MAXIMUM-A-POSTERIORI SIGNAL RECOVERY WITH PRIOR INFORMATION: APPLICATIONS TO COMPRESSIVE SENSING

Ali Bereyhi and Ralf R. Müller

Friedrich-Alexander Universität Erlangen-Nürnberg
ali.bereyhi@fau.de and ralf.r.mueller@fau.de

ABSTRACT

This paper studies the asymptotic performance of maximum-a-posteriori estimation in the presence of prior information. The problem arises in several applications such as recovery of signals with non-uniform sparsity pattern from underdetermined measurements. With prior information, the maximum-a-posteriori estimator might have asymmetric penalty. We consider a generic form of this estimator and study its performance via the replica method. Our analyses demonstrate an asymmetric form of the decoupling property in the large-system limit. Employing our results, we further investigate the performance of weighted zero-norm minimization for recovery of a non-uniform sparse signal. Our investigations illustrate that for a given distortion, the minimum number of required measurements can be significantly reduced by choosing weighting coefficients optimally.

Index Terms— Maximum-a-posteriori estimation, compressive sensing, weighted norm minimization, decoupling property, replica method

1. INTRODUCTION

The problem of estimating $\boldsymbol{x} \in \mathbb{X}^N$, for some $\mathbb{X} \subset \mathbb{R}$, from

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

with $\mathbf{A} \in \mathbb{R}^{K \times N}$ and $\boldsymbol{z} \sim \mathcal{N}(0, \lambda_0 \mathbf{I}_K)$, arises in various applications. In presence of prior information, the Maximum-A-Posteriori (MAP) estimation approach might deal with an asymmetric penalty term appearing due to the non-identical prior distributions. In this paper, we intend to investigate the asymptotic performance of this class of estimators which encloses several reconstruction schemes in signal processing.

Particular examples of these estimators are the weighted norm minimization schemes [1] in compressive sensing [2, 3] which are employed for recovery of signals with non-uniform sparsity patterns. In this problem, the signal consists of multiple sparse blocks whose sparsity factors are different. A restricted class of such non-uniform sparse settings, in which the signal support is partially known, was addressed in [4], and the modified-CS scheme was proposed for signal recovery. Weighted ℓ_1-norm minimization was further invoked in [5] for non-uniform sparse recovery in which different blocks of signal samples have different sparsity factors. More general settings were investigated in recent studies; see [6–11] and the references therein.

Due to the nonlinear nature of the MAP estimator, basic tools fail to investigate its large-system performance. Several studies thus invoked the replica method for investigation. This method was developed for analysis of spin glasses [12] in the physics literature and accepted as an efficient mathematical tool in information theory; e.g., [13]. The method was moreover employed to investigate the performance of various recovery schemes in large compressive sensing systems [14–17]. For non-uniform sparse models, the method was employed in [18] to study the performance of weighted ℓ_1-norm minimization recovery considering noise-free measurements. In this paper, we consider a generic class of estimators which includes formerly studied schemes such as weighted ℓ_1-norm minimization and also encloses several other settings whose performances have not yet been addressed in the literature. Invoking our results we derive an asymmetric version of the MAP decoupling principle which extends the results of [19, 20] to a larger class of estimators.

2. PROBLEM FORMULATION

Consider (1) with $K/N = \alpha < \infty$ as $N \uparrow \infty$. Let $[N] :=\{1, \dots, N\}$ be partitioned into disjoint subsets $\mathbb{N}_j$ for $j \in [J]$. J is assumed to be fixed and bounded meaning that $J/N \downarrow 0$ as N grows large. The signal $\boldsymbol{x}_{N\times 1}$ is divided into J blocks. The block j is denoted by $\mathbb{B}_j(\boldsymbol{x})$ and contains entries whose indices are in $\mathbb{N}_j$, i.e., $\mathbb{B}_j(\boldsymbol{x}) := \{x_n : n \in \mathbb{N}_j\}$. We use the notation $j(n)$ to denote the index of the block to which x_n belongs, i.e., $x_n \in \mathbb{B}_{j(n)}(\boldsymbol{x})$. The entries of $\boldsymbol{x}$ are independent, and $x_n \sim \mathrm{p}_{j(n)}(x_n; \rho_n)$ where $\{\rho_n\}$ is a deterministic sequence over $[N]$. The signal is reconstructed from $\boldsymbol{y}$ as

$$\hat{\boldsymbol{x}} = \operatorname*{argmin}_{\boldsymbol{v} \in \mathbb{X}^n} \frac{1}{2\lambda} \|\boldsymbol{y} - \mathbf{A}\boldsymbol{v}\|^2 + u(\boldsymbol{v}; \boldsymbol{c}) \tag{2}$$

This work was supported by the German Research Foundation, Deutsche Forschungsgemeinschaft (DFG), under Grant No. MU 3735/2-1.

where λ is the estimation parameter, $\boldsymbol{c}_{N\times 1}$ contains weighting coefficients $\{c_n\}$, and $u(\boldsymbol{v};\boldsymbol{c})$ is a penalty function with decoupling property, i.e., there exist $\{u_j(v_n;c_n)\}$ such that

$$u(\boldsymbol{v};\boldsymbol{c}) = \sum_{n=1}^{N} u_{j(n)}(v_n;c_n). \tag{3}$$

$\mathbf{A}$ is assumed to be random, such that $\mathbf{J} = \mathbf{A}^\mathsf{T}\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{U}^\mathsf{T}$ with $\mathbf{U}$ being Haar distributed and $\mathbf{D}$ denoting the diagonal matrix of eigenvalues. A trivial example is a matrix with independent and identically distributed (i.i.d.) entries. The empirical distribution of eigenvalues when $N \uparrow \infty$ is denoted by $\mathrm{p}_{\mathbf{J}}(\lambda)$. For this distribution, the Stieltjes transform is given by $\mathrm{G}_{\mathbf{J}}(s) = \mathbb{E}\left\{(\lambda - s)^{-1}\right\}$ where $\lambda \sim \mathrm{p}_{\mathbf{J}}(\lambda)$, and the R-transform is defined as $\mathrm{R}_{\mathbf{J}}(\omega) \coloneqq \mathrm{G}_{\mathbf{J}}^{-1}(-\omega) - \omega^{-1}$ with $\mathrm{G}_{\mathbf{J}}^{-1}(\cdot)$ being the inverse with respect to (w.r.t.) composition.

The setting recovers several problems in signal processing. An example is recovery of non-uniform sparse signals from noisy measurements in compressive sensing: Let $J = 1$ and

$$\mathrm{p}_1(x_n;\rho_n) = \rho_n \mathrm{q}(x_n) + (1-\rho_n)\delta(x_n) \tag{4}$$

for some distribution $\mathrm{q}(x_n)$. Then, $\boldsymbol{x}$ models a sparse signal with non-uniform sparsity pattern whose non-zero entries are distributed with $\mathrm{q}(x_n)$. Consequently, by setting $u(v_n;c_n) = c_n|v_n|$, the estimator reduces to the weighted ℓ_1-norm minimization recovery scheme. For $\lambda_0 = 0$, the setup recovers the formerly studied noise-free case, e.g., [5, 8, 18], when $\lambda \downarrow 0$.

In order to quantify the large-system performance of this setting, we define the weighted distortion as follows.

Definition 1 (Weighted Distortion)**:** Let $\boldsymbol{w}_{N\times 1}$ enclose the coefficients $\{w_n\}$. The weighted distortion w.r.t. the distortion function $\mathsf{d}(\cdot;\cdot)$ for a given $\boldsymbol{w}$ reads

$$\mathsf{D}(\boldsymbol{x};\hat{\boldsymbol{x}}|\boldsymbol{w}) \coloneqq \frac{1}{N}\sum_{n=1}^{N} w_n \mathbb{E}\left\{\mathsf{d}(x_n;\hat{x}_n)\right\}. \tag{5}$$

Moreover, the asymptotic weighted distortion is given by taking the limit $N \uparrow \infty$, i.e., $\mathsf{D}_{\boldsymbol{w}} \coloneqq \lim_{N\uparrow\infty} \mathsf{D}(\boldsymbol{x};\hat{\boldsymbol{x}}|\boldsymbol{w})$.

The weighted distortion recovers various forms of recovery distortions. For instance, setting $w_n = 1$ and $\mathsf{d}(x_n;\hat{x}_n) = |x_n - \hat{x}_n|^2$, $\mathsf{D}_{\boldsymbol{w}}$ determines the asymptotic Mean Square Error (MSE). Moreover, it evaluates the average error probability by setting $\mathsf{d}(x_n;\hat{x}_n) = \mathbf{1}\left\{x_n = \hat{x}_n\right\}$ with $\mathbf{1}\left\{\cdot\right\}$ being the indicator function. The main goal of this study is to derive the weighted distortion in its generic form when N grows large.

3. ASYMPTOTIC PERFORMANCE

Invoking the replica method, $\mathsf{D}_{\boldsymbol{w}}$ is derived in a closed form. The derivations are briefly sketched in Section 5. For the sake of compactness, we state the basic form of the result known as the "Replica Symmetry (RS) solution". Our derivations are however in a general form enclosing "Replica Symmetry Breaking (RSB) solutions".

3.1. Asymptotic Weighted Distortion

$\mathsf{D}_{\boldsymbol{w}}$ in the large-system limit can be expressed in terms of an equivalent scalar system. For $j \in [J]$, we define the scalar estimator $\mathrm{g}_j^{\mathsf{dec}}(\cdot;c)$ which for given c and θ reads

$$\mathrm{g}_j^{\mathsf{dec}}(y;c) = \operatorname*{argmin}_{v\in\mathbb{X}} \frac{1}{2\theta}(y-v)^2 + u_j(v;c) \tag{6}$$

$\mathrm{g}_j^{\mathsf{dec}}(y;c)$ represents a estimator which recovers a scalar from the single measurement y using the one-dimensional form of MAP formulation in (2) with the weighting coefficient c and estimation parameter θ. In order to state the result, we moreover define the effective noise variance θ_0 and the equivalent estimation parameter θ for some scalars χ and p as

$$\theta = \left[\mathrm{R}_{\mathbf{J}}(-\frac{\chi}{\lambda})\right]^{-1}\lambda \tag{7a}$$

$$\theta_0 = \left[\mathrm{R}_{\mathbf{J}}(-\frac{\chi}{\lambda})\right]^{-2}\frac{\partial}{\partial\chi}\left[(\lambda_0\chi - \lambda\mathsf{p})\mathrm{R}_{\mathbf{J}}(-\frac{\chi}{\lambda})\right] \tag{7b}$$

where $\mathrm{R}_{\mathbf{J}}(\cdot)$ denotes the R-transform of $\mathrm{p}_{\mathbf{J}}(t)$ defined in the previous section. One should note that θ and θ_0 are controlled by χ and p and are functions of the true estimation parameter λ, statistics of $\mathbf{A}$ and the true noise variance λ_0.

Proposition 1: Let $z^{\mathsf{dec}} \sim \mathcal{N}(0,\theta_0)$, and for each $n \in [N]$, define the decoupled estimation g_n as

$$\mathrm{g}_n \coloneqq \mathrm{g}_{j(n)}^{\mathsf{dec}}(x_n + z^{\mathsf{dec}};c_n). \tag{8}$$

Then, under some assumptions[1], $\mathsf{D}_{\boldsymbol{w}}$ is given by

$$\mathsf{D}_{\boldsymbol{w}} = \left\langle w_n \mathbb{E}\left\{\mathsf{d}\left(x_n;\mathrm{g}_n\right)\right\}\right\rangle_{[N]} \tag{9}$$

where we define $\langle f(a_n)\rangle_{\mathbb{N}} \coloneqq |\mathbb{N}|^{-1}\sum_{n\in\mathbb{N}} f(a_n)$. The variables p and χ which determine θ and θ_0 are moreover calculated from the fixed-point equations

$$\mathsf{p} = \left\langle \mathbb{E}\left\{(\mathrm{g}_n - \mathrm{x}_n)^2\right\}\right\rangle_{[N]}, \tag{10a}$$

$$\frac{\theta_0}{\theta}\chi = \left\langle \mathbb{E}\left\{(\mathrm{g}_n - \mathrm{x}_n)\, z^{\mathsf{dec}}\right\}\right\rangle_{[N]}. \tag{10b}$$

Proof: The proof is briefly sketched in Section 5. The details of the proof, however, are skipped due to the page limitation.

3.2. Asymmetric Decoupling Property

Proposition 1 determines the asymptotic weighted distortion by averaging the scalar systems shown in Fig. 1 over n w.r.t. $\boldsymbol{w}$. In fact by setting $\mathrm{x}_n = x_n$ and $\hat{\mathrm{x}}_n = \mathrm{g}_n$ in this diagram, one observes that $\mathsf{D}_{\boldsymbol{w}}$ is the weighted average of input-output distortions. These scalar systems can be further shown to describe input-output marginal distributions. This

[1] These assumptions are mainly replica continuity and the replica symmetry which are later introduced in Section 5.

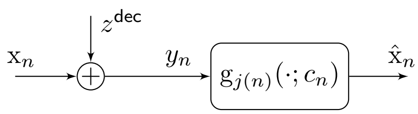

Fig. 1: Asymmetric Decoupling Property: The decoupled systems are dependent on the index n in general.

observation states that the estimator exhibits the decoupling property in the large-system limit. To illustrate this property, let us denote the marginal joint distribution of $(\hat{x}_n, x_n)$ with $\mathrm{q}_N(\hat{x}_n, x_n)$ where the subscript indicates the dependency of the distribution on N. The asymptotic decoupling property mainly claims that as N grows, $\mathrm{q}_N(\hat{x}_n, x_n)$ converge to a deterministic distribution described by the input-output distribution of the scalar system in Fig. 1. The previously studied forms of the property, e.g., [19, 20], have considered identically distributed source entries, i.e., $\mathrm{p}_{j(n)}(\cdot;\rho_n) = \mathrm{p}(\cdot;\rho)$ and $u_{j(n)}(\cdot;c_n) = u(\cdot;c)$ for some constants ρ and c. For this case, the limiting distribution is shown to be independent of n, and thus, the equivalent scalar systems are the same. The decoupled system derived in this paper, however, can vary from one index to another. We therefore refer to this form of decoupling as the "asymmetric decoupling property" which recovers the previous "symmetric" forms. The property is stated in the following. The proof follows the moment method and takes a similar path as in [16] with some modifications. It is however omitted for the sake of compactness.

Asymmetric Decoupling: For $n \in [N]$, $(\hat{x}_n, x_n)$ converges in distribution to the pair $(\hat{\mathrm{x}}_n, \mathrm{x}_n)$ in Fig. 1 with $\mathrm{g}_j^{\mathsf{dec}}(\cdot; c_n)$ and z^{dec} being given in Proposition 1.

4. APPLICATIONS OF THE MAIN RESULTS

The asymptotic results presented in Section 3 can be employed to investigate various estimation problems. In the sequel, we give some examples in compressive sensing.

4.1. Recovery of Non-uniform Sparse Signals

Stochastic signals with non-uniform sparsity patterns are described by our setting when $J = 1$ and the signal entries x_n are distributed as in (4). Several recovery schemes, some of which have not been addressed in the literature, can then be investigated by choosing corresponding utility functions. A trivial approach is to let the utility function be

$$u(\boldsymbol{v}; \boldsymbol{c}) = \sum_{n=1}^{N} c_n |v_n|^p. \tag{11}$$

Using Proposition 1, the large-system performance of these recovery schemes can be studied w.r.t. various forms of distortions. Moreover, the optimal choices for $\{c_n\}$ can be found in terms of the priors $\{\rho_n\}$, such that the average distortion is minimized. This investigation widens the scope of analyses in [18] to noisy scenarios and various recovery schemes. Moreover, it enables us to extend the recent study in [21] to cases with prior information on the sparsity pattern. To discuss further the application of the results in recovery of non-uniform sparse signals, we consider the following example.

Example 1: Assume that $\boldsymbol{x}$ is a sparse-Gaussian signal with a non-uniform sparsity pattern, i.e., $J = 1$ and the distribution of x_n for $n \in [N]$ are given by $\mathrm{p}_1(x_n; \rho_n)$ in (4) with $\mathrm{q}(x_n)$ being the zero-mean and unit-variance Gaussian distribution. To recover the signal, we employ the weighted zero-norm recovery scheme which is given by setting $u(v_n; c_n) = c_n \mathbf{1}\{v_n \neq 0\}$ in (2). Proposition 1 enables us to investigate the recovery performance in this case and also evaluate the optimal choice of $\{c_n\}$ in terms of $\{\rho_n\}$. For the sake of simplicity, consider the scenario in which $\mathbf{A}$ is an i.i.d. matrix whose entries are zero-mean with variance $1/K$. In this case, $\mathrm{p}_{\mathbf{J}}$ follows the Marcenko-Pastur law [22], and thus, $\mathrm{R}_{\mathbf{J}}(\omega) = \alpha(\alpha - \omega)^{-1}$ which implies $\theta = \lambda + \alpha^{-1}\chi$ and $\theta_0 = \lambda + \alpha^{-1}\mathsf{p}$. Moreover,

$$\mathrm{g}^{\mathsf{dec}}(y_n; c_n) = \begin{cases} y_n & |y_n| > t_n \\ 0 & |y_n| \leq t_n \end{cases} \tag{12}$$

where $t_n := \sqrt{2\theta c_n}$. Consequently, the asymptotic distortion w.r.t. some given distortion function and $\boldsymbol{w}$ is determined by Proposition 1. As (12) shows, weighted zero-norm recovery decouples asymptotically into a set of hard thresholding operators whose threshold levels depend on weights c_n. By setting $c_n = 1$ and $\rho_n = \rho$ for all $n \in [N]$, the decoupled setups reduce to the symmetric setups reported in [19, 20].

To investigate the performance of weighted zero-norm recovery numerically, we consider the configuration in which

$$\rho_n = \begin{cases} \rho_0 & n \in [N/B], \\ \rho_1 & n \in [N/B+1:N], \end{cases} \tag{13}$$

for some $\rho_0, \rho_1 \in [0, 1]$ and some integer B being a divisor of N. Here, $[M : N]$ denotes $\{M, \dots, N\}$. Moreover, we set

$$c_n = \begin{cases} 1 & n \in [N/B], \\ c & n \in [N/B+1:N], \end{cases} \tag{14}$$

for some c. We denote the asymptotic average MSE by $\mathsf{mse} := \lim_{N\uparrow\infty} \mathbb{E}\left\{\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|^2\right\}/N$. Moreover, for a given mse_0, we define the threshold compression rate $R_{\mathrm{t}}(\mathsf{mse}_0)$ to be the maximum possible inverse load factor $\alpha^{-1} = N/K$ which results in $\mathsf{mse} \leq \mathsf{mse}_0$. Fig. 2 shows the threshold compression rate as a function of c for $\mathsf{mse}_0 = -25$ dB. The curves have been plotted for $\rho_0 = 0.1$ considering various choices of ρ_1 and B. The noise power is set to be $\lambda_0 = 0.01$ and λ is tuned such that the MSE is minimized at each load factor. As

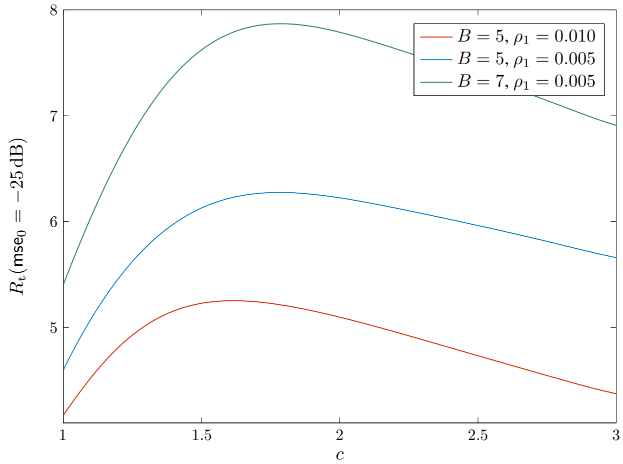

Fig. 2: $R_{\mathrm{t}}(\mathsf{mse}_0)$ at $\mathsf{mse}_0 = -25$ dB versus the norm-weight of the more sparse block in Example 1. As the figure shows, either the growth in the size of the block or reduction in its sparsity increases the degradation caused by uniform recovery, i.e., $c = 1$.

the figure shows, the optimal choice of c can significantly increase the threshold compression rate. The curves moreover indicate that as B grows or ρ_1 reduces the gap between the optimal $R_{\mathrm{t}}(\mathsf{mse}_0)$, maximized over c, and the threshold compression rate at $c = 1$ increases. This observation is due to the fact that the growth in B or the reduction in ρ_1 imposes more asymmetry into the setting, and therefore, increases the loss caused by uniform recovery, i.e., $c = 1$.

4.2. Non-uniform Sparsity with Multiple Prior States

The non-uniform sparsity model can be extended to signals with multiple prior states by considering $J > 1$. In this case,

$$\mathrm{p}_{j(n)}(x_n; \rho_n) = \rho_n \mathrm{q}_{j(n)}(x_n) + (1 - \rho_n)\delta(x_n) \tag{15}$$

represents a signal with non-uniform sparsity pattern whose non-zero entries are taken from multiple possible prior distributions. This model describes a scenario in which multiple uncorrelated non-uniform sparse signals are simultaneously measured, e.g., a network of independent sensors with different prior distributions. An efficient approach for signal recovery in this case is to set $u_{j(n)}(v_n; c_n) = f_{j(n)}(v_n) + c_n \mathbf{1}\{v_n \neq 0\}$ for some $f_j(v)$. Similar to the case with single state sources, the optimal choice of $\{c_n\}$ as well as the asymptotic distortion is determined using Proposition 1.

5. LARGE-SYSTEM ANALYSIS

In this section, we briefly sketch the derivations based on the replica method. Consider $\mathcal{E}(\boldsymbol{v}) = \|\boldsymbol{y} - \mathbf{A}\boldsymbol{v}\|^2/2\lambda + u(\boldsymbol{v}; \boldsymbol{c})$, and define $\mathcal{Z}(\beta, h) = \sum_{\boldsymbol{v}} \exp\{-\beta\mathcal{E}(\boldsymbol{v}) + hN\mathsf{D}(\boldsymbol{x}; \boldsymbol{v}|\boldsymbol{w})\}$. One can then employ large deviation arguments and write

$$\mathsf{D}_{\boldsymbol{w}} = \lim_{N\uparrow\infty} \lim_{\beta\uparrow\infty} \frac{\partial}{\partial h} \mathcal{F}(\beta, h)|_{h=0}. \tag{16}$$

where $\mathcal{F}(\beta, h) = \mathbb{E}\{\log \mathcal{Z}(\beta, h)\}/N$. As evaluating a logarithmic expectation is not a trivial task, we invoke the replica method. The main idea comes from the Riesz equality [23] which states $\mathbb{E}\{\log \mathrm{x}\} = \lim_{m\downarrow 0} \log \mathbb{E}\{\mathrm{x}^m\}/m$. Using this equality, $\mathsf{D}_{\boldsymbol{w}}$ is determined in terms of the mth moment of $\mathcal{Z}(\beta, h)$. Nevertheless, the moments need to be determined for real values of m which is still challenging. This challenge is addressed by assuming "replica continuity" which means that $\mathbb{E}\{\mathcal{Z}^m(\beta, h)\}$ analytically continues from $m \in \mathbb{Z}^+$ to $m \in \mathbb{R}^+$. After calculating the moments, $\mathsf{D}_{\boldsymbol{w}}$ is given by

$$\mathsf{D}_{\boldsymbol{w}} = \lim_{\beta\uparrow\infty} \lim_{m\downarrow 0} \sum_{\mathbf{v}} \left\langle \mathbb{E}\left\{ w_n \mathsf{d}(\mathbf{v}; \mathbf{x}_n) \mathrm{p}_n^{\beta}(\mathbf{v}|\mathbf{x}_n) \right\} \right\rangle_{[N]}, \tag{17}$$

for $\mathbf{v} \in \mathbb{X}^m$, where $\mathbf{x}_n$ is an $m \times 1$ vector with all the entries being x_n and $\mathsf{d}(\mathbf{v}; \mathbf{x}_n) := \sum_{a=1}^{m} \mathsf{d}(\mathrm{v}_a; x_n)$; moreover,

$$\mathrm{p}_n^{\beta}(\mathbf{v}|\mathbf{x}_n) = \frac{e^{-\beta\left[(\mathbf{v}-\mathbf{x}_n)^{\mathsf{H}}\mathbf{R}(\mathbf{v}-\mathbf{x}_n) + u_{j(n)}(\mathbf{v}; c_n)\right]}}{\sum_{\mathbf{v}} e^{-\beta\left[(\mathbf{v}-\mathbf{x}_n)^{\mathsf{H}}\mathbf{R}(\mathbf{v}-\mathbf{x}_n) + u_{j(n)}(\mathbf{v}; c_n)\right]}}. \tag{18}$$

with $u_j(\mathbf{v}; c_n) := \sum_{a=1}^{m} u_j(\mathrm{v}_a; c_n)$, and $\mathbf{R} := \mathbf{T}\mathrm{R}_{\mathbf{J}}(-\beta\mathbf{T}\mathbf{Q})$ for $\mathbf{T} = \dfrac{1}{2\lambda}(\mathbf{I}_m - \beta\dfrac{\lambda_0}{\lambda}\mathbf{1}_m)$ and some $\mathbf{Q}_{m\times m}$ which satisfies

$$\mathbf{Q} = \sum_{\mathbf{v}} \left\langle \mathbb{E}\left\{ \mathrm{p}_n^{\beta}(\mathbf{v}|\mathbf{x}_n)(\mathbf{v} - \mathbf{x}_n)(\mathbf{v} - \mathbf{x}_n)^{\mathsf{H}} \right\} \right\rangle_{[N]}. \tag{19}$$

In (17), the general replica solution is given. The explicit determination of $\mathsf{D}_{\boldsymbol{w}}$, however, needs $\mathbf{Q}$ to be found such that (19) is fulfilled. To do so, we need to suppose a structure for $\mathbf{Q}$. The basic structure is given by RS as $\mathbf{Q} = \chi\beta^{-1}\mathbf{I}_m + \mathsf{p}\mathbf{1}_m$ for some χ and p. By substituting $\mathbf{Q}$ in (17), Proposition 1 is concluded after some lines of derivations. The RSB solutions are further derived by extending the RS structure to

$$\mathbf{Q} = \frac{\chi}{\beta}\mathbf{I}_m + \sum_{\kappa=1}^{b} \mathsf{c}_\kappa \mathbf{I}_{\frac{m\beta}{\mu_\kappa}} \otimes \mathbf{1}_{\frac{\mu_\kappa}{\beta}} + \mathsf{p}\mathbf{1}_m, \tag{20}$$

for some integer b. The derivations under RSB follow [16, Appendix D] and are omitted due to the page limitation.

6. CONCLUSION

In this paper, we have studied the asymptotic performance of a class of MAP-based signal recovery schemes when prior information is available for reconstruction. Our analysis has demonstrated an asymmetric version of the decoupling principle for these estimators which generalizes the formerly studied forms of MAP decoupling [19, 20]. Invoking the results, we have investigated the performance of weighted zero-norm minimization for recovery of a signal with non-uniform sparsity pattern. The results of this paper can be further employed to study various problems. A particular example in compressive sensing is to extend the scope of investigations in [21] to signals with non-uniform sparsity patterns and study the impact of replacing the ℓ_1-norm with an ℓ_p-norm in the weighted norm minimization scheme for $0 \leq p \leq 1$. Currently, the work in this direction has been started.

7. REFERENCES

[1] Emmanuel J Candes, Michael B Wakin, and Stephen P Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *Journal of Fourier analysis and applications*, vol. 14, no. 5, pp. 877–905, 2008.

[2] David L Donoho, "Compressed sensing," *IEEE Trans. on Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[3] Emmanuel J Candes, Justin K Romberg, and Terence Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. on pure and applied mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.

[4] Namrata Vaswani and Wei Lu, "Modified-CS: Modifying compressive sensing for problems with partially known support," *IEEE Trans. on Signal Processing*, vol. 58, no. 9, pp. 4595–4607, 2010.

[5] M Amin Khajehnejad, Weiyu Xu, A Salman Avestimehr, and Babak Hassibi, "Weighted ℓ_1 minimization for sparse recovery with prior information," in *Inf. Theory (ISIT), IEEE International Symposium on*. 2009, pp. 483–487.

[6] Laurent Jacques, "A short note on compressed sensing with partially known signal support," *Signal Processing*, vol. 90, no. 12, pp. 3308–3312, 2010.

[7] Hassan Mansour and Özgür Yilmaz, "Support driven reweighted ℓ_1 minimization," in *Acoustics, Speech and Signal Processing (ICASSP), IEEE International Conference on*. 2012, pp. 3309–3312.

[8] Jonathan Scarlett, Jamie S Evans, and Subhrakanti Dey, "Compressed sensing with prior information: Information-theoretic limits and practical decoders," *IEEE Trans. on Signal Processing*, vol. 61, no. 2, pp. 427–439, 2013.

[9] Samet Oymak, M Amin Khajehnejad, and Babak Hassibi, "Recovery threshold for optimal weight ℓ_1 minimization," in *Inf. Theory (ISIT), IEEE International Symposium on*. 2012, pp. 2032–2036.

[10] Bubacarr Bah and Rachel Ward, "The sample complexity of weighted sparse approximation.," *IEEE Trans. Signal Processing*, vol. 64, no. 12, pp. 3145–3155, 2016.

[11] Holger Rauhut and Rachel Ward, "Interpolation via weighted ℓ_1 minimization," *Applied and Computational Harmonic Analysis*, vol. 40, no. 2, pp. 321–351, 2016.

[12] Samuel Frederick Edwards and Phil W Anderson, "Theory of spin glasses," *Journal of Physics F: Metal Physics*, vol. 5, no. 5, pp. 965, 1975.

[13] Toshiyuki Tanaka, "A statistical-mechanics approach to large-system analysis of CDMA multiuser detectors," *IEEE Trans. on Inf. Theory*, vol. 48, no. 11, pp. 2888–2910, 2002.

[14] Antonia M Tulino, Giuseppe Caire, Sergio Verdu, and Shlomo Shamai, "Support recovery with sparsely sampled free random matrices," *IEEE Trans. on Inf. Theory*, vol. 59, no. 7, pp. 4243–4271, 2013.

[15] Mikko Vehkaperä, Yoshiyuki Kabashima, and Saikat Chatterjee, "Analysis of regularized LS reconstruction and random matrix ensembles in compressed sensing," *IEEE Trans. on Inf. Theory*, vol. 62, no. 4, pp. 2100–2124, 2016.

[16] Ali Bereyhi, Ralf R Müller, and Hermann Schulz-Baldes, "Statistical mechanics of MAP estimation: General replica ansatz," *arXiv preprint arXiv:1612.01980*, 2016.

[17] Ali Bereyhi, Ralf R Müller, and Hermann Schulz-Baldes, "Replica symmetry breaking in compressive sensing," in *IEEE Inf. Theory and Applications Workshop (ITA)*. 2017, pp. 1–7.

[18] Toshiyuki Tanaka and Jack Raymond, "Optimal incorporation of sparsity information by weighted ℓ_1 optimization," in *Information Theory (ISIT), IEEE Int. Sym, on*. 2010, pp. 1598–1602.

[19] Sundeep Rangan, Alyson K Fletcher, and Vivek Goyal, "Asymptotic analysis of MAP estimation via the replica method and applications to compressed sensing," in *IEEE Trans. on Inf. Theory*, 2012, pp. 1902–1923.

[20] Ali Bereyhi, Ralf Müller, and Hermann Schulz-Baldes, "RSB decoupling property of MAP estimators," in *IEEE Inf. Theory Workshop (ITW)*. 2016, pp. 379–383.

[21] Le Zheng, Arian Maleki, Haolei Weng, Xiaodong Wang, and Teng Long, "Does ℓ_p-minimization outperform ℓ_1-minimization?," *IEEE Trans. on Information Theory*, 2017.

[22] Ralf R Müller, Giusi Alfano, Benjamin M Zaidel, and Rodrigo de Miguel, "Applications of large random matrices in communications engineering," *arXiv preprint arXiv:1310.5479*, 2013.

[23] Frédéric Riesz, "Sur les valeurs moyennes des fonctions," *Journal of the London Mathematical Society*, vol. 1, no. 2, pp. 120–121, 1930.