sigma_e^2}}\right) > 0\,,$$

it is worth noting that if a support $\mathbf{q}$ satisfies the JMAP Upper Singular Conditions for $\mathcal{J}(\mathbf{q};\kappa(\lambda))$, then $\mathbf{q}$ also satisfies the MMAP Upper Singular Conditions for $\mathcal{M}(\mathbf{q};\lambda)$. Conversely, if $\mathbf{q}$ satisfies the MMAP Lower Singular Conditions for $\mathcal{M}(\mathbf{q};\lambda)$, then $\mathbf{q}$ satisfies the JMAP Lower Singular Conditions for $\mathcal{J}(\mathbf{q};\kappa(\lambda))$. Thanks to the two previous Lemmas, this remark allows one to directly state the following Theorem

Theorem 2.4. *Suppose that the rank of* $\mathbf{H}$ *is* $\min(M,N)$. *Let* $\mathbf{q}^*$ *be a support and let Let* z_k^* *be the singular values of* $\mathbf{HQ}^*$, *such that* $\mathbf{HQ}^* = \mathbf{U}^*\mathrm{Diag}\left[z_k^*\right](\mathbf{V}^*)^T$ *with orthogonal (unitary) matrices* $\mathbf{U}^* \in \mathbb{R}^{MM}$ *and* $\mathbf{V}^* \in \mathbb{R}^{NN}$. *Let* $\mathbf{q}$ *be any support and let* z_k *be the singular values of* $\mathbf{HQ}$, *such that* $\mathbf{HQ} = \mathbf{U}\mathrm{Diag}\left[z_k\right]\mathbf{V}^T$ *with orthogonal (unitary) matrices* $\mathbf{U}$ *and* $\mathbf{V}$. *We denote by* $\mathbf{u}_k^*$ *(resp.* $\mathbf{u}_k$*) the columns of* $\mathbf{U}^*$ *(resp.* $\mathbf{U}$*).*

If $\mathbf{q}^*$ *satisfies the JMAP Upper Singular Conditions for* $\mathcal{J}(\mathbf{q};\kappa(\lambda)$ *and the MMAP Lower Singular Conditions for* $\mathcal{M}(\mathbf{q};\lambda)$, *then* $\mathbf{q}^*$ *is a global minimizer of* $\mathcal{J}(\mathbf{q};\kappa(\lambda))$ *and a global minimizer of* $\mathcal{M}(\mathbf{q};\lambda)$.

That is, minimizing the JMAP with the right value of the hyper-parameter $\kappa(\lambda)$ may lead to a minimizer of the MMAP with hyper-parameter λ. The difficulty comes to identify when the set of support satisfying both the Upper Singular Conditions for $\mathcal{J}(\mathbf{q};\kappa(\lambda))$ and the MMAP Lower Singular Conditions for $\mathcal{M}(\mathbf{q};\lambda)$ may be empty or not.

In their study, Argyriou et al. (2012) proved that the Lasso and the k-support norm are the most effective convex relaxations of the ℓ_0 penalty and the ℓ_2-constrained ℓ_0 penalty, respectively. These convex relaxations enable one to design efficient and guaranteed optimization solvers. In the following section, the discrete Bernoulli distribution is replaced with a continuous distribution, which allows the former relaxations to be recovered as special cases. The joint-MAP and marginal-MAP estimators are further modified to accommodate such continuous priors. In Barbault et al. (2023), we proposed an EM approach to obtain a suboptimal solution to the marginal-MAP problem, and showed some improvement in support estimation when the operator $\mathbf{H}$ is highly correlated. Moreover, in Section 4, we will explain why the marginal-MAP approach may be particularly effective for highly correlated operators $\mathbf{H}$.

3. Continuous relaxation of the Bernoulli prior

Back to the spike and slab model, we may consider a continuous model for $\mathbf{q}$. We consider the following family of continuous priors parametrized by $\rho \geq 0$

$$p(\mathbf{q}) = C(\rho,\alpha)\exp\left(-\rho\sum_{n=1}^{N} q_n^{\alpha}\right)\mathbb{1}_{\mathcal{B}}(\mathbf{q})\,, \tag{19}$$

where $\alpha > 0$, $\mathcal{B} = [0,a]^N$ with $a > 0$, or $\mathcal{B} = [0,+\infty[^N$, and $C(\rho,\alpha)$ is a normalizing constant in order to have a unit mass. When $\mathcal{B} = [0,+\infty[^N$, we set $a = +\infty$. Similar to Section 2, $\mathbf{r}$ is a Gaussian process independent of $\mathbf{q}$, with $\mathbf{r} \sim \mathcal{N}(0,\sigma_r^2 I_N)$.

Some special cases will be of particular interest. If $\alpha = 1$ and $\mathcal{B} = [0,1]^N$, one recovers the continuous Bernoulli distribution studied by Loaiza-Ganem and Cunningham (2019), while if $\mathcal{B} = [0,+\infty[^N$ we have a simple exponential distribution. If $\alpha = 2$ and $\mathcal{B} = [0,+\infty[^N$, one recovers the half-normal distribution or the truncated half-normal distribution as soon as $a < +\infty$.

3.1. Continuous joint-MAP

Similar to Section 2, joint-MAP estimation can be stated as the joint minimization problem

$$\min_{\mathbf{r}\in\mathbb{R}^N,\mathbf{q}\in\mathcal{B}} \mathcal{J}_{\alpha}(\mathbf{r},\mathbf{q};\rho)\,, \tag{20}$$

with, still by using $\mathbf{Q} = \mathrm{Diag}\,[\mathbf{q}]$,

$$\mathcal{J}_\alpha(\mathbf{r},\mathbf{q};\rho) \triangleq \frac{1}{2\sigma_e^2}\|\mathbf{y}-\mathbf{H}\mathbf{Q}\mathbf{r}\|_2^2 + \frac{1}{2\sigma_r^2}\|\mathbf{r}\|_2^2 + \rho\|\mathbf{q}\|_\alpha^\alpha \,. \tag{21}$$

As in the discrete BG case studied in Section 2, we can rewrite the joint-MAP estimator of $\mathbf{q}$ as the minimizer of

$$\mathcal{J}_\alpha\left(\mathbf{q};\rho\right) \triangleq \mathcal{J}_\alpha\left(\hat{\mathbf{r}}(\mathbf{q}),\mathbf{q};\rho\right) \,, \tag{22}$$

with $\hat{\mathbf{r}}(\mathbf{q})$ defined in (8) (by replacing $\hat{\mathbf{q}}$ and $\hat{\mathbf{Q}}$ by $\mathbf{q}$ and $\mathbf{Q}$ therein). Then, this joint-MAP criterion can also be written as:

$$\mathcal{J}_\alpha\left(\mathbf{q};\rho\right) = \frac{1}{2}\mathbf{y}^t\Gamma_y(\mathbf{q})^{-1}\mathbf{y} + \rho\|\mathbf{q}\|_\alpha^\alpha \,, \tag{23}$$

with $\Gamma_y(\mathbf{q})$ defined in (9). One can remark that, when $\alpha \to 0$, then $\mathcal{J}_\alpha\left(\mathbf{q};\rho\right) \to \mathcal{J}_0\left(\mathbf{q};\rho\right)$ with $\mathcal{J}_0$ as defined in (11).

Notice that for binary maps $\mathbf{q} \in \{0,1\}^N$, $\mathcal{J}_\alpha(\mathbf{q};\rho) = \mathcal{J}_0(\mathbf{q};\rho)$ with functional $\mathcal{J}_0$ defined in Equation (5). Moreover, when $a \geq 1$, it is clear that $\{0,1\}^N \subset \mathcal{B}$ and then:

$$\min_{\mathbf{q}\in\mathcal{B}} \mathcal{J}_\alpha(\mathbf{q};\rho) \leq \min_{\mathbf{q}\in\{0,1\}^N} \mathcal{J}_0(\mathbf{q};\rho) \,. \tag{24}$$

This lower bound was exploited for $\alpha = 2$ and $a = 1$ within the so-called perspective relaxation of the BG joint-MAP studied by Pilanci et al. (2015).

The next theorem states that the joint minimization in $(\mathbf{r},\mathbf{q})$ is equivalent to minimizing an appropriate functional in $\mathbf{x} = \mathbf{Q}\mathbf{r} \in \mathbb{R}^N$.

Theorem 3.1. *Let $\alpha > 0$. Define*

$$\mathfrak{J}_\alpha(\mathbf{x};\rho) \triangleq \frac{1}{2\sigma_e^2}\|\mathbf{y}-\mathbf{H}\mathbf{x}\|_2^2 + \frac{1}{2}\sum_{n=1}^N \phi_{\alpha,a}(x_n;\sigma_r^2;\rho)$$

with

$$\phi_{\alpha,a}(x;\sigma_r^2;\rho) = \begin{cases} \left(\frac{1}{\alpha\sigma_r^2}\right)^{\frac{\alpha}{\alpha+2}}(\alpha+2)\,\rho^{\frac{2}{\alpha+2}}|x_k|^{\frac{2\alpha}{\alpha+2}} & \text{if } |x| \leq a^{\frac{\alpha+2}{2}}\sqrt{\alpha\rho\sigma_r^2}, \\ \frac{x^2}{\sigma_r^2 a^2} + 2\rho a^\alpha & \text{if } |x| > a^{\frac{\alpha+2}{2}}\sqrt{\alpha\rho\sigma_r^2}. \end{cases}$$

Then

$$\min_{\mathbf{q}\in\mathcal{B}} \mathcal{J}_\alpha(\mathbf{q};\rho) = \min_{\mathbf{x}\in\mathbb{R}^N} \mathfrak{J}_\alpha(\mathbf{x};\rho).$$

Specifically, if $(\mathbf{r}^*,\mathbf{q}^*)$ is a global minimizer of $\mathcal{J}_\alpha$, then $\mathbf{x}^* = \mathbf{Q}^*\mathbf{r}^*$ is a global minimizer of $\tilde{\mathcal{J}}_\alpha$. Conversely, let $\mathbf{x}^*$ be a global minimizer of $\tilde{\mathcal{J}}_\alpha$. Define for all k,

$$q_k^* = \begin{cases} \left(\frac{1}{\alpha\rho\sigma_r^2}\right)^{\frac{1}{\alpha+2}}|x_k^*|^{\frac{2}{\alpha+2}} & \text{if } |x_k^*| \leq a^{\frac{\alpha+2}{2}}\sqrt{\alpha\rho\sigma_r^2}, \\ a & \text{if } |x_k^*| > a^{\frac{\alpha+2}{2}}\sqrt{\alpha\rho\sigma_r^2}, \end{cases}$$

and $r_k^* = \frac{x_k^*}{q_k^*}$ if $q_k^* \neq 0$, and 0 otherwise. Then, $(\mathbf{q}^*,\mathbf{r}^*)$ is a global minimizer of $\mathcal{J}_\alpha$.

This theorem states that the joint minimization of $\mathcal{J}_\alpha(\mathbf{r},\mathbf{q},\rho)$ can be replaced by the minimization of $\mathfrak{J}_\alpha(\mathbf{x},\rho)$. Then, one can recover the minimizers $(\mathbf{q},\mathbf{r})$ of $\mathcal{J}_\alpha$ from the minimizer $\mathbf{x}$ of $\tilde{\mathcal{J}}_\alpha$. Moreover, one can remark than $\tilde{J}_\alpha$ is convex as soon as $\alpha \geq 2$. Indeed, let us first notice that $\phi_{\alpha,a}(x;\sigma_r^2;\rho)$ is continuous at both points x such that $|x| = a^{\frac{\alpha+2}{2}}\sqrt{\alpha\rho\sigma_r^2}$, with $\phi_{\alpha,a}(x;\sigma_r^2;\rho) = a^\alpha\rho(\alpha+2)$. We can further notice that when $\alpha \geq 2$, $\frac{2\alpha}{\alpha+2} \geq 1$, so $\phi_{\alpha,a}$ is piecewise convex. At last, we check that the left and right derivatives of $\phi_{\alpha,a}$ at $|x| = a^{\frac{\alpha+2}{2}}\sqrt{\alpha\rho\sigma_r^2}$ are equal.

Some special cases are of particular interest when $\alpha = 2$, *i.e.*, for a half-Gaussian distribution on $\mathbf{q}$. For $\mathcal{B} = [0,1]^N$, we have

$$\mathfrak{J}_2(\mathbf{x};\rho) = \frac{1}{2\sigma_e^2}\|\mathbf{y}-\mathbf{H}\mathbf{x}\|_2^2 + 2\rho\,\mathrm{Rhub}\left(\frac{\mathbf{x}}{\sqrt{2\rho\sigma_r^2}}\right), \tag{25}$$

with Rhub the reverse Huber function Pilanci et al. (2015) defined componentwise by:

$$\mathrm{Rhub}(x) = \begin{cases} |x| & \text{if } |x| \le 1, \\ \frac{x^2+1}{2} & \text{if } |x| \ge 1. \end{cases}$$

Moreover, for $\mathcal{B} = [0,+\infty[^N$, we have

$$\mathfrak{J}_2(\mathbf{x};\rho) = \frac{1}{2\sigma_e^2}\|\mathbf{y}-\mathbf{H}\mathbf{x}\|_2^2 + \sqrt{\frac{2\rho}{\sigma_r^2}}\,\|\mathbf{x}\|_1. \tag{26}$$

Hence, the well-known Lasso problem can be interpreted as a "continuous" spike and slap model with a half-Gaussian prior on $\mathbf{q}$. (Jiang and Nadarajah, 2019) give the characteristic function of the product of a normally distributed random variable and an independent random variable. In the framework of continuous spike and slab model, the characteristic function of $x_n = q_n r_n$ reads:

$$\mathbb{E}\left\{e^{itx_n}\right\} = \frac{1}{\sqrt{1+\frac{\sigma_r^2}{2\rho}t^2}}, \tag{27}$$

which does not correspond to a well-known distribution, as far as we know. More specifically, $\mathbf{x}$ does not follow a Laplacian distribution (one can remark that the characteristic function is in $L^2(\mathbb{R})$ but not in $L^1(\mathbb{R})$).

Still using Jiang and Nadarajah (2019), if one chooses a Rayleigh distribution of scale parameter $\frac{1}{\epsilon} > 0$ as a prior on q_n, that is

$$p(q_n) = 2\epsilon^2 q_n e^{-\epsilon^2 q_n^2},\ q_n \ge 0\,, \tag{28}$$

then $x_n = q_n r_n$ follows a Laplacian distribution with scale parameter $\sqrt{\frac{\sigma_r^2}{2\epsilon^2}}$. While the half Gaussian and Rayleigh priors on $\mathbf{q}$ both lead to the Lasso, it is interesting to notice that a Rayleigh distribution has a mode in $\epsilon > 0$, and the half-Gaussian distribution has a mode in 0. Moreover, with the half-Gaussian prior on $\mathbf{q}$, one can recover $(\mathbf{q},\mathbf{r})$ from $\mathbf{x}$ with the closed form given in Theorem 3.1, which is not the case anymore with the Rayleigh prior on $\mathbf{q}$.

3.2. *Continuous marginal-MAP*

The marginal-MAP estimator related to the continuous spike and slab model with a prior Equation (19) reads

$$\mathcal{M}_\alpha(\mathbf{q};\rho) \triangleq \frac{1}{2}\mathbf{y}^t\Gamma_y(\mathbf{q})^{-1}\mathbf{y} + \frac{1}{2}\log|\Gamma_y(\mathbf{q})| + \rho\|\mathbf{q}\|_\alpha^\alpha\,. \tag{29}$$

Here again, as soon as $a \ge 1$, the BG-marginal-MAP criterion in Equation (14) can be bounded from below:

$$\min_{\mathbf{q}\in\mathcal{B}} \mathcal{M}_\alpha(\mathbf{q};\rho) \le \min_{\mathbf{q}\in\{0,1\}^N} \mathcal{M}(\mathbf{q};\rho)\,. \tag{30}$$

with $\mathcal{M}_\alpha(\mathbf{q};\rho) = \mathcal{M}(\mathbf{q};\rho)$ for all $\mathbf{q} \in \{0,1\}^N$. Similar to the BG model, see Equation (15), we have

$$\mathcal{M}_\alpha(\mathbf{q};\rho) = \mathcal{J}_\alpha(\mathbf{q};\rho) + \frac{1}{2}\log\left|\Gamma_y(\mathbf{q})\right| \tag{31}$$

where $\mathcal{J}_\alpha(\mathbf{q};\rho) = \mathcal{J}_\alpha(\hat{\mathbf{r}}(\mathbf{q});\mathbf{q};\rho)$ with $\hat{\mathbf{r}}(\mathbf{q})$ defined in Equation (8). The bounds of Theorem 2.1 are adapted, the upper-bound being looser.

Theorem 3.2. *Let $\mathcal{J}_\alpha(\mathbf{q};\rho)$ and $\mathcal{M}_\alpha(\mathbf{q};\rho)$ denote the continuous joint-MAP and marginal-MAP criteria, respectively. Assume that for all n, $q_n \leq a$ with $a > 0$, and let $L = \|\mathbf{H}\mathbf{H}^T\|$ be the largest eigenvalue of $\mathbf{H}\mathbf{H}^T$. Then,*

$$\mathcal{J}_\alpha(\mathbf{q};\rho) + \frac{M}{2}\log(\sigma_e^2) \leq \mathcal{M}_\alpha(\mathbf{q};\rho) \leq \mathcal{J}_\alpha(\mathbf{q};\rho) + \frac{M}{2}\log(\sigma_e^2 + a^2 L\sigma_r^2).$$

In Appendix B.2, we give the minimizers of the continuous Joint-MAP and the marginal-MAP in the case of orthogonal matrices $\mathbf{H}$ for $\alpha = 1$ and $\alpha = 2$.

The marginal-MAP approach is at the heart of the so-called Sparse Bayesian Learning (SBL) proposed by Tipping (2001) and further studied in the sparse inverse problem context by Wipf and Rao (2004). SBL aims to maximize the so-called Type-II likelihood:

$$\boldsymbol{\upsilon}^{II} = \underset{\boldsymbol{\upsilon}}{\operatorname{argmax}}\, p(\boldsymbol{\upsilon}|\mathbf{y}) = \underset{\boldsymbol{\upsilon}}{\operatorname{argmax}}\, p(\mathbf{y}|\boldsymbol{\upsilon})p(\boldsymbol{\upsilon}) = \underset{\boldsymbol{\upsilon}}{\operatorname{argmin}}\, \frac{1}{2}\mathbf{y}^t\boldsymbol{\Upsilon}_{\mathbf{y}}(\boldsymbol{\upsilon})^{-1}\mathbf{y} + \frac{1}{2}\log|\boldsymbol{\Upsilon}_{\mathbf{y}}| + R(\boldsymbol{\upsilon}) \tag{32}$$

with $\boldsymbol{\Upsilon}_{\mathbf{y}}(\boldsymbol{\upsilon}) = \sigma_e^2\mathbf{I} + \mathbf{H}\mathrm{Diag}\,[\boldsymbol{\upsilon}]\,\mathbf{H}^T$, $R(\boldsymbol{\upsilon}) = -\log p(\boldsymbol{\upsilon})$. For fixed parameter σ_r^2, if the prior on $\boldsymbol{\upsilon}$ is chosen following Equation (19), that is such that

$$p(\boldsymbol{\upsilon}) = C(\rho\sigma^\alpha, \alpha)\exp\left(-\rho\sigma^\alpha\sum_{n=1}^{N}\boldsymbol{\upsilon}^{\frac{\alpha}{2}}\right), \tag{33}$$

hence we have

$$R(\boldsymbol{\upsilon}) = -\log(p(\boldsymbol{\upsilon})) = \rho\left(\sigma_r^2\right)^{-\frac{\alpha}{2}}\|\boldsymbol{\upsilon}\|_\alpha^{\frac{\alpha}{2}} + \text{Constant}. \tag{34}$$

By conditioning on $\mathbf{q}$, he generative prior model for signal $\mathbf{x}$ reads $\mathbf{x}|\mathbf{q} \sim \mathcal{N}(\mathbf{0}, \sigma_r^2\mathbf{Q}^T\mathbf{Q})$. Then, the change of variable $\upsilon[n] = q[n]^2\sigma_r^2$ for all n in Equation (29) shows that the marginal-MAP estimator related to the continuous spike and slab model is the type-II estimator used in the SBL.

In most applications, SBL is applied without choosing any prior on $\boldsymbol{\upsilon}$ and reduces to minimizing

$$\mathcal{SBL}(\boldsymbol{\upsilon}) = \frac{1}{2}\mathbf{y}^t\boldsymbol{\Upsilon}_{\mathbf{y}}(\boldsymbol{\upsilon})^{-1}\mathbf{y} + \frac{1}{2}\log|\boldsymbol{\Upsilon}_{\mathbf{y}}(\boldsymbol{\upsilon})|\,. \tag{35}$$

It is interesting to notice that by setting $\rho = 0$ in Equation (19), the marginal-MAP criterion $\mathcal{M}_\alpha$ is lower bounded by the SBL criterion for all $\alpha > 0$:

$$\min_{\boldsymbol{\upsilon}\in\mathbb{R}_+^N}\mathcal{SBL}(\boldsymbol{\upsilon}) = \min_{\mathbf{q}\in\mathbb{R}_+^N}\mathcal{M}_\alpha(\mathbf{q};0) \leq \min_{\boldsymbol{\upsilon}\in[0,\sigma_r^2]^N}\mathcal{SBL}(\boldsymbol{\upsilon}) = \min_{\mathbf{q}\in[0,1]^N}\mathcal{M}_\alpha(\mathbf{q};0) \leq \min_{\mathbf{q}\in\{0,1\}^N}\mathcal{M}_0(\mathbf{q};0) \tag{36}$$

because the term $\rho\|\mathbf{q}\|_\alpha^\alpha$ vanishes within Equation (29). Furthermore, (Wipf and Rao, 2004) have shown that SBL leads to a sparse estimation: that is, even with no prior on $\mathbf{q}$, one still gets a sparse estimation of the support $\mathbf{q}$ using the marginal-MAP approach with $\rho = 0$.

4. Randomized rounding

Randomized rounding is an optimization technique introduced by Raghavan and Tompson (1987) that connects continuous and discrete domains. When dealing with solutions that exist in a continuous space (such as $[0,1]^N$), we can use randomized rounding to convert fractional values into discrete values within $\{0,1\}^N$. This method provides approximation guarantees and has been applied to sparse regression models that use a Bernoulli-Gaussian prior, as demonstrated by Pilanci et al. (2015). Their work provided probabilistic bounds on the global solution of an ℓ_0 constrained problem, which can be seen as a ℓ_0 constrained joint-MAP. We first extend their result to the regularized ℓ_0 problem, which corresponds to the joint-MAP explained in Section 2. We then use the randomized rounding strategy to solve the marginal-MAP problem as well.

In the following, we make the hypothesis that the matrix $\mathbf{H}$ as normalized columns.

4.1. Randomized rounding for the joint-MAP

Building upon (Pilanci et al., 2015), we adapt the randomized rounding approach to generate a Boolean solution $\tilde{\mathbf{q}} \in \{0,1\}^N$ from a fractional solution $\hat{\mathbf{q}} \in [0,1]^N$ as follows

$$\forall k,\ \mathbb{P}(\tilde{q}[k] = 1) = (\hat{q}[k])^2\ . \tag{37}$$

Hence, by construction, $\tilde{\mathbf{q}}$ is a vector of i.i.d. random variables and we have:

$$\mathbb{E}\left\{\|\tilde{\mathbf{q}}\|_0\right\} = \|\hat{\mathbf{q}}^2\|_1 \tag{38}$$

Where, with a slight abuse of notation, we denote by $\hat{\mathbf{q}}^2$ the vector $\hat{\mathbf{q}}$ where all the elements are squared. Furthermore, we observe the straightforward relationship, for all $\tilde{\mathbf{q}} \in \{0,1\}^N$, for all $\alpha > 0$

$$\mathcal{J}_\alpha(\tilde{\mathbf{q}}, \rho) = \mathcal{J}_0(\tilde{\mathbf{q}}, \rho) \tag{39}$$

where criteria $\mathcal{J}_0$ and $\mathcal{J}_\alpha$ have been defined in (11) and (22), respectively.

We first introduce the following key lemma that is essential in providing bounds for both the joint-MAP and the marginal-MAP.

Lemma 4.1. *Let $\hat{\mathbf{q}} \in [0,1]^N$. Let $\tilde{\mathbf{q}}$ be a N-sample of independent Bernoulli distributed random variables such that for all k, $\mathbb{P}(\tilde{q}[k] = 1) = (\hat{q}[k])^2$. Let $r = \#\{k \text{ s.t. } \hat{q}[k] \in (0,1)\}$. Suppose that $\alpha \leq 2$. Then, with probability at least $1 - {}^{2}/_{\min\{r,N\}^{c_1}}$, we have*

$$\mathcal{J}_0(\tilde{\mathbf{q}}, \rho) - \mathcal{J}_\alpha(\hat{\mathbf{q}}, \rho) < \sqrt{c}\frac{\sqrt{r \log \min\{r, N\}}}{\sigma_r^2}$$

where c_1 and c are constants (see Appendix A.6 for the details).

When $\alpha \leq 2$, sparse solutions are ensured. Conversely, for $\alpha > 2$, $\tilde{J}_\alpha(\mathbf{x})$ becomes differentiable everywhere, leading to non-sparse minimizers in general. Decreasing α is expected to yield sparser minimizers with fewer fractional coefficients.

This lemma allows us to directly state the following theorem, akin to the work by Pilanci et al. (2015, Theorem 3), but adapted to the regularized form of the ℓ_0 joint-MAP problem instead of its constrained variant and extended to any $\mathcal{J}_\alpha$ relaxation with $\alpha \leq 2$, minimized with respect to $\mathbf{q} \in [0,1]^N$.

Theorem 4.2. *Let $\alpha \leq 2$. Let $\mathbf{q}_\alpha^J = \operatorname{argmin} \mathcal{J}_\alpha(\mathbf{q}, \rho)$ and let $\tilde{\mathbf{q}}_\alpha^J$ be a N-sample of independant Bernoulli distributed random variables such that for all k, $\mathbb{P}(\tilde{q}_\alpha^J[k] = 1) = (q_\alpha^J[k])^2$. Let $\mathbf{q}_0^J = \operatorname{argmin} \mathcal{J}_0(\mathbf{q}, \rho)$. Then, with probability at least $1 - \frac{2}{\min\{n,r\}^{c_J}}$, we have*

$$\mathcal{J}_0(\tilde{\mathbf{q}}_\alpha^J, \rho) - \mathcal{J}_0(\mathbf{q}_0^J; \rho) < \sqrt{c}\frac{\sqrt{r^J \log \min\{r^J, N\}}}{\sigma_r^2}$$

with $r^J = \#\left\{k \text{ s.t. } q_\alpha^J[k] \in (0,1)\right\}$.

This theorem is a direct consequence of Lemma 4.1. Indeed, since $\mathbf{q}_\alpha^J$ is by definition a global minimizer of $\mathcal{J}_\alpha$ and $\mathcal{J}_\alpha$ is a continuous relaxation of $\mathcal{J}_0$ we have that $\mathcal{J}_\alpha(\mathbf{q}_\alpha^J, \rho) \leq \mathcal{J}_0(\mathbf{q}_0^J, \rho)$. Moreover, $\mathbf{q}_0^J$ being a global minimizer of $\mathcal{J}_0$, we have that $\mathcal{J}_0(\mathbf{q}_0^J, \rho) \leq \mathcal{J}_0(\tilde{\mathbf{q}}_\alpha^J, \rho)$. Hence,

$$\mathcal{J}_\alpha(\mathbf{q}_\alpha^J, \rho) \leq \mathcal{J}_0(\mathbf{q}_0^J, \rho) \leq \mathcal{J}_0(\tilde{\mathbf{q}}_\alpha^J, \rho)\ . \tag{40}$$

4.2. Randomized rounding for the marginal-MAP

We extend the results provided in Theorem 4.2 for the joint-MAP to the marginal-MAP in next Theorem. We still have the natural relationship for all $\tilde{\mathbf{q}} \in \{0,1\}^N$, for all $\alpha > 0$

$$\mathcal{M}_\alpha(\tilde{\mathbf{q}}, \lambda) = \mathcal{M}_0(\tilde{\mathbf{q}}, \lambda) \,. \tag{41}$$

Theorem 4.3. *Let $\alpha \leq 2$. Let $\mathbf{q}_\alpha^M = \operatorname{argmin} \mathcal{M}_\alpha(\mathbf{q}, \lambda)$ and let $\tilde{\mathbf{q}}_\alpha^M$ be a N-sample of independent Bernoulli distributed random variables such that $\mathbb{P}(\tilde{q}_\alpha^M[k] = 1) = (q_\alpha^M[k])^2 \forall k$. Let $\mathbf{q}_0^M = \operatorname{argmin} \mathcal{M}_0(\mathbf{q}, \lambda)$. Then, denoting by $\gamma^2 = \operatorname{var} \log |\Gamma_y(\tilde{\mathbf{q}}_\alpha^M)|$, with probability at least $\max\left\{1 - \frac{\gamma^2}{\delta}, 1 - \frac{2}{\min\{r,N\}^{cM}}\right\}$, we have*

$$\mathcal{M}_0(\tilde{\mathbf{q}}_\alpha^M, \lambda) - \mathcal{M}_0(\mathbf{q}_0^M; \lambda) < \sqrt{c}\frac{\sqrt{r^M \log \min\{r^M, N\}}}{\sigma_r^2} + \delta$$

with $r^M = \#\left\{k \text{ s.t. } q_\alpha^M[k] \in (0,1)\right\}$.

The part of the bound relative to δ comes from the Chebychev inequality (see the proof for the details), which is then inherently loose but serves as a valuable lens. The non-convex nature of $\mathcal{M}_\alpha(\mathbf{q}, \lambda)$ makes finding a global minimizer considerably more difficult compared to obtaining a minimizer of $\mathcal{J}_\alpha(\mathbf{q}, \rho)$. However, certain scenarios allow the application of randomized rounding to a minimizer of $\mathcal{J}_\alpha(\mathbf{q}, \kappa(\lambda))$ to yield an estimation of the minimizer of $\mathcal{M}_0(\mathbf{q}; \lambda)$. Notably, if

$$\mathcal{J}_\alpha(\mathbf{q}_\alpha^J, \kappa(\lambda)) < \mathcal{M}_0(\mathbf{q}_0^M; \lambda) - \frac{M}{2}\log(\sigma_e^2) \,, \tag{42}$$

then, as per Lemma 4.1, it directly follows that, with a probability at least $1 - \frac{2}{\min\{r^J,N\}^{cJ}}$

$$\mathcal{J}_0(\tilde{\mathbf{q}}_\alpha^J, \kappa(\lambda)) + \frac{M}{2}\log(\sigma_e^2) - \mathcal{M}_0(\mathbf{q}_0^M; \lambda) < \sqrt{c}\frac{\sqrt{r^J \log \min\{r^J, N\}}}{\sigma_r^2} \,. \tag{43}$$

A less stringent condition emerges as

$$\mathcal{J}_\alpha(\mathbf{q}_\alpha^J, \kappa(\lambda)) < \mathcal{M}_0(\tilde{\mathbf{q}}_\alpha^M; \lambda) - \frac{M}{2}\log(\sigma_e^2) \,. \tag{44}$$

In such instances, the following relation holds

$$\begin{aligned}\mathcal{J}_0(\tilde{\mathbf{q}}_\alpha^J, \kappa(\lambda)) - \mathcal{M}_0(\mathbf{q}_0^M; \lambda) + \frac{M}{2}\log(\sigma_e^2) &< \mathcal{J}_0(\tilde{\mathbf{q}}_\alpha^J, \kappa(\lambda)) - \mathcal{M}_0(\tilde{\mathbf{q}}_\alpha^M; \lambda) + \frac{M}{2}\log(\sigma_e^2) \quad + \mathcal{M}_0(\tilde{\mathbf{q}}_\alpha^M; \lambda) - \mathcal{M}_0(\mathbf{q}_0^M; \lambda) \\ &< \mathcal{J}_0(\tilde{\mathbf{q}}_\alpha^J, \kappa(\lambda)) - \mathcal{J}_0(\mathbf{q}_\alpha^J, \kappa(\lambda)) + \mathcal{M}_0(\tilde{\mathbf{q}}_\alpha^M; \lambda) - \mathcal{M}_0(\mathbf{q}_0^M; \lambda) \,. \end{aligned} \tag{45}$$

Thus, with probabilty at least $\max\left\{1 - \frac{2}{\min\{r^J,N\}^{cJ}}, 1 - \frac{\gamma^2}{\delta}, 1 - \frac{2}{\min\{r^M,N\}^{cM}}\right\}$, we have that

$$\mathcal{J}_0(\tilde{\mathbf{q}}_\alpha^J, \kappa(\lambda)) - \mathcal{M}_0(\mathbf{q}_0^M; \lambda) + \frac{M}{2}\log(\sigma_e^2) < \sqrt{c}\frac{\sqrt{r^J \log \min\{r^J, N\}}}{\sigma_r^2} + \sqrt{c}\frac{\sqrt{r^M \log \min\{r^M, N\}}}{\sigma_r^2} + \delta \tag{46}$$

with $r^J = \#\left\{k \text{ s.t. } q_\alpha^J[k] \in (0,1)\right\}$ and $r^M = \#\left\{k \text{ s.t. } q_\alpha^M[k] \in (0,1)\right\}$

When $\frac{1}{2}\log|\Gamma_y(\mathbf{q}_0^M)|$ closely approaches its maximum value, that is, $\log\left(\frac{\sigma_e^2 + L\sigma_r^2}{\sigma_e^2}\right)\|\mathbf{q}_0^M\|_0 + \frac{M}{2}\log(\sigma_e^2)$, – implying that $\mathbf{HQ}_0^M$ is not "excessively non-orthogonal" – the condition $\mathcal{J}_\alpha(\mathbf{q}_\alpha^J; \kappa(\lambda)) < \mathcal{M}_0(\mathbf{q}_0^M; \lambda) - \frac{M}{2}\log(\sigma_e^2)$ should be met, as we have

$$\begin{aligned}\mathcal{J}_\alpha(\mathbf{q}_\alpha^J; \kappa(\lambda)) - \mathcal{M}_0(\mathbf{q}_0^M; \lambda) + \frac{M}{2}\log(\sigma_e^2) &= \mathcal{J}_\alpha(\mathbf{q}_\alpha^J; \kappa(\lambda)) - \mathcal{J}_0(\mathbf{q}_0^M; \kappa(\lambda)) + \log\left(\frac{\sigma_e^2 + L\sigma_r^2}{\sigma_e^2}\right)\|\mathbf{q}_0^M\|_0 \\ &\quad + \frac{M}{2}\log(\sigma_e^2) - \frac{1}{2}\log|\Gamma_y(\mathbf{q}_0^M)| \end{aligned} \tag{47}$$

Hence, the uniform rounding of the minimizer of $\mathcal{J}_\alpha(\mathbf{q}, \kappa(\lambda))$ is able to provide a reliable estimation of the minimizer of the original Bernoulli-Gaussian marginal-MAP estimation of the support of the solution.

5. Conclusion

Inverse problems typically focus on signal recovery, but selecting the support of a sparse signal is also crucial and directly related to variable selection in statistics. In this context, we studied the two possible joint or a marginal-MAP criteria to estimate the support of a BG signal. We also introduced a continuous relaxation approach, where a continuous parametric model replaces the binary process that models the support. This approach shows that commonly used relaxations affect the posterior distribution of the solution.

We acknowledge the effectiveness of both the LASSO (and possibly nonconvex variants) and SBL approaches. However, we have discovered a strong correlation between both. It should be noted that the SBL method is commonly used in solving Magneto/Electroencephalography inverse problems due to the high correlation of the Lead field operator. This success can be attributed to marginal-MAP optimization, which could better identify the sources' support than joint-MAP.

By incorporating randomized rounding techniques, we efficiently minimize the convex joint-MAP with optimal hyper-parameters, facilitating robust estimation in the marginal case. However, high correlations in the matrix H can present challenges in the estimation process. Future efforts will focus on refining the bounds derived from Chebyshev's inequality and exploring more efficient algorithms to minimize the marginal-MAP, whether in its continuous or binary form. These endeavors aim to enhance the accuracy and effectiveness of support estimation and signal recovery techniques in sparse signal processing applications. Finally, Bayesian methods can address uncertainty quantification, an important problem that should be considered in future work. A measure of uncertainty on the selected support is undoubtedly a key to making these methods even more appealing in some applications like medical imaging.

Acknowledgements

This work was supported by the French National Agency for Research through the BMWs project (ANR-20-CE45-0018)

Appendix A. Proofs

Appendix A.1. Proof of Theorem 2.1

Hereafter, we denote $\mathbf{H_q} \in \mathbb{R}^{M\times\|\mathbf{q}\|_0}$ the matrix gathering the nonzero columns of $\mathbf{HQ}$, *i.e.*, the columns $\mathbf{h}_i$ of $\mathbf{H}$ for which $q_i = 1$. We have that

$$\mathcal{M}_0(\mathbf{q};\lambda) = \mathcal{J}_0(\mathbf{q};\lambda) + \frac{1}{2}\log\left|\Gamma_y(\mathbf{q})\right|$$

with $\Gamma_y(\mathbf{q})$ defined in (9).

The matrices being positive semi-definite, Minkowsky's determinant inequality implies that

$$\left|\sigma_e^2\mathbf{I}_M\right| \le \left|\sigma_e^2\mathbf{I}_M\right| + \left|\sigma_r^2\mathbf{HQQ}^T\mathbf{H}^T\right| \le \left|\Gamma_y(\mathbf{q})\right| \ , \tag{A.1}$$

then the lower bound of Theorem 2.1 is found:

$$M\log(\sigma_e^2) \le \log\left|\Gamma_y(\mathbf{q})\right| . \tag{A.2}$$

To prove the upper bound, let us denote by z_k ($k = 1,\dots,\|\mathbf{q}\|_0$) the singular values of $\mathbf{H_q}$, sorted in descending order. Clearly, $\|\mathbf{HQ}\| = \|\mathbf{H_q}\| = \max_k\{z_k\}$. According to the sub-multiplicative property of matrix norms, we have

$$\|\mathbf{HQ}\|^2 = \|\mathbf{HQQ}^T\mathbf{H}^T\| \le \|\mathbf{H}\|^2 . \|\mathbf{Q}\|^2 = \|\mathbf{HH}^T\| . \|\mathbf{Q}\|^2 \le \|\mathbf{HH}^T\| = L. \tag{A.3}$$

So for all k, $z_k \le \sqrt{L}$.

It follows from (9) that the eigenvalues of $\Gamma_y(\mathbf{q})$ are equal to $\sigma_e^2+\sigma_r^2 z_k^2$ (for $k = 1,\dots,\|\mathbf{q}\|_0$) and σ_e^2 (eigenvalue with multiplicity $\min(M,N)-\|\mathbf{q}\|_0$). The former are upper bounded by $\sigma_e^2+\sigma_r^2 L$. Defining the matrix $\mathbf{\Xi} = \mathrm{Diag}(\xi_1,\dots,\xi_M)$ such that $\xi_k = L$ for $k = 1,\dots,\|\mathbf{q}\|_0$ and $\xi_k = 0$ otherwise, we get

$$\left|\Gamma_y(\mathbf{q})\right| \le \left|\sigma_e^2\mathbf{I}_M + \sigma_r^2\mathbf{\Xi}\right| \tag{A.4}$$

and then

$$\log\left|\Gamma_y(\mathbf{q})\right| \le \log(\sigma_e^2 + L\sigma_r^2)\|\mathbf{q}\|_0 + (M-\|\mathbf{q}\|_0)\log(\sigma_e^2) = \log\left(\frac{\sigma_e^2+L\sigma_r^2}{\sigma_e^2}\right)\|\mathbf{q}\|_0 + M\log(\sigma_e^2) \ , \tag{A.5}$$

hence the upper bound.

When all the nonzero columns of $\mathbf{H_q}$ are orthogonal with the same norm $\sqrt{L}$, we directly have the equality in Equation (A.4), hence in Equation (A.5), thanks to the Sylvester determinant Theorem. Conversely, when the latter equalities hold, we have that

$$\left|\Gamma_y(\mathbf{q})\right| = (\sigma_e^2)^{M-\|\mathbf{q}\|_0}\prod_{k=1}^{\|\mathbf{q}\|_0}\left(\sigma_e^2+\sigma_r^2 z_k^2\right) = \left|\sigma_e^2\mathbf{I}_M + \sigma_r^2\mathbf{\Xi}\right| = (\sigma_e^2)^{M-\|\mathbf{q}\|_0}\prod_{k=1}^{\|\mathbf{q}\|_0}\left(\sigma_e^2+\sigma_r^2 L\right). \tag{A.6}$$

Therefore, for all k, $z_k^2 = L$. The eigenvalues of $\mathbf{H}_\mathbf{q}^T\mathbf{H_q}$ being all equal to $z_k^2 = L$, we get

$$\mathbf{H}_\mathbf{q}^T\mathbf{H_q} = L\,\mathbf{I}_{\|\mathbf{q}\|_0} \ . \tag{A.7}$$

In other words, all the non-zeros columns of $\mathbf{HQ}$ are orthogonal with the same norm $\sqrt{L}$.

Appendix A.2. Proof of Lemma 2.2

Let $\mathbf{q} \in \{0,1\}^N$ be an arbitrary support such that $\|\mathbf{q}\|_0 \le \min(M,N)$. Let $\mathbf{r} = \hat{\mathbf{r}}(\mathbf{q})$ denote the related amplitude vector where operator $\hat{\mathbf{r}}$ was defined in Equation (7). Let $\mathbf{H_q} = \mathbf{HQ} = \mathbf{UZV}^T$ with unitary matrices $\mathbf{U} \in \mathbb{R}^{M\times M}$, $\mathbf{V} \in \mathbb{R}^{N\times N}$ and a diagonal (rectangular) matrix $\mathbf{Z} \in \mathbb{R}^{M\times N}$. Expanding Equation (8), we have

$$\mathbf{r}(\mathbf{q}) = \sigma_r^2\left(\sigma_r^2\mathbf{H}_\mathbf{q}^T\mathbf{H_q} + \sigma_e^2\mathbf{I}_N\right)^{-1}\mathbf{H}_\mathbf{q}^T\mathbf{y} = \sigma_r^2\left(\sigma_r^2\mathbf{VZ}^T\mathbf{ZV}^T + \sigma_e^2\mathbf{I}_N\right)^{-1}\mathbf{VZ}^T\mathbf{U}^T\mathbf{y} \tag{A.8}$$

$$= \sigma_r^2\left(\sigma_r^2\mathbf{VZ}^T\mathbf{ZV}^T + \sigma_e^2\mathbf{I}_N\right)^{-1}\mathbf{VZ}^T\mathbf{U}^T\mathbf{y} \tag{A.9}$$

Then, $\mathcal{J}_0(\mathbf{q};\lambda)$ writes

$$\mathcal{J}_0(\mathbf{q};\lambda) = \frac{1}{2\sigma_e^2}\|\mathbf{y}-\mathbf{HQr}\|^2 + \frac{1}{2\sigma_r^2}\|\mathbf{r}\|_2^2 + \lambda\|\mathbf{q}\|_0 \tag{A.10}$$

$$= \frac{1}{2\sigma_e^2}\left\|\mathbf{y}-\mathbf{UZV}^T\sigma_r^2\left(\sigma_r^2\mathbf{VZ}^T\mathbf{ZV}^T+\sigma_e^2\mathbf{I}_N\right)^{-1}\mathbf{VZ}^T\mathbf{U}^T\mathbf{y}\right\|^2 \tag{A.11}$$

$$+ \frac{1}{2\sigma_r^2}\left\|\sigma_r^2\left(\sigma_r^2\mathbf{VZ}^T\mathbf{ZV}^T+\sigma_e^2\mathbf{I}_N\right)^{-1}\mathbf{VZ}^T\mathbf{U}^T\mathbf{y}\right\|_2^2 + \lambda\|\mathbf{q}\|_0 \tag{A.12}$$

$$= \frac{1}{2\sigma_e^2}\left\|\mathbf{U}^T\mathbf{y}-\mathbf{Z}\sigma_r^2\left(\sigma_r^2\mathbf{Z}^T\mathbf{Z}+\sigma_e^2\mathbf{I}_N\right)^{-1}\mathbf{Z}^T\mathbf{U}^T\mathbf{y}\right\|^2 + \frac{1}{2\sigma_r^2}\left\|\sigma_r^2\left(\sigma_r^2\mathbf{Z}^T\mathbf{Z}+\sigma_e^2\mathbf{I}_N\right)^{-1}\mathbf{Z}^T\mathbf{U}^T\mathbf{y}\right\|_2^2 + \lambda\|\mathbf{q}\|_0 \tag{A.13}$$

We denote by $\{z_k\}_{k=1}^{\|\mathbf{q}\|_0}$ the singular values in $\mathbf{Z}$, which are all non-zeros since $\|\mathbf{q}\|_0 \leq \min(M,N)$ and by assumption, $\operatorname{rank}(\mathbf{H}) = \min(M,N)$. Hence

$$\mathcal{J}_0(\mathbf{q};\lambda) = \sum_{k=1}^{\|\mathbf{q}\|_0}\left\{\frac{1}{2\sigma_e^2}\left(1-\sigma_r^2\left(\sigma_r^2 z_k^2+\sigma_e^2\right)^{-1}{z_k}^2\right)^2\langle\mathbf{u}_k,\mathbf{y}\rangle^2 + \frac{z_k^2\sigma_r^2}{2}\left(z_k^2\sigma_r^2+\sigma_e^2\right)^{-2}\langle\mathbf{u}_k,\mathbf{y}\rangle^2 + \lambda\right\} + \frac{1}{2\sigma_e^2}\sum_{k=\|\mathbf{q}\|_0+1}^{M}\langle\mathbf{u}_k,\mathbf{y}\rangle^2 \tag{A.14}$$

$$= \sum_{k=1}^{\|\mathbf{q}\|_0}\left\{\frac{1}{2}\frac{\sigma_e^2}{\left(z_k^2\sigma_r^2+\sigma_e^2\right)^2}\langle\mathbf{u}_k,\mathbf{y}\rangle^2 + \frac{1}{2}\frac{z_k^2\sigma_r^2}{\left(z_k^2\sigma_r^2+\sigma_e^2\right)^2}\langle\mathbf{u}_k,\mathbf{y}\rangle^2 + \lambda\right\} + \frac{1}{2\sigma_e^2}\sum_{k=\|\mathbf{q}\|_0+1}^{M}\langle\mathbf{u}_k,\mathbf{y}\rangle^2 \tag{A.15}$$

$$= \sum_{k=1}^{\|\mathbf{q}\|_0}\left\{\frac{1}{2}\left(\frac{1}{z_k^2\sigma_r^2+\sigma_e^2}\right)\langle\mathbf{u}_k,\mathbf{y}\rangle^2 + \lambda\right\} + \frac{1}{2\sigma_e^2}\sum_{k=\|\mathbf{q}\|_0+1}^{M}\langle\mathbf{u}_k,\mathbf{y}\rangle^2. \tag{A.16}$$

As $\mathbf{U}$ is unitary, we have $\|\mathbf{y}\|^2 = \|\mathbf{U}^T\mathbf{y}\|^2$, and then

$$\mathcal{J}_0(\mathbf{q};\lambda) = \sum_{k=1}^{\|\mathbf{q}\|_0}\left\{\lambda - \frac{1}{2\sigma_e^2}\left(\frac{z_k^2\sigma_r^2}{z_k^2\sigma_r^2+\sigma_e^2}\right)\langle\mathbf{u}_k,\mathbf{y}\rangle^2\right\} + \frac{1}{2\sigma_e^2}\|\mathbf{y}\|_2^2. \tag{A.17}$$

Let $\mathbf{q}^{\mathcal{J}}$ be a global minizer. Suppose that for all k,

$$2\sigma_e^2\lambda > \left(\frac{(z_k^{\mathcal{J}})^2\sigma_r^2}{(z_k^{\mathcal{J}})^2\sigma_r^2+\sigma_e^2}\right)\langle\mathbf{u}_k^{\mathcal{J}},\mathbf{y}\rangle^2. \tag{A.18}$$

Then, we have

$$\mathcal{J}_0(\mathbf{0};\lambda) = \frac{1}{2\sigma_e^2}\|\mathbf{y}\|_2^2 < \mathcal{J}_0(\mathbf{q}^{\mathcal{J}};\lambda) = \sum_{k=1}^{\|\mathbf{q}^{\mathcal{J}}\|_0}\left\{\lambda - \frac{1}{2\sigma_e^2}\left(\frac{(z_k^{\mathcal{J}}\sigma_r)^2}{(z_k^{\mathcal{J}}\sigma_r)^2+\sigma_e^2}\right)\langle\mathbf{u}_k^{\mathcal{J}},\mathbf{y}\rangle^2\right\} + \frac{1}{2\sigma_e^2}\|\mathbf{y}\|_2^2. \tag{A.19}$$

which is impossible. Then, necessarily,

$$2\sigma_e^2\lambda \leq \max_{k\in\{1,\ldots,\|\mathbf{q}^{\mathcal{J}}\|_0\}}\left\{\left(\frac{(z_k^{\mathcal{J}}\sigma_r)^2}{(z_k^{\mathcal{J}}\sigma_r)^2+\sigma_e^2}\right)\langle\mathbf{u}_k^{\mathcal{J}},\mathbf{y}\rangle^2\right\}. \tag{A.20}$$

Now, still with $\mathbf{q}^{\mathcal{J}}$ a global minizer, let $\mathbf{q}$ be any possible support. In one hand, suppose that $\|\mathbf{q}\|_0 \leq \|\mathbf{q}^{\mathcal{J}}\|_0$. Then, we have

$$\mathcal{J}_0(\mathbf{q}^{\mathcal{J}};\lambda) = \sum_{k=1}^{\|\mathbf{q}^{\mathcal{J}}\|_0}\left\{\lambda - \frac{1}{2\sigma_e^2}\left(\frac{(z_k^{\mathcal{J}}\sigma_r)^2}{(z_k^{\mathcal{J}}\sigma_r)^2+\sigma_e^2}\right)\langle\mathbf{u}_k^{\mathcal{J}},\mathbf{y}\rangle^2\right\} + \frac{1}{2\sigma_e^2}\|\mathbf{y}\|_2^2. \tag{A.21}$$

$$< \mathcal{J}_0(\mathbf{q};\lambda) = \sum_{k=1}^{\|\mathbf{q}\|_0}\left\{\lambda - \frac{1}{2\sigma_e^2}\left(\frac{(z_k\sigma_r)^2}{(z_k\sigma_r)^2+\sigma_e^2}\right)\langle\mathbf{u}_k,\mathbf{y}\rangle^2\right\} + \frac{1}{2\sigma_e^2}\|\mathbf{y}\|_2^2. \tag{A.22}$$

Hence,

$$\left(\|\mathbf{q}^{\mathcal{J}}\|_0 - \|\mathbf{q}\|_0\right) 2\sigma_e^2\lambda < \sum_{k=\|\mathbf{q}\|_0+1}^{\|\mathbf{q}^{\mathcal{J}}\|_0} \left(\frac{(z_k^{\mathcal{J}}\sigma_r)^2}{(z_k^{\mathcal{J}}\sigma_r)^2+\sigma_e^2}\right)\langle \mathbf{u}_k{}^{\mathcal{J}}, \mathbf{y}\rangle^2 + \sum_{k=1}^{\|\mathbf{q}\|_0}\left\{\left(\frac{(z_k^{\mathcal{J}}\sigma_r)^2}{(z_k^{\mathcal{J}}\sigma_r)^2+\sigma_e^2}\right)\langle \mathbf{u}_k^{\mathcal{J}}, \mathbf{y}\rangle^2 - \left(\frac{(z_k\sigma_r)^2}{(z_k\sigma_r)^2+\sigma_e^2}\right)\langle \mathbf{u}_k, \mathbf{y}\rangle^2\right\} \tag{A.23}$$

In the other hand, suppose that $\|\mathbf{q}\|_0 \geq \|\mathbf{q}^{\mathcal{J}}\|_0$. Then we have

$$\left(\|\mathbf{q}\|_0 - \|\mathbf{q}^{\mathcal{J}}\|_0\right) 2\sigma_e^2\lambda > \sum_{k=\|\mathbf{q}^{\mathcal{J}}\|_0+1}^{\|\mathbf{q}\|_0} \left(\frac{(z_k\sigma_r)^2}{(z_k\sigma_r)^2+\sigma_e^2}\right)\langle \mathbf{u}_k, \mathbf{y}\rangle^2 + \sum_{k=1}^{\|\mathbf{q}^{\mathcal{J}}\|_0}\left\{\left(\frac{(z_k^{\mathcal{J}}\sigma_r)^2}{(z_k^{\mathcal{J}}\sigma_r)^2+\sigma_e^2}\right)\langle \mathbf{u}_k^{\mathcal{J}}, \mathbf{y}\rangle^2 - \left(\frac{(z_k\sigma_r)^2}{(z_k\sigma_r)^2+\sigma_e^2}\right)\langle \mathbf{u}_k, \mathbf{y}\rangle^2\right\} \tag{A.24}$$

In the other direction, if $\mathbf{q}^{\mathcal{J}}$ satisfies Equation (A.23) for any $\mathbf{q}$ such that $\|\mathbf{q}\|_0 \leq \|\mathbf{q}^{\mathcal{J}}\|_0$ and Equation (A.24) for any $\mathbf{q}$ such that $\|\mathbf{q}\|_0 \geq \|\mathbf{q}^{\mathcal{J}}\|_0$, then we have

$$\mathcal{J}(\mathbf{q}^{\mathcal{J}};\lambda) \leq \mathcal{J}(\mathbf{q};\lambda) \tag{A.25}$$

for any $\mathbf{q}$, hence $\mathbf{q}^{\mathcal{J}}$ is a global minimizer.

Appendix A.3. Proof of Lemma 2.3

As for the previous Theorem let $\mathbf{q} \in \{0,1\}^N$ be an arbitrary support such that $\|\mathbf{q}\|_0 \leq \min(M,N)$ and let $\mathbf{HQ} = \mathbf{UZV}^T$. We have

$$\Gamma_y(\mathbf{q}) = \sigma_e^2\mathbf{I}_M + \sigma_r^2\mathbf{HQQ}^T\mathbf{H}^T = \sigma_e^2\mathbf{I}_M + \sigma_r^2\mathbf{UZZ}^T\mathbf{U}^T \tag{A.26}$$

and then

$$\left|\Gamma_y(\mathbf{q})\right| = \prod_{k=1}^{\|\mathbf{q}\|_0}((z_k\sigma_r)^2+\sigma_e^2)\prod_{k=\|\mathbf{q}\|_0+1}^{M}\sigma_e^2 = \sigma_e^{2M-\|\mathbf{q}\|_0}\prod_{k=1}^{\|\mathbf{q}\|_0}((z_k\sigma_r)^2+\sigma_e^2) \tag{A.27}$$

that is,

$$\log\left|\Gamma_y(\mathbf{q})\right| = (M-\|\mathbf{q}\|_0)\log(\sigma_e^2) + \sum_{k=1}^{\|\mathbf{q}\|_0}\log((z_k\sigma_r)^2+\sigma_e^2). \tag{A.28}$$

According to Equation (15),

$$\mathcal{M}_0(\mathbf{q};\lambda) = \mathcal{J}_0(\mathbf{q};\lambda) + \frac{1}{2}\log\left|\Gamma_y(\mathbf{q})\right| \tag{A.29}$$

then (see proof of Lemma 2.2)

$$\mathcal{M}_0(\mathbf{q};\lambda) = \mathcal{J}_0(\mathbf{q};\lambda) + (M-\|\mathbf{q}\|_0)\log\left(\sqrt{\sigma_e^2}\right) + \sum_{k=1}^{\|\mathbf{q}\|_0}\log\left(\sqrt{(z_k\sigma_r)^2+\sigma_e^2}\right) \tag{A.30}$$

$$= \sum_{k=1}^{\|\mathbf{q}\|_0}\left\{\frac{1}{2}\left(\frac{1}{(z_k\sigma_r)^2+\sigma_e^2}\right)\langle \mathbf{u}_k, \mathbf{y}_k\rangle^2 + \lambda q_k^*\right\} + \sum_{k=1}^{\|\mathbf{q}\|_0}\log\left(\sqrt{\frac{(z_k\sigma_r)^2+\sigma_e^2}{\sigma_e^2}}\right) + \frac{M}{2}\log\left(\sigma_e^2\right) + \frac{1}{2\sigma_e^2}\sum_{k=\|\mathbf{q}\|_0+1}^{M}\langle \mathbf{u}_k, \mathbf{y}_k\rangle^2. \tag{A.31}$$

$$= \sum_{k=1}^{\|\mathbf{q}\|_0}\left\{\lambda + \log\left(\sqrt{\frac{(z_k\sigma_r)^2+\sigma_e^2}{\sigma_e^2}}\right) - \frac{1}{2\sigma_e^2}\left(\frac{(z_k\sigma_r)^2}{(z_k\sigma_r)^2+\sigma_e^2}\right)\langle \mathbf{u}_k, \mathbf{y}\rangle^2\right\} + \frac{1}{2\sigma_e^2}\|\mathbf{y}\|_2^2 + \frac{M}{2}\log\left(\sigma_e^2\right). \tag{A.32}$$

The conditions follow as in the proof of Lemma 2.2.

Appendix A.4. Proof of Theorem 3.1

Let

$$\mathcal{J}_\alpha(\mathbf{r}, \mathbf{q}; \rho) \triangleq \frac{1}{2\sigma_e^2} \|\mathbf{y} - \mathbf{HQr}\|_2^2 + \frac{1}{2\sigma_r^2} \|\mathbf{r}\|^2 + \rho \|\mathbf{q}\|_\alpha^\alpha \tag{A.33}$$

When $q_n = 0$, the criterion reaches its minimum value when $r_n = 0$.

Let $\mathbf{x} = \mathbf{Qr}$ in Equation (A.33). Using the reparametrization $(\mathbf{x}, \mathbf{q})$, minimizing $\mathcal{J}_\alpha(\mathbf{r}, \mathbf{q}; \rho)$ is equivalent to minimizing:

$$\mathcal{K}_\alpha(\mathbf{x}, \mathbf{q}) \triangleq \frac{1}{2\sigma_e^2} \|\mathbf{y} - \mathbf{Hx}\|_2^2 + \frac{1}{2\sigma_r^2} \sum_{k=1}^{N} \frac{x_k^2}{q_k^2} + \rho \|\mathbf{q}\|_\alpha^\alpha. \tag{A.34}$$

where we use the abuse of notation $\frac{x_k^2}{q_k^2} = 0$ when $x_k = 0$ and $q_k = 0$. Indeed, we consider that $r_n = x_n/q_n$ for all q_n and choose to define it as zero when q_n vanishes.

Minimizing (A.34) with respect to $\mathbf{q}$ leads to finding

$$\hat{q}_k(x_k) = \underset{q_k \in [0,a]}{\operatorname{argmin}} \frac{1}{2\sigma_r^2} \left(\frac{x_k}{q_k}\right)^2 + \rho q_k^\alpha. \tag{A.35}$$

When $x_k \neq 0$, as $\rho > 0$, one can easily check that the function $q_k \mapsto \frac{1}{2\sigma_r^2} \frac{x_k^2}{q_k^2} + \rho q_k^\alpha$ is decreasing and then increasing on $[0, \infty)$. Its minimum is reached when the first order derivative vanishes, that is, when

$$-\frac{x_k^2}{\sigma_r^2} q_k^{-3} + \rho\alpha q_k^{\alpha-1} = 0, \tag{A.36}$$

that is, when

$$q_k^{\alpha+2} = \frac{x_k^2}{\alpha\rho\sigma_r^2}. \tag{A.37}$$

When minimization over q_k is carried out for $\hat{q}_k(x_k) \in [0, a]$, the solution is given by (A.37) when

$$a^{\alpha+2} \geq \frac{x_k^2}{\alpha\rho\sigma_r^2} \quad \Leftrightarrow \quad |x_k| \leq a^{\frac{\alpha+2}{2}} \sqrt{\alpha\rho\sigma_r^2} \tag{A.38}$$

When this condition is not met, the minimizer is $\hat{q}_k(x_k) = a$. Summarizing,

$$\hat{q}_k(x_k) = \begin{cases} \left(\frac{1}{\alpha\rho\sigma_r^2}\right)^{\frac{1}{\alpha+2}} |x_k|^{\frac{2}{\alpha+2}} & \text{if } |x_k| \leq a^{\frac{\alpha+2}{2}} \sqrt{\alpha\rho\sigma_r^2} \\ a & \text{if } |x_k| > a^{\frac{\alpha+2}{2}} \sqrt{\alpha\rho\sigma_r^2}. \end{cases} \tag{A.39}$$

The minimum value of the criterion in Equation (A.35) is thus equal to

$$\frac{1}{2\sigma_r^2} \left(\frac{x_k}{\hat{q}_k}\right)^2 + \rho \hat{q}_k^\alpha = \frac{x_k^2}{2\sigma_r^2 a^2} + \rho a^\alpha \tag{A.40}$$

when $|x_k| > a^{\frac{\alpha+2}{2}}\sqrt{\alpha\rho\sigma_r^2}$. When $|x_k| \leq a^{\frac{\alpha+2}{2}}\sqrt{\alpha\rho\sigma_r^2}$, the minimum value reads:

$$\frac{1}{2\sigma_r^2}\left(\frac{x_k}{\hat{q}_k}\right)^2 + \rho\hat{q}_k^\alpha = \frac{1}{2\sigma_r^2}\left(\frac{x_k}{\left(\frac{1}{\alpha\rho\sigma_r^2}\right)^{\frac{1}{\alpha+2}}|x_k|^{\frac{2}{\alpha+2}}}\right)^2 + \rho\left(\left(\frac{1}{\alpha\rho\sigma_r^2}\right)^{\frac{1}{\alpha+2}}|x_k|^{\frac{2}{\alpha+2}}\right)^\alpha \tag{A.41}$$

$$= \frac{1}{2\sigma_r^2}\frac{|x_k|^{2-\frac{4}{\alpha+2}}}{\left(\frac{1}{\alpha\rho\sigma_r^2}\right)^{\frac{2}{\alpha+2}}} + \rho\left(\frac{1}{\alpha\rho\sigma_r^2}\right)^{\frac{\alpha}{\alpha+2}}|x_k|^{\frac{2\alpha}{\alpha+2}} \tag{A.42}$$

$$= \left(\frac{1}{2\sigma_r^2}\left(\alpha\rho\sigma_r^2\right)^{\frac{2}{\alpha+2}} + \rho\left(\alpha\rho\sigma_r^2\right)^{-\frac{\alpha}{\alpha+2}}\right)|x_k|^{\frac{2\alpha}{\alpha+2}} \tag{A.43}$$

$$= \left(\frac{1}{2}(\alpha\rho)^{\frac{2}{\alpha+2}} + \rho(\alpha\rho)^{-\frac{\alpha}{\alpha+2}}\right)(\sigma_r^2)^{-\frac{\alpha}{\alpha+2}}|x_k|^{\frac{2\alpha}{\alpha+2}} \tag{A.44}$$

$$= \left(\frac{1}{2}\alpha^{\frac{2}{\alpha+2}} + \alpha^{-\frac{\alpha}{\alpha+2}}\right)\rho^{\frac{2}{\alpha+2}}(\sigma_r^2)^{-\frac{\alpha}{\alpha+2}}|x_k|^{\frac{2\alpha}{\alpha+2}} \tag{A.45}$$

$$= \frac{1}{2}\alpha^{-\frac{\alpha}{\alpha+2}}(\alpha+2)\rho^{\frac{2}{\alpha+2}}(\sigma_r^2)^{-\frac{\alpha}{\alpha+2}}|x_k|^{\frac{2\alpha}{\alpha+2}} \tag{A.46}$$

$$= \frac{1}{2}\left(\frac{1}{\alpha\sigma_r^2}\right)^{\frac{\alpha}{\alpha+2}}(\alpha+2)\rho^{\frac{2}{\alpha+2}}|x_k|^{\frac{2\alpha}{\alpha+2}} \tag{A.47}$$

Let us denote by $\mathcal{X}$ the set of indices $k \in [1;N]$ such that $k \in \mathcal{X}$ if x_k satisfies Equation (A.38). Re-injecting the latter expression into (A.34) yields:

$$\mathcal{K}_\alpha(\mathbf{x}, \hat{\mathbf{q}}(\mathbf{x})) = \frac{1}{2\sigma_e^2}\|\mathbf{y}-\mathbf{H}\mathbf{x}\|_2^2 + \frac{1}{2}\left(\frac{1}{\alpha\sigma_r^2}\right)^{\frac{\alpha}{\alpha+2}}(\alpha+2)\rho^{\frac{2}{\alpha+2}}\sum_{k\in\mathcal{X}}|x_k|^{\frac{2\alpha}{\alpha+2}} + \frac{1}{2}\sum_{k\notin\mathcal{X}}\left(\frac{x_k^2}{\sigma_r^2a^2} + 2\rho a^\alpha\right) \tag{A.48}$$

which identifies with

$$\tilde{\mathcal{J}}_\alpha(\mathbf{x}) \triangleq \frac{1}{2\sigma_e^2}\|\mathbf{y}-\mathbf{H}\mathbf{x}\|_2^2 + \frac{1}{2}\sum_{k=1}^N \phi_{\alpha,a}(x_k;\sigma_r^2;\rho) \tag{A.49}$$

with function $\phi_{\alpha,a}$ defined in the statement of Theorem 3.1.

If $a = +\infty$, the result simplifies to

$$\tilde{\mathcal{J}}_\alpha(\mathbf{x}) = \frac{1}{2\sigma_e^2}\|\mathbf{y}-\mathbf{H}\mathbf{x}\|_2^2 + \frac{1}{2}\left(\frac{1}{\alpha\rho\sigma_r^2}\right)^{\frac{\alpha}{\alpha+2}}(\alpha+2)\rho\sum_{k=1}^N|x_k|^{\frac{2\alpha}{\alpha+2}}. \tag{A.50}$$

Specifically, for $\alpha = 2$, we get on $[0,+\infty[$:

$$\tilde{\mathcal{J}}_\alpha(\mathbf{x}) = \frac{1}{2\sigma_e^2}\|\mathbf{y}-\mathbf{H}\mathbf{x}\|_2^2 + \sqrt{\frac{2\rho}{\sigma_r^2}}\|\mathbf{x}\|_1, \tag{A.51}$$

and on $[0,1]$, that is, for $a = 1$:

$$\tilde{\mathcal{J}}_\alpha(\mathbf{x}) = \frac{1}{2\sigma_e^2}\|\mathbf{y}-\mathbf{H}\mathbf{x}\|_2^2 + \sqrt{\frac{2\rho}{\sigma_r^2}}\sum_{k=1;|x_k|\leq\sqrt{2\rho\sigma_r^2}}^N|x_k| + \frac{1}{2}\sum_{k=1;|x_k|>\sqrt{2\rho\sigma_r^2}}^N\left(\frac{1}{\sigma_r^2}x_k^2 + 2\rho\right) \tag{A.52}$$

$$= \frac{1}{2\sigma_e^2}\|\mathbf{y}-\mathbf{H}\mathbf{x}\|_2^2 + 2\rho\mathrm{Rhub}\left(\frac{1}{\sqrt{2\rho\sigma_r^2}}\mathbf{x}\right) \tag{A.53}$$

A graphical illustration of $\phi_{\alpha,1}(x)$ is provided in Fig. A.1.

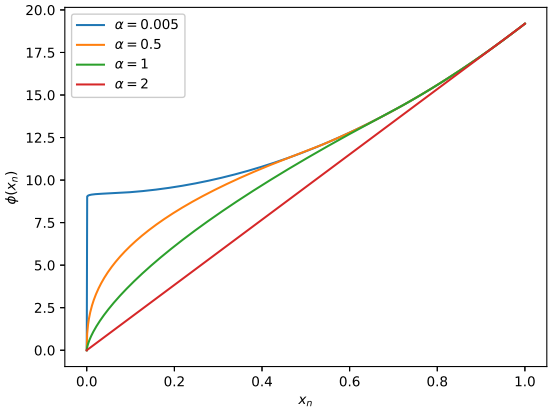

Figure A.1: Regularization functions for different α and $a = 1$.

Appendix A.5. Proof of Theorem 3.2

The proof is similar to that of Theorem 2.1. By assumption, we have that for all k, $q_k < a$, then $\|Q^T Q\| \leq a^2$. Hence, by taking $\boldsymbol{\Xi} = \text{Diag}\,[\xi_1, \ldots, \xi_n]$ the diagonal matrix such that $\xi_k = a^2 L$ if $z_k \neq 0$ and $\xi_k = 0$ if $z_k = 0$, we only need to majorize $\|\mathbf{q}\|_0$ by N in Equation (A.5).

Appendix A.6. Proof of Lemma 4.1

Noticing that for $0 < \alpha \leq 2$

$$\|\hat{\mathbf{q}}\|_\alpha^\alpha = \|\hat{\mathbf{q}}^2\|_{\frac{\alpha}{2}}^{\frac{\alpha}{2}} \geq \|(\hat{\mathbf{q}})^2\|_1 = \mathbb{E}\left\{\|\tilde{\mathbf{q}}\|_0\right\}, \tag{A.54}$$

we have

$$\mathcal{J}_0(\tilde{\mathbf{q}}, \rho) - \mathcal{J}_\alpha(\hat{\mathbf{q}}, \rho) = \mathbf{y}^T \left(\Gamma_y^{-1}(\tilde{\mathbf{q}}) - \Gamma_y^{-1}(\hat{\mathbf{q}})\right)\mathbf{y} + \rho\left(\|\tilde{\mathbf{q}}\|_0 - \|\hat{\mathbf{q}}\|_\alpha^\alpha\right) \tag{A.55}$$

$$\leq \mathbf{y}^T \left(\Gamma_y^{-1}(\tilde{\mathbf{q}}) - \Gamma_y^{-1}(\hat{\mathbf{q}})\right)\mathbf{y} + \rho\left(\|\tilde{\mathbf{q}}\|_0 - \|\hat{\mathbf{q}}^2\|_1\right) \tag{A.56}$$

On the one hand, since the columns of $\mathbf{H}$ are normalized, Pilanci et al. (2015) have shown (see the proof of (Pilanci et al., 2015, Theorem 3)), that

$$\mathbb{P}\left(\mathbf{y}^T \left(\Gamma_y^{-1}(\tilde{\mathbf{q}}) - \Gamma_y^{-1}(\hat{\mathbf{q}})\right)\mathbf{y} \leq \frac{\sqrt{rt}}{\sigma_r^2}\right) \geq 1 - 2\min\{r, N\} e^{-\frac{t^2}{16}}, \tag{A.57}$$

with $r = \#\{k \text{ s.t. } \hat{q}[k] \in (0, 1)\}$.

And in the other hand, using the Hoeffding inequality

$$\mathbb{P}\left(\left|\|\tilde{\mathbf{q}}\|_0 - \|\hat{\mathbf{q}}^2\|_1\right| \leq \delta\right) \geq 1 - 2e^{-\frac{2\delta^2}{N}}. \tag{A.58}$$

Taking $t = \sqrt{c \log \min\{r, N\}}$ we have

$$1 - 2\min\{r, N\} e^{-\frac{t^2}{16}} = 2\min\{r, N\} e^{-\frac{c \log \min\{r,N\}}{16}} = 1 - 2\min\{r, N\} e^{\log(\min\{r,N\}^{-\frac{c}{16}})} \tag{A.59}$$

$$= 1 - \frac{2}{\min\{r, N\}^{\frac{c}{16}-1}} \tag{A.60}$$

and in the other side, taking $\delta = \frac{\sqrt{r}t}{\sigma_r^2} = \sqrt{c\frac{r}{\sigma_r^2}\log\min\{r,N\}}$, we have

$$1 - 2e^{-\frac{2\delta^2}{N}} = 1 - 2e^{-2c\frac{r}{N\sigma_r^2}\log\min\{r,N\}} \tag{A.61}$$

$$= 1 - \frac{2}{\min\{r,N\}^{\frac{2cr}{N\sigma_r^2}}} \tag{A.62}$$

Then, with probability at least $\max\left\{1 - \frac{2}{\min\{r,N\}^{\frac{c}{16}-1}}, 1 - \frac{2}{\min\{r,N\}^{\frac{2cr}{N\sigma_r^2}}}\right\}$, we can conclude that

$$\mathcal{J}_0(\tilde{\mathbf{q}},\rho) - \mathcal{J}_\alpha(\hat{\mathbf{q}},\rho) \geq \sqrt{c}\frac{\sqrt{r\log\min\{r,N\}}}{\sigma_r^2}\,. \tag{A.63}$$

Appendix A.7. Proof of Theorem 4.3

We have

$$\mathcal{M}_0(\tilde{\mathbf{q}}^M,\lambda) - \mathcal{M}_0(\mathbf{q}^M,\lambda) \leq \mathcal{M}_0(\tilde{\mathbf{q}}^M,\lambda) - \mathcal{M}_\alpha(\mathbf{q}_\alpha^M,\lambda) \tag{A.64}$$

$$\leq \mathcal{J}_0(\tilde{\mathbf{q}}^M,\lambda) - \mathcal{J}_\alpha(\mathbf{q}_\alpha^M,\lambda) + \frac{1}{2}\log|\Gamma_y(\tilde{\mathbf{q}}^M)| - \frac{1}{2}\log|\Gamma_y(\mathbf{q}_\alpha^M)|\,. \tag{A.65}$$

As $\tilde{\mathbf{q}}^M[k] \in \{0,1\}^N$, we get

$$\mathbb{E}\left\{\Gamma_y(\tilde{\mathbf{q}}^M)\right\} = \mathbb{E}\left\{\sigma_e^2\mathbf{I} + \sigma_r^2\mathbf{H}\mathrm{Diag}\left[(\tilde{\mathbf{q}}^M)^2\right]\mathbf{H}^T\right\} = \sigma_e^2\mathbf{I} + \sigma_r^2\mathbf{H}\mathrm{Diag}\left[\mathbb{E}\left\{\tilde{\mathbf{q}}^M\right\}\right]\mathbf{H}^T \tag{A.66}$$

$$= \sigma_e^2\mathbf{I} + \sigma_r^2\mathbf{H}\mathrm{Diag}\left[(\mathbf{q}_\alpha^M)^2\right]\mathbf{H}^T \tag{A.67}$$

$$= \Gamma_y(\mathbf{q}_\alpha^M). \tag{A.68}$$

By Jensen's inequality (the logdet function being concave), we get

$$\mathbb{E}\left\{\log|\Gamma_y(\tilde{\mathbf{q}}^M)|\right\} \leq \log|\mathbb{E}\left\{\Gamma_y(\tilde{\mathbf{q}}^M)\right\}| = \log|\Gamma_y(\mathbf{q}_\alpha^M)|\,, \tag{A.69}$$

hence

$$\mathcal{M}_0(\tilde{\mathbf{q}}^M,\lambda) - \mathcal{M}_0(\mathbf{q}^M,\lambda) \leq \mathcal{J}_0(\tilde{\mathbf{q}}^M,\lambda) - \mathcal{J}_\alpha(\mathbf{q}_\alpha^M,\lambda) + \frac{1}{2}\log|\Gamma_y(\tilde{\mathbf{q}}^M)| - \frac{1}{2}\mathbb{E}\left\{\log|\Gamma_y(\tilde{\mathbf{q}}^M)|\right\}\,. \tag{A.70}$$

In one hand, as $|\Gamma_y(\tilde{\mathbf{q}}^M)|$ is upper and lower bounded by positive constants, we can apply the Bienaymé-Tchebychev inequality. Hence

$$\mathbb{P}\left(\left|\log|\Gamma_y(\tilde{\mathbf{q}}^M)| - \mathbb{E}\left\{\log|\Gamma_y(\tilde{\mathbf{q}}^M)|\right\}\right| \leq \delta\right) \geq 1 - \frac{\gamma^2}{\delta} \tag{A.71}$$

with $\gamma^2 = \mathrm{var}\log|\Gamma_y(\tilde{\mathbf{q}}^M)| < +\infty$.

On the other hand, using Lemma 4.1, we have with probability at least $1 - \frac{2}{\min\{r^M,N\}^{c_M}}$

$$\mathcal{J}_0(\tilde{\mathbf{q}}^M,\lambda) - \mathcal{J}_\alpha(\mathbf{q}_\alpha^M,\lambda) \leq \sqrt{c}\frac{\sqrt{r\log\min\{r^M,N\}}}{\sigma_r^2} \tag{A.72}$$

Finally, with probability at least $\max\left\{1 - \frac{\gamma^2}{\delta}, 1 - \frac{2}{\min\{r^M,N\}^{c_M}}\right\}$

$$\mathcal{M}_0(\tilde{\mathbf{q}}^M,\lambda) - \mathcal{M}_0(\mathbf{q}^M,\lambda) \leq \sqrt{c}\frac{\sqrt{r\log\min\{r^M,N\}}}{\sigma_r^2} + \delta \tag{A.73}$$

Appendix B. Joint and marginal-MAP in the orthogonal case

The case where matrix $\mathbf{H}$ is orthogonal is equivalent to a denoising problem (up to the replacement of $\mathbf{y}$ by $\mathbf{H}^t\mathbf{y}$). Therefore, we address the denoising problem which reads (with $N = M$):

$$y_n = q_n r_n + e_n, \quad \forall n \in \{1, \dots, N\}. \tag{B.1}$$

Appendix B.1. BG model

Joint-MAP. The joint-MAP criterion reads:

$$\mathcal{J}_0(\mathbf{q}, \mathbf{r}; \lambda) = \sum_{n=1}^{N} \left\{ \frac{1}{2\sigma_e^2}(y_n - q_n r_n)^2 + \frac{1}{2\sigma_r^2} r_n^2 + \lambda q_n \right\} \tag{B.2}$$

$$\triangleq \sum_{n=1}^{N} \mathcal{J}_0(q_n, r_n; \lambda). \tag{B.3}$$

Minimizing $\mathcal{J}_0(q_n, r_n; \lambda)$ with respect to (q_n, r_n) implies that $q_n = 0 \Rightarrow \hat{r}_n = 0$. For a given $q_n = 1$, minimizing in r_n yields

$$\hat{r}_n(q_n) = \frac{\sigma_r^2}{\sigma_r^2 + \sigma_e^2} y_n \,. \tag{B.4}$$

If $q_n = 0$, then

$$\mathcal{J}_0(0, \hat{r}_n(q_n); \lambda) = \frac{y_n^2}{2\sigma_e^2} \,. \tag{B.5}$$

If $q_n = 1$ then

$$\mathcal{J}_0(1, \hat{r}_n(q_n); \lambda) = \frac{1}{2\sigma_e^2}\left(y_n - \frac{\sigma_r^2}{\sigma_r^2 + \sigma_e^2}\, y_n\right)^2 + \frac{1}{2\sigma_r^2}\left(\frac{\sigma_r^2}{\sigma_r^2 + \sigma_e^2}\, y_n\right)^2 + \lambda \tag{B.6}$$

$$= \frac{1}{2}\, \frac{1}{\sigma_r^2 + \sigma_e^2}\, y_n^2 + \lambda. \tag{B.7}$$

We have $q_n = 1$ if and only if

$$\mathcal{J}_0(1, \hat{r}_n(1); \lambda) < \mathcal{J}_0(0, \hat{r}_n(0); \lambda), \tag{B.8}$$

that is, if

$$\frac{1}{2}\, \frac{1}{\sigma_r^2 + \sigma_e^2}\, y_n^2 + \lambda < \frac{1}{2\sigma_e^2}\, y_n^2 \tag{B.9}$$

i.e.,

$$\lambda < \frac{1}{2\sigma_e^2}\, \frac{\sigma_r^2}{\sigma_r^2 + \sigma_e^2}\, y_n^2. \tag{B.10}$$

Marginal-MAP. The marginal-MAP criterion reads

$$\mathcal{M}_0(\mathbf{q}; \lambda) = \frac{1}{2}\mathbf{y}^T \Gamma_y(\mathbf{q})^{-1}\mathbf{y} + \frac{1}{2}\log|\Gamma_y(\mathbf{q})| + \lambda\|\mathbf{q}\|_0 \tag{B.11}$$

with

$$\Gamma_y(\mathbf{q}) = \sigma_e^2\mathbf{I} + \sigma_r^2\mathbf{Q}\mathbf{Q}^T = \mathrm{Diag}\left[\sigma_e^2 + \sigma_r^2 q_n^2\right] \tag{B.12}$$

as $q_n \in \{0, 1\}$. Hence,

$$\mathcal{M}_0(\mathbf{q}; \lambda) = \sum_{n=1}^{N} \left\{ \frac{1}{2} \frac{y_n^2}{\sigma_e^2 + \sigma_r^2 q_n} + \frac{1}{2} \log(\sigma_e^2 + \sigma_r^2 q_n) + \lambda q_n \right\} \tag{B.13}$$

$$\triangleq \sum_{n=1}^{N} \mathcal{M}_0(q_n; \lambda). \tag{B.14}$$

If $q_n = 0$, then

$$\mathcal{M}_0(q_n; \lambda) = \frac{1}{2\sigma_e^2} y_n^2 + \frac{1}{2} \log(\sigma_e^2). \tag{B.15}$$

If $q_n = 1$, then

$$\mathcal{M}_0(q_n; \lambda) = \frac{1}{2} \frac{1}{\sigma_r^2 + \sigma_e^2} y_n^2 + \lambda + \frac{1}{2} \log\left(\sigma_r^2 + \sigma_e^2\right). \tag{B.16}$$

Hence, $q_n = 1$ if and only if

$$\mathcal{M}_0(1; \lambda) < \mathcal{M}_0(0; \lambda) \tag{B.17}$$

that is

$$\frac{1}{2} \frac{1}{\sigma_r^2 + \sigma_e^2} y_n^2 + \lambda + \log\left(\sqrt{\frac{\sigma_r^2 + \sigma_e^2}{\sigma_e^2}} \right) < \frac{1}{2\sigma_e^2} y_n^2 \tag{B.18}$$

i.e.,

$$\lambda + \log\left(\sqrt{\frac{\sigma_r^2 + \sigma_e^2}{\sigma_e^2}} \right) < \frac{1}{2\sigma_e^2} \frac{\sigma_r^2}{\sigma_e^2 + \sigma_r^2} y_n^2. \tag{B.19}$$

Appendix B.2. Continuous model

Joint-MAP. The joint-MAP criterion reads, with $\alpha > 0$:

$$\mathcal{J}_\alpha(\mathbf{q}, \mathbf{r}; \rho) = \sum_{n=1}^{N} \left\{ \frac{1}{2\sigma_e^2} (y_n - q_n r_n)^2 + \frac{1}{2\sigma_r^2} r_n^2 + \rho q_n^\alpha \right\} \tag{B.20}$$

$$\triangleq \sum_{n=1}^{N} \mathcal{J}_\alpha(q_n, r_n; \rho) \tag{B.21}$$

For fixed q_n, minimization of $\mathcal{J}_\alpha(q_n, r_n; \rho)$ with respect to r_n is an unconstrained least squares problem (since the dependency of $\mathcal{J}_\alpha$ upon r_n is quadratic). Therefore, the optimal parameter $\hat{r}_n(q_n)$ satisfies:

$$\frac{\partial}{\partial r_n} \mathcal{J}_\alpha(q_n, \hat{r}_n; \rho) = -\frac{q_n}{\sigma_e^2} (y_n - q_n \hat{r}_n) + \frac{1}{\sigma_r^2} \hat{r}_n = 0, \tag{B.22}$$

which yields

$$\hat{r}_n = \frac{\sigma_r^2 q_n y_n}{\sigma_e^2 + \sigma_r^2 q_n^2}. \tag{B.23}$$

Now, let us find the minimum of $\mathcal{J}_\alpha(q_n, \hat{r}_n(q_n); \rho)$. We have:

$$\mathcal{J}_\alpha(q_n, \hat{r}_n(q_n); \rho) = \frac{1}{2\sigma_e^2}\left(y_n - q_n \frac{\sigma_r^2 q_n y_n}{\sigma_e^2 + \sigma_r^2 q_n^2}\right)^2 + \frac{1}{2\sigma_r^2}\left(\frac{\sigma_r^2 q_n y_n}{\sigma_e^2 + \sigma_r^2 q_n^2}\right)^2 + \rho q_n^\alpha \tag{B.24}$$

$$= \frac{y_n^2}{2\sigma_e^2}\left(\frac{\sigma_e^2}{\sigma_e^2 + \sigma_r^2 q_n^2}\right)^2 + \frac{y_n^2}{2\sigma_r^2}\left(\frac{\sigma_r^2 q_n}{\sigma_e^2 + \sigma_r^2 q_n^2}\right)^2 + \rho q_n^\alpha \tag{B.25}$$

$$= \frac{\sigma_e^2 + q_n^2 \sigma_r^2}{(\sigma_e^2 + \sigma_r^2 q_n^2)^2} \frac{y_n^2}{2} + \rho q_n^\alpha \tag{B.26}$$

$$= \frac{1}{\sigma_e^2 + \sigma_r^2 q_n^2} \frac{y_n^2}{2} + \rho q_n^\alpha. \tag{B.27}$$

Hereafter, we assume that the minimization with respect to q_n is done for $q_n \in [0, +\infty[$ without loss of generality (the extension to the case $q_n \in [0, a]$ with $a < \infty$ is straightforward). We calculate:

$$\frac{\partial}{\partial q_n} \mathcal{J}_\alpha(q_n, \hat{r}_n(q_n); \rho) = -\sigma_r^2 y_n^2 q_n (\sigma_e^2 + \sigma_r^2 q_n^2)^{-2} + \alpha\rho q_n^{\alpha-1}. \tag{B.28}$$

Given that $\rho > 0$, this derivative vanishes when:

$$q_n^{\alpha-1}(\sigma_e^2 + \sigma_r^2 q_n^2)^2 - \frac{\sigma_r^2}{\alpha\rho} y_n^2 q_n = 0, \tag{B.29}$$

$$\Leftrightarrow \sigma_r^4 q_n^{\alpha+3} + 2\sigma_e^2 \sigma_r^2 q_n^{\alpha+1} + \sigma_e^4 q_n^{\alpha-1} - \frac{\sigma_r^2}{\alpha\rho} y_n^2 q_n = 0, \tag{B.30}$$

$$\Leftrightarrow q_n^{\alpha+3} + 2\frac{\sigma_e^2}{\sigma_r^2} q_n^{\alpha+1} + \left(\frac{\sigma_e^2}{\sigma_r^2}\right)^2 q_n^{\alpha-1} - \frac{1}{\alpha\rho} \frac{y_n^2}{\sigma_r^2} q_n = 0. \tag{B.31}$$

If $\alpha = 1$, we get:

$$q_n^4 + 2\frac{\sigma_e^2}{\sigma_r^2} q_n^2 - \frac{1}{\rho} \frac{y_n^2}{\sigma_r^2} q_n + \left(\frac{\sigma_e^2}{\sigma_r^2}\right)^2 = 0\,. \tag{B.32}$$

If $\alpha = 2$, we get:

$$q_n^5 + 2\frac{\sigma_e^2}{\sigma_r^2} q_n^3 + \left(\frac{\sigma_e^2}{\sigma_r^2}\right)^2 q_n - \frac{1}{2\rho} \frac{y_n^2}{\sigma_r^2} q_n = q_n \left(q_n^4 + 2\frac{\sigma_e^2}{\sigma_r^2} q_n^2 + \left(\frac{\sigma_e^2}{\sigma_r^2}\right)^2 - \frac{1}{2\rho} \frac{y_n^2}{\sigma_r^2}\right) = 0. \tag{B.33}$$

So we have either $q_n = 0$ or

$$q_n^2 = \frac{1}{\sqrt{2\rho\sigma_r^2}} |y_n| - \frac{\sigma_e^2}{\sigma_r^2} \tag{B.34}$$

as soon as $|y_n| > \sqrt{\frac{2\rho}{\sigma_r^2}}\, \sigma_e^2$ (to ensure positivity of the latter expression). From the sign of the derivative of $\mathcal{J}_\alpha(q_n, \hat{r}_n(q_n); \rho)$, we have that $\mathcal{J}_\alpha$ for $\alpha = 1$ or $\alpha = 2$ is decreasing for $q_n^2 < \frac{1}{\sqrt{2\rho\sigma_r^2}} |y_n| - \frac{\sigma_e^2}{\sigma_r^2}$ and increasing for $q_n^2 > \frac{1}{\sqrt{2\rho\sigma_r^2}} |y_n| - \frac{\sigma_e^2}{\sigma_r^2}$. Thus, the minimizer $\hat{q}_n$ is given by Equation (B.34).

When $|y_n| \leq \sqrt{\frac{2\rho}{\sigma_r^2}}\, \sigma_e^2$, $q_n \mapsto \mathcal{J}_\alpha(q_n, \hat{r}_n(q_n); \rho)$ with $\alpha = 1$ or $\alpha = 2$ is non-decreasing over $\mathbb{R}_+$ because its derivative is non-negative. Overall, we get

$$\hat{q}_n^2 = \begin{cases} \sqrt{\frac{1}{2\rho\sigma_r^2}}\, |y_n| - \frac{\sigma_e^2}{\sigma_r^2} & \text{if } y_n^2 > 2\rho\sigma_r^2 \left(\sigma_e^2/\sigma_r^2\right)^2, \\ 0 & \text{if } y_n^2 \leq 2\rho\sigma_r^2 \left(\sigma_e^2/\sigma_r^2\right)^2. \end{cases} \tag{B.35}$$

Marginal-MAP. The marginal-MAP criterion reads

$$\mathcal{M}_\alpha(\mathbf{q};\rho) = \frac{1}{2}\mathbf{y}^T\Gamma_y(\mathbf{q})^{-1}\mathbf{y} + \frac{1}{2}\log|\Gamma_y(\mathbf{q})| + \rho\|\mathbf{q}\|_\alpha^\alpha \tag{B.36}$$

with

$$\Gamma_y(\mathbf{q}) = \sigma_e^2\mathbf{I} + \sigma_r^2\mathbf{Q}\mathbf{Q}^T = \mathrm{Diag}\left[\sigma_e^2 + \sigma_r^2 q_n^2\right] . \tag{B.37}$$

Hence,

$$\mathcal{M}_\alpha(\mathbf{q};\rho) = \sum_{n=1}^N \left\{ \frac{1}{2}\frac{y_n^2}{\sigma_e^2+\sigma_r^2 q_n^2} + \frac{1}{2}\log(\sigma_e^2+\sigma_r^2 q_n^2) + \rho q_n^\alpha \right\} \tag{B.38}$$

$$\triangleq \sum_{n=1}^N \mathcal{M}_\alpha(q_n;\rho). \tag{B.39}$$

Again, we assume that $q_n \in [0,+\infty[$. The derivative reads

$$\frac{\partial}{\partial q_n}\mathcal{M}_\alpha(q_n;\rho) = -\sigma_r^2 q_n y_n^2(\sigma_e^2+\sigma_r^2 q_n^2)^{-2} + \sigma_r^2 q_n(\sigma_e^2+\sigma_r^2 q_n^2)^{-1} + \alpha\rho q_n^{\alpha-1} \tag{B.40}$$

$$= (\sigma_r^2 q_n)\left((\sigma_e^2+\sigma_r^2 q_n^2)^{-1} - y_n^2(\sigma_e^2+\sigma_r^2 q_n^2)^{-2} + \alpha\frac{\rho}{\sigma_r^2}q_n^{\alpha-2}\right). \tag{B.41}$$

First we notice that $q_n = 0$ is always a stationary point. The other stationary points satisfy:

$$(\sigma_e^2+\sigma_r^2 q_n^2)^{-1} - y_n^2(\sigma_e^2+\sigma_r^2 q_n^2)^{-2} + \alpha\frac{\rho}{\sigma_r^2}q_n^{\alpha-2} = 0. \tag{B.42}$$

If $\alpha = 2$, we get for $\rho > 0$:

$$(\sigma_e^2+\sigma_r^2 q_n^2)^{-1} - y_n^2(\sigma_e^2+\sigma_r^2 q_n^2)^{-2} + 2\frac{\rho}{\sigma_r^2} = 0 \tag{B.43}$$

$$\Leftrightarrow 2\frac{\rho}{\sigma_r^2}(\sigma_e^2+\sigma_r^2 q_n^2)^2 + (\sigma_e^2+\sigma_r^2 q_n^2) - y_n^2 = 0 \tag{B.44}$$

$$\Leftrightarrow (\sigma_e^2+\sigma_r^2 q_n^2) = -\frac{\sigma_r^2}{4\rho} \pm \frac{\sigma_r^2}{4\rho}\sqrt{1+8\rho\frac{y_n^2}{\sigma_r^2}} \tag{B.45}$$

$$\Leftrightarrow q_n^2 = -\frac{1}{4\rho} \pm \frac{1}{4\rho}\sqrt{1+8\rho\frac{y_n^2}{\sigma_r^2}} - \frac{\sigma_e^2}{\sigma_r^2} \tag{B.46}$$

Here $\rho > 0$. To satisfy the positivity constraint on q_n^2, we must have:

$$\hat{q}_n^2 = \frac{1}{4\rho}\sqrt{1+8\rho\frac{y_n^2}{\sigma_r^2}} - \frac{1}{4\rho} - \frac{\sigma_e^2}{\sigma_r^2} > 0. \tag{B.47}$$

This condition rereads:

$$\frac{1}{4\rho}\sqrt{1+8\rho\frac{y_n^2}{\sigma_r^2}} > \frac{1}{4\rho} + \frac{\sigma_e^2}{\sigma_r^2} \tag{B.48}$$

$$\Leftrightarrow\ 8\rho\frac{y_n^2}{\sigma_r^2} > \left(1+4\rho\frac{\sigma_e^2}{\sigma_r^2}\right)^2 - 1 \tag{B.49}$$

$$\Leftrightarrow\ 8\rho\frac{y_n^2}{\sigma_r^2} > 8\rho\frac{\sigma_e^2}{\sigma_r^2}\left(1+2\rho\frac{\sigma_e^2}{\sigma_r^2}\right) \tag{B.50}$$

$$\Leftrightarrow\ y_n^2 > \sigma_e^2\left(1+2\rho\frac{\sigma_e^2}{\sigma_r^2}\right) \tag{B.51}$$

When the latter condition is met, it is easy to check (from the sign of the derivative of $\mathcal{M}_2(q_n;\rho)$ that $\hat{q}_n$ defined in Equation (B.47) is the minimizer of $\mathcal{M}_2$. Otherwise, $\mathcal{M}_2$ is increasing over $\mathbb{R}_+$ and the minimizer is equal to 0.

Summarizing, when $\alpha = 2$ and $\rho > 0$, the estimator is given by

$$\hat{q}_n^2 = \begin{cases} \frac{1}{4\rho}\sqrt{1+8\rho\frac{y_n^2}{\sigma_r^2}} - \frac{1}{4\rho} - \frac{\sigma_e^2}{\sigma_r^2} & \text{if } y_n^2 > \sigma_e^2\left(1+2\rho\frac{\sigma_e^2}{\sigma_r^2}\right) \\ 0 & \text{if } y_n^2 \leq \sigma_e^2\left(1+2\rho\frac{\sigma_e^2}{\sigma_r^2}\right) \end{cases} \tag{B.52}$$

In the limit case $\rho \to 0$, we obtain

$$\hat{q}_n^2 = \begin{cases} (y_n^2 - \sigma_e^2)/\sigma_r^2 & \text{if } y_n^2 > \sigma_e^2 \\ 0 & \text{if } y_n^2 \leq \sigma_e^2 \end{cases} \tag{B.53}$$

References

Argyriou, A., Foygel, R., Srebro, N., 2012. Sparse prediction with the k-support norm. Advances in Neural Information Processing Systems 25.

Barbault, P., Kowalski, M., Soussen, C., 2023. Marginal map estimation of a bernoulli-gaussian signal: continuous relaxation approach, in: 2023 31st European Signal Processing Conference (EUSIPCO), IEEE. pp. 1833–1837.

Blumensath, T., Davies, M.E., 2009. Iterative hard thresholding for compressed sensing. Appl. Comp. Harmonic Anal. 27, 265–274.

Bruce, A.G., Gao, H.Y., 1996. Understanding waveshrink: Variance and bias estimation. Biometrika 83, 727–745.

Champagnat, F., Goussard, Y., Idier, J., 1996. Unsupervised deconvolution of sparse spike trains using stochastic approximation. IEEE Trans. Signal Process. 44, 2988–2998.

Goussard, Y., Demoment, G., Idier, J., 1990. A new algorithm for iterative deconvolution of sparse spike trains, in: Proc. IEEE ICASSP, pp. 1547–1550.

Gribonval, R., 2011. Should penalized least squares regression be interpreted as maximum a posteriori estimation? IEEE Transactions on Signal Processing 59, 2405–2410.

Gribonval, R., Machart, P., 2013. Reconciling "priors" & "priors" without prejudice? Advances in Neural Information Processing Systems 26.

Gribonval, R., Nikolova, M., 2021. On Bayesian estimation and proximity operators. Applied and Computational Harmonic Analysis 50, 49–72.

Hazimeh, H., Mazumder, R., Saab, A., 2021. Sparse regression at scale: Branch-and-bound rooted in first-order optimization. Mathematical Programming , 1–42.

Ishwaran, H., Rao, J.S., 2005. Spike and slab variable selection: frequentist and Bayesian strategies. The Annals of Statistics 33, 730–773.

Jiang, X., Nadarajah, S., 2019. On characteristic functions of products of two random variables. Wireless Personal Communications 108, 1157–1177.

Kowalski, M., 2014. Thresholding rules and iterative shrinkage/thresholding algorithm: A convergence study, in: 2014 IEEE International Conference on Image Processing (ICIP), IEEE. pp. 4151–4155.

Loaiza-Ganem, G., Cunningham, J.P., 2019. The continuous Bernoulli: fixing a pervasive error in variational autoencoders. Advances in Neural Information Processing Systems 32.

McDonald, A.M., Pontil, M., Stamos, D., 2016. New perspectives on k-support and cluster norms. The Journal of Machine Learning Research 17, 5376–5413.

Mendel, J.M., 1983. Optimal seismic deconvolution. Academic Press, New York.

Pilanci, M., Wainwright, M.J., El Ghaoui, L., 2015. Sparse learning via boolean relaxations. Mathematical Programming 151, 63–87.

Protter, M., Yavneh, I., Elad, M., 2010. Closed-form MMSE estimation for signal denoising under sparse representation modeling over a unitary dictionary. IEEE Trans. Signal Process. 58, 3471–3484.

Raghavan, P., Tompson, C.D., 1987. Randomized rounding: a technique for provably good algorithms and algorithmic proofs. Combinatorica 7, 365–374.

Robert, C.P., 1997. The Bayesian Choice. A Decision-Theoretic Motivation. Springer Texts in Statistics, Springer Verlag, New York.

Soussen, C., Idier, J., Brie, D., Duan, J., 2011. From Bernoulli-Gaussian deconvolution to sparse signal restoration. IEEE Trans. Signal Process. 59, 4572–4584.

Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological) 58, 267–288.

Tibshirani, R., 2010. Strong rules for discarding predictors in lasso-type problems. URL: https://tibshirani.su.domains/ftp/strongtalk.pdf.

Tipping, M.E., 2001. Sparse Bayesian learning and the relevance vector machine. J. Mach. Learn. Res. 1, 211–244.

Turek, J.S., Yavneh, I., Elad, M., 2011. On MMSE and MAP denoising under sparse representation modeling over a unitary dictionary. IEEE Transactions on Signal Processing 59, 3526–3535.

Wipf, D.P., Rao, B.D., 2004. Sparse Bayesian learning for basis selection. IEEE Trans. Signal Process. 52, 2153–2164.

Wright, S.J., Nowak, R.D., Figueiredo, M.A., 2009. Sparse reconstruction by separable approximation. IEEE Transactions on signal processing 57, 2479–2493.

Yen, T.J., 2011. A majorization–minimization approach to variable selection using spike and slab priors. The Annals of Statistics 39, 1748–1775.