Analysis of MMSE Estimation for Compressive Sensing of Block Sparse Signals

Mikko Vehkaperä, Saikat Chatterjee, and Mikael Skoglund
School of Electrical Engineering and the ACCESS Linnaeus Center
KTH Royal Institute of Technology, SE-10044, Stockholm, Sweden
E-mail: {mikkov,sach,skoglund}@kth.se

Abstract—Minimum mean square error (MMSE) estimation of block sparse signals from noisy linear measurements is considered. Unlike in the standard compressive sensing setup where the non-zero entries of the signal are independently and uniformly distributed across the vector of interest, the information bearing components appear here in large mutually dependent clusters. Using the replica method from statistical physics, we derive a simple closed-form solution for the MMSE obtained by the optimum estimator. We show that the MMSE is a version of the Tse-Hanly formula with system load and MSE scaled by parameters that depend on the sparsity pattern of the source. It turns out that this is equal to the MSE obtained by a genie-aided MMSE estimator which is in fact in advance about the exact locations of the non-zero blocks. The asymptotic results obtained by the non-rigorous replica method are found to have an excellent agreement with finite sized numerical simulations.

I. INTRODUCTION

Compressive sensing (CS) [1], [2] tackles the problem of recovering a high-dimensional sparse vector from a set of linear measurements. Typically the number of observations is much less than the number of elements in the vector of interest, making naive reconstruction attempts inefficient. In addition to being an under-determined problem, the measurements may suffer from additive noise. Under such conditions, the signal model for the noisy CS measurements can be written as

\[ y = Ax + n \in \mathbb{R}^M, \]

where \( x \in \mathbb{R}^N \) is the sparse vector of interest, \( A \in \mathbb{R}^{M \times N} \) the measurement matrix, and \( n \in \mathbb{R}^M \) represents the measurement noise. By assumption, \( M < N \) and only some of the elements of \( x \) are non-zero. The task of CS is then to infer \( x \), given \( A, y \) and possibly some information about the sparsity of \( x \) and the statistics of the noise \( n \).

In this paper, the vector \( x \) is assumed to have a special block sparse structure. Such sparsity patterns have recently been found, e.g., in multiband signals and multipath communication channels (see, e.g., [3]–[9] and references therein). More precisely, the source is considered to be \( K \) block sparse so that for any realization of \( x \), its information bearing entries occur in at most \( K \) non-overlapping clusters. This is markedly different from the conventional sparsity assumption in CS, where the individual non-zero components are independently distributed over \( x \) and uniformly distributed over \( A \). Given the \( K \) block sparse source, we study the minimum mean square error (MMSE) estimation of \( x \), assuming full knowledge of the statistics of the input \( x \) and the noise \( n \). Albeit this is an optimistic scenario for practical CS problems, it provides a lower bound on the MSE for any other reconstruction method. Also, knowing the benefits of having the statistics of the system at the estimator gives a hint how much the sub-optimum blind schemes could improve if they were to learn the parameters of the problem.

The main result of the paper is the closed-form MMSE for the CS of block sparse signals. The solution turns out to be of a particularly simple form, namely, the Tse-Hanly formula [7] where the system load and MSE are scaled by parameters that depend on the sparsity pattern of the source. This is found to be equal to the MSE obtained by a genie-aided MMSE estimator that is informed in advance the locations of the non-zero blocks. The result implies that if the statistics of the block sparse CS problem are known, the MMSE is independent of the knowledge about the positions of the non-vanishing blocks.

Finally, we remark that the analysis in the paper are obtained via the replica method (RM) from statistical physics. Albeit the RM is non-rigorous, it has been used with great success for the large system analysis of, e.g., multi-antenna systems [8], [9], code division multiple access [10], [11], vector precoding [12], iterative receivers [13] and compressed sensing [14]–[17]. The main difference here compared to [9]–[17] is that the elements of the \( K \) block sparse vector \( x \) are neither independent nor identically distributed. This requires a slight modification to the standard replica treatment, as detailed in the Appendix.

A. Notation

The probability density function (PDF) of a random vector (RV) \( x \in \mathbb{R}^N \) (assuming one exists) is written as \( p(x) \), and conditional densities are denoted \( p(x | \cdot \cdot \cdot) \). The related PDFs postulated by the estimator are denoted \( q(\hat{x}) \) and \( q(\hat{x} | \cdot \cdot \cdot) \), respectively. For further discussion on the so-called generalized posterior mean estimation using true and postulated probabilities, see for example, [10], [11], [13]. We denote \( y \sim p(y) = g_N(y | \mu; \Sigma) \) for a RV \( y \) that is drawn according to the \( N \)-dimensional Gaussian density \( g_N(y | \mu; \Sigma) \) with mean \( \mu \in \mathbb{R}^N \) and covariance \( \Sigma \in \mathbb{R}^{N \times N} \). For a vector \( x \) that is drawn according to a Gaussian mixture density, we have...
\[
x \sim p(x) = \sum_{r=1}^{R} \omega_r g_N(x \mid \mu_r; \Sigma_r),
\]
where the density parameters \(\omega_r\) satisfy \(\sum_{r=1}^{R} \omega_r = 1\) and \(\omega_r \geq 0\) for all \(r = 1, \ldots, R\).

We write \(1_Q \in \mathbb{R}^{Q \times Q}\) for the all-ones vector having \(Q\) elements, and given vectors \(\{d_r \in \mathbb{R}^{R} \}_{r=1}^{R}\), the diagonal matrix \(D = \text{diag}(d_1, \ldots, d_R) \in \mathbb{R}^{R \times R}\) has vector \(\{d_1^T \ldots d_R^T\}^T \in \mathbb{R}^{NR}\) on the main diagonal and zeros elsewhere. Superscript \(^T\) denotes the transpose of a matrix.

\[
x \sim p(x) = \sum_{r=1}^{R} \omega_r g_N(x \mid \mu_r; \Sigma_r),
\]
where the density parameters \(\omega_r\) satisfy \(\sum_{r=1}^{R} \omega_r = 1\) and \(\omega_r \geq 0\) for all \(r = 1, \ldots, R\).

We write \(1_Q \in \mathbb{R}^{Q \times Q}\) for the all-ones vector having \(Q\) elements, and given vectors \(\{d_r \in \mathbb{R}^{R} \}_{r=1}^{R}\), the diagonal matrix \(D = \text{diag}(d_1, \ldots, d_R) \in \mathbb{R}^{R \times R}\) has vector \(\{d_1^T \ldots d_R^T\}^T \in \mathbb{R}^{NR}\) on the main diagonal and zeros elsewhere. Superscript \(^T\) denotes the transpose of a matrix.

\[
x \sim p(x) = \sum_{r=1}^{R} \omega_r g_N(x \mid \mu_r; \Sigma_r),
\]
where the density parameters \(\omega_r\) satisfy \(\sum_{r=1}^{R} \omega_r = 1\) and \(\omega_r \geq 0\) for all \(r = 1, \ldots, R\).

We write \(1_Q \in \mathbb{R}^{Q \times Q}\) for the all-ones vector having \(Q\) elements, and given vectors \(\{d_r \in \mathbb{R}^{R} \}_{r=1}^{R}\), the diagonal matrix \(D = \text{diag}(d_1, \ldots, d_R) \in \mathbb{R}^{R \times R}\) has vector \(\{d_1^T \ldots d_R^T\}^T \in \mathbb{R}^{NR}\) on the main diagonal and zeros elsewhere. Superscript \(^T\) denotes the transpose of a matrix.
Given the MMSE estimate of Proposition 1, we are now interested in computing the per-component MSE

$$\text{mse}(\sigma^2) = E \| x - \langle \langle \hat{x} \rangle \rangle_p \|^2 / N,$$

(19)

where the dimensions of $A$ grow large with fixed ratio $\beta = N/M$, and the number of blocks $R = N/Q$ stays finite. The desired result is obtained in two steps: 1) the replica method is used in Sec. III-A to show that the original problem can be transformed to a set of simpler ones in the large system limit; 2) the solutions to the transformed problems are given in Sec. III-B.

### III. MAIN RESULTS

#### A. Equivalent AWGN Channel

Let the indexes $k = 1, \ldots, K$ and $l = 1, \ldots, L_k$ be as in the previous section. Define a set of additive white Gaussian noise (AWGN) channels for all $k$ and $l$

$$z_{k,l} = x_{k,l} + \xi_{k,l} \eta \in \mathbb{R}^N, \quad \eta \sim g_N(\eta | 0; I_N),$$

(20)

where $x_{k,l}$ is a zero-mean Gaussian with covariance $D_{k,l}$ and $\xi_{k,l} > 0$. Let the events of observing the $(k, l)$th channel be independent and occur with probability $\omega_{k,l}$, for all $k$ and $l$. Furthermore, let

$$\langle \ldots \rangle_q^{(k,l)} = \int \cdots \int q(\tilde{x}_{k,l} | z_{k,l}) d\tilde{x}_{k,l}$$

(21)

be an expectation operator similar to (10), but related to the $(k, l)$th AWGN channel (20). The MMSE estimate of $x_{k,l}$ given the channel outputs $z_{k,l}$ is then given by

$$\langle \hat{x}_{k,l} \rangle_p^{(k,l)} = D_{k,l} (D_{k,l} + \xi_{k,l} I_N)^{-1} z_{k,l}.$$  

(22)

We denote the per-component MSE of the estimates $\langle \hat{x}_{k,l} \rangle_p^{(k,l)}$ as

$$\text{mse}_q^{(k,l)}(\xi_{k,l}^2) = E \| x_{k,l} \|^2 - E \langle \langle \hat{x}_{k,l} \rangle_p^{(k,l)} \|^2 / N,$$

(23)

where the expectation is w.r.t. the joint distribution of all variables associated with (20). The per-component MSE averaged over the realizations of the channels (20) is thus

$$\text{mse}_q(\xi_{k,l}^2) = \sum_{k=1}^K \sum_{l=1}^{L_k} \omega_{k,l} \text{mse}_q^{(k,l)}(\xi_{k,l}^2).$$  

(24)

**Claim 1.** Let $x$ be (approximately) $K$ block sparse as given in Definition 1. In the large system limit when $M, N, Q \to \infty$ with finite and fixed ratios $\beta = N/M$ and $R = N/Q$,

$$\text{mse}(\sigma^2) \to \text{mse}_q(\xi_{k,l}^2),$$

(25)

where $\text{mse}(\sigma^2)$ is given in (19) and $\text{mse}_q(\xi_{k,l}^2)$ in (24). The noise variance $\xi_{k,l}^2 > 0$, on the other hand, is the solution to the fixed point equation

$$\xi_{k,l}^2 = \sigma^2 + \beta \text{mse}_q^{(k,l)}(\xi_{k,l}^2),$$

(26)

for all $k$ and $l$.

**Proof:** The proof is based on the replica method (see, e.g., [8]–[17] for similar results in communication theory and signal processing) from statistical physics. The main difference to the standard approach is that here the elements of the input vector $x$ are neither independent nor identically distributed. Thus, the decoupling result proved for the CDMA systems [11] cannot be straightforwardly extended to our case. Alternative derivation is sketched below and in part in the Appendix.

To start, let us define a modified partition function related to the posterior mean estimator (10) as

$$Z(y, A, \lambda) = \int e^{\lambda^T \hat{x} q(y | A, \hat{x}) q(\hat{x}) d\hat{x}},$$

(27)

where $\lambda \in \mathbb{R}^N$ is a constant vector. The posterior mean estimator of $x$ given in (10) can then be written as the gradient with respect to (w.r.t) $\lambda$ at $\lambda = 0$ of the free energy, i.e.,

$$\langle \langle \hat{x} \rangle \rangle_q = \nabla_\lambda \log Z(y, A, \lambda) \bigg|_{\lambda=0}.$$

(28)

Similarly, if (28) is the optimum MMSE estimate, that is $q = p$, we denote the free energy density

$$f(y, A, \lambda) = \frac{1}{N} \log Z(y, A, \lambda),$$

(29)

the average per-component MMSE is given by

$$\text{mse} = \text{tr} \left( E \{ \nabla^2_{\lambda\lambda} f(y, A, \lambda) \big|_{\lambda=0} \} \right),$$

(30)

where the expectations are w.r.t. the joint density of $(y, A)$. Unlike in (19), however, direct computation of the free energy (density) is not possible here. We thus resort to the non-rigorous RM to calculate (27) and then use the relation (30) to obtain the final result. The details are given in the Appendix.

### B. Performance of MMSE Estimation of Block Sparse Signals

Claim 1 asserts that the MSE of the estimator (11) in the original setting (1) can be obtained in the large system limit from (24). Given the Claim 1 holds, a bit of algebra gives the following proposition.

**Proposition 2.** Let $\sigma^2 = 1$, $\xi_{k,l} > 0$ and $N \to \infty$. Then the per-component MSE (25) is given by

$$\text{mse}_q^{(k,l)}(\xi_{k,l}^2) = k \frac{\xi_{k,l}^2 \delta^2}{R \xi_{k,l}^2 + 1} + \frac{R - k}{R} \frac{\xi_{k,l}^2 \delta^2}{\xi_{k,l}^2 + \delta^2},$$

(31)

where $\xi_{k,l}^2$ is the solution of (26). When the source is strictly block sparse, that is $\delta^2 \to 0^+$,

$$\xi_{k,l}^2 = \frac{1}{2} \left( -1 + \beta_k + \sigma^2 + \sqrt{4 \sigma^4 + (1 + \beta \xi_{k,l}^2)^2} \right),$$

(32)

where $\xi_{k,l}^2 = \xi_{k,l}^2 \forall l = 1, \ldots, L_k$ and we denoted $\beta_k = \frac{\beta}{R}$ for notational convenience. Thus, given Claim 1 holds, the MSE of the block sparse system is given by

$$\text{mse}(\sigma^2) = \sum_{k=1}^K \omega_k \frac{k \xi_{k,l}^2}{R \xi_{k,l}^2 + 1},$$

(33)

in the large system limit.

**Remark 1.** Note that the MSE is independent of the distribution $\left\{ \omega_k \right\}_{k=1}^K$ that makes up $\omega_k$ (see [5]).
system limit $N \to \infty$. Then (29) can be written as
\begin{equation}
  f = \lim_{N \to \infty} \frac{1}{N} \frac{1}{u \to 0} \frac{\partial}{\partial y} \log \mathbb{E}_{y, A} \left\{ Z(y, A, \lambda)^u \right\},
\end{equation}
where $u$ is a real parameter. The replica trick consists of treating $u$ as an integer while calculating the expectations, but taking the limit as if $u$ was real valued outside the expectation. The second step is to exchange the limits and write the power of $u$ inside the expectation using the set $\{x_{[a]}\}_{a=1}^u$ of replicated random vectors, resuling to,
\begin{equation}
  f_{rm} = \lim_{u \to 0} \frac{\partial}{\partial u} \lim_{N \to \infty} \frac{1}{N} \log \Xi_N^{(u)}(\lambda),
\end{equation}
where $x_{[a]}$ are IID with density $p(x)$ and
\begin{equation}
  \Xi_N^{(u)}(\lambda) = \mathbb{E}_{y, A} \left\{ \int \prod_{a=1}^u p(x_{[a]}) e^{\lambda^T x_{[a]} p(y | A, x_{[a]})} dx_{[a]} \right\}.
\end{equation}
Unfortunately, these steps are non-rigorous and there is no general proof yet under which conditions $f_{rm}$ equals $f$. For more discussion and details, see, e.g., [8] [13].

Let $x_{[0]}$ be the true vector of interest, independent of $\{x_{[a]}\}_{a=1}^u$ and distributed as $x$. Plugging $y = Ax_{[0]} + n$ to (36), the average over the additive noise vector $n \sim g_M(n | 0; \sigma^2 I_M)$ can be assessed using (15). Furthermore, recalling that the true and postulated source vectors have GM densities (4), we obtain (37) at the top of the next page, where $\mathbb{E}_{x_{[a]}}$ denotes expectation over the vectors $x_{[a]} \sim g_N(x_{[a]} | 0; D_{k,l})$, $a = 0, 1, \ldots, u$. Next, let $v = ([v_0]^T \ldots [v_u]^T)^T \in \mathbb{R}^{M(u+1)}$, be a RV composed of $u+1$ sub-vectors $v_a = \beta^{-1/2} A x_{[a]} \in \mathbb{R}^M$. Also denote $Q_{k,l} = Q_{k,l}^{(u)} \otimes I_M$ where $Q_{k,l}^{(u)} \in \mathbb{R}^{(u+1) \times (u+1)}$ and the $(a,b)$th element of $Q_{k,l}^{(u)}$ is given by $Q_{k,l}^{[a,b]} = x_{[a]}^T x_{[b]}/N$ where $x_{[a]}, x_{[b]} \sim g_N(x | 0; D_{k,l})$ for all $a, b = 0, 1, \ldots, u$. Then, (37) can be written in the form
\begin{equation}
  \Xi_N^{(u)}(\lambda) = \left[ \frac{(2\pi\sigma^2)^{-u} \pi^{M(u+1)} N^{u+1}}{u+1} \right]^{\frac{1}{2}} \sum_{k=1}^K \sum_{l=1}^{L_k} \omega_{k,l} \times E_{x_{[a]}} \left\{ \mathbb{E}_{y \sim g_M(v|0,Q_{k,l}^{(u)})} \left\{ e^{\frac{1}{2} y^T \Sigma y} \right\} \right\},
\end{equation}
where $\Sigma = (\beta/\sigma^2) [I_u - 1_u 1_u^T/(1+u)] \in \mathbb{R}^{(u+1) \times (u+1)}$. Using (15) to integrate over the Gaussian RV $v$ in (39) yields
\begin{equation}
  \Xi_N^{(u)}(\lambda) = \sum_{k=1}^K \sum_{l=1}^{L_k} \omega_{k,l} E_{x_{[a]}} \left\{ e^{N \beta^{-1} G^{(u)}(Q_{k,l}^{(u)})} \right\},
\end{equation}
where
\begin{equation}
  e^{G^{(u)}(Q_{k,l}^{(u)})} = \sqrt{\frac{(2\pi\sigma^2)^{-u}}{(1+u) \det(I + \Sigma Q_{k,l}^{(u)})}}.
\end{equation}
To compute the expectations w.r.t. $\{x_{[a]}\}_{a=0}^u$ for all $k = 1, \ldots, K$, $l = 1, \ldots, L_k$ in (40), we write the measure of the

![Graph](image_url)

Fig. 1. Per-component MSE for the CS of block sparse signal. Solid lines are obtained from Proposition 2 with $\delta^2 \to 0^+$ and markers depict numerical Monte Carlo simulations where $\sigma^2 = 1$, $\delta^2 = 10^{-9}$ and $N = 1200$.

**Remark 2.** As $\delta^2 \to 0^+$, the noise variance (32) becomes the Tse-Hanly solution [7] for equal power users but with a modified user load $\beta_k = \frac{k}{K} \beta$. The same noise variance is obtained by a genie-aided MMSE receiver, conditioned on the event that $x$ is sampled from one of the $L_k$ mixtures indexed by $k$. The MSE (33), on the other hand, is a summation of the related MSEs but weighted with the probability of having $k$ non-zero blocks in a realization of the source vector $x$. Thus, there is no loss in not knowing the positions of the zero blocks in advance if we use the optimum MMSE receiver for very large $K$ block sparse systems. Note that the equivalent AWGN channel model in Sec. III-A already implies this point. For practical settings with finite sized sensing matrices, however, this model does not strictly hold.

**Corollary 1.** The MMSE estimator (11) has the same MSE in the large system limit as a genie-aided MMSE estimator that knows in advance the positions of the non-zero blocks in $x$.

To empirically verify the analytical results, we have plotted in Fig. 1 the MSE of estimator (11), obtained via computer simulations. The theoretical MSE given in Proposition 2 is given as well. In all simulation cases we have set $\omega_{k,l} = \omega = 1/\sum_{k=1}^K L_k$ so that $\omega_k = \omega L_k$. For the selected cases the theory matches Monte Carlo simulations very well.

**IV. CONCLUSIONS**

Minimum mean square error estimation of block sparse signals from noisy linear measurements was considered. The main result of the paper is the closed-form MMSE for the CS of such signals. The solution turned out to be of a particularly simple form, namely, the Tse-Hanly formula with a scaling by parameters that depend on the sparsity pattern of the source. The result implies that if the statistics of the block sparse CS problem are known, the MMSE is independent of the knowledge about the positions of the non-vanishing blocks.

**APPENDIX**

**DERIVATION OF CLAIM 1**

Let us assume that the free energy density (29) is self-averaging w.r.t. the quenched randomness $(A, y)$ in the large
matrix $\mathbf{Q}^{(u)}_{k,l}$ as

$$
\mu_N(\mathbf{Q}^{(u)}_{k,l}) = E_{\{x_{[a]}}\{e^{X^T x_{[a]} - NQ^{(u)}_{k,l}} \delta(x^T x_{[b]} = NQ^{(u)}_{k,l}) \right\}, \quad (42)
$$

and integrate w.r.t. (42). Writing the Dirac measures in terms of (inverse) Laplace transform and invoking saddle point integration (see Appendix A for details), we get

$$
\Xi^{(u)}_N(\lambda) = \sum_{k=1}^{K} \sum_{l=1}^{L_k} \omega_{k,l} \sum_{n=1}^{N} \mathbf{x}^{(u)}_{n} e^{N T^{(u)}_{k,l}(\lambda)}, \quad (43)
$$

where $\hat{\mathbf{Q}}^{(u)}_{k,l} \in \mathbb{R}^{(u+1) \times (u+1)}$ is a symmetric matrix. To obtain (43), we defined an auxiliary function

$$
T^{(u)}_{k,l}(\lambda) = \inf_{\hat{\mathbf{Q}}^{(u)}_{k,l}} \left\{ \beta^{-1} G^{(u)}(\mathbf{Q}^{(u)}_{k,l}) \right\} - \inf_{\hat{\mathbf{Q}}^{(u)}_{k,l}} \left\{ \text{tr}(\mathbf{Q}^{(u)}_{k,l} \hat{\mathbf{Q}}^{(u)}_{k,l}) \right\}, \quad (44)
$$

where

$$
G^{(u)}(p_{k,l}, q_{k,l}) = E_{\{x_{[a]}}\left\{ e^{X^T x_{[a]} - N \mathbf{x}^{(u)}_{n} e^{N T^{(u)}_{k,l}(\lambda)}} \right\}, \quad (45)
$$

denotes the moment generating function (MGF) of (42). We also wrote $\mathbf{x}^{(u)}_{n} = [x_{[a]}_{1,n} \cdots x_{[a]}_{u,n}] \in \mathbb{R}^{(u+1), n} = r \in \{r = 1, \ldots, R \}$ and $q = 1, \ldots, Q$ in the notation of (3).

To make the optimization problems in (44) tractable, we assume that their solutions are the replica symmetric (RS) matrices (see, e.g., [8]–[13] on discussion about this assumption)

$$
\mathbf{Q}^{(u)}_{k,l} = (p_{k,l} - q_{k,l}) \mathbf{I}_{u+1} + q_{k,l} \mathbf{I}_{u+1}^T, \quad (46)
$$

$$
\mathbf{Q}^{(u)}_{k,l} = (p_{k,l} - q_{k,l}) \mathbf{I}_{u+1} + q_{k,l} \mathbf{I}_{u+1}^T, \quad (47)
$$

respectively, where $p_{k,l}, q_{k,l}, \hat{p}_{k,l}, \hat{q}_{k,l}$ are real parameters. Under the RS assumption, we get the simplifications

$$
\text{tr}(\mathbf{Q}^{(u)}_{k,l} \mathbf{Q}^{(u)}_{k,l}) = (u+1)(p_{k,l}^{2} + aq_{k,l}^{2}), \quad (48)
$$

$$
\frac{u}{q} \rightarrow 0, \quad (49)
$$

and

$$
G^{(u)}(p_{k,l}, q_{k,l}) = -\frac{u}{2} \log[a^2 + \beta(p_{k,l} - q_{k,l})] \quad -\frac{u}{2} \log[2\pi a^2] \quad = \frac{u}{2} \log(\frac{u}{q} + 1), \quad (50)
$$

$$
\frac{u}{q} \rightarrow 0, \quad (51)
$$

From the first extremum in (44) one obtains $\hat{p}_{k,l} = 0$ and

$$
\hat{q}_{k,l} = [\sigma^2 + \beta(p_{k,l} - q_{k,l})]^{-1}, \quad (52)
$$

where $p_{k,l}$ and $q_{k,l}$ are left as arbitrary but fixed parameters for now. To proceed with the second optimization problem in (44), we need to evaluate the MGF (45) under the RS assumption.

Using (15) right-to-left in (45), using the RS assumption (46) – (47) and recalling that the replicas $\{x_{[a]}\}_{a=0}^{u}$ are IID yields after some algebra

$$
\phi^{(u)}_{k,l}(\hat{q}_{k,l}, \lambda; N) = C^{(u)}_{N}(\hat{q}_{k,l}) \int \mathbf{E}_{x} \left\{ g_{N}(x_{k,l} \mid x; \hat{q}_{k,l}^{-1} \mathbf{I}_{N}) \right\} \mathbf{E}_{\hat{x}} \left\{ e^{N \lambda g_{N}(x_{k,l} \mid x; \hat{q}_{k,l}^{-1} \mathbf{I}_{N})} \right\} \mathbf{d}z_{k,l}, \quad (53)
$$

where the expectations $E^{(k,l)}$ are w.r.t. zero-mean Gaussian RVs with covariance $D_{k,l}$. The normalization factor $C^{(u)}_{N}(\hat{q}_{k,l}) = [(1+u)(2\pi \hat{q}_{k,l}^{-1})]^{N/2}$ is due the introduction of the Gaussian densities in (53). Since

$$
\phi^{(u)}_{k,l}(\hat{q}_{k,l}, \lambda; N) \overset{u \rightarrow 0}{\longrightarrow} 1, \quad (54)
$$

the second optimization in (44) reduces to the conditions

$$
p_{k,l} = \lim_{N \rightarrow \infty} N^{-1} E^{(k,l)} \left\{ ||x||^2 \right\}, \quad (55)
$$

$$
q_{k,l} = \lim_{N \rightarrow \infty} N^{-1} E^{(k,l)} \left\{ ||\hat{x}||^2 \right\}, \quad (56)
$$

where $\lambda = 0$ and $u \rightarrow 0$. The expectations in (55) and (56) are w.r.t. $z_{k,l}$, and the independent zero-mean Gaussian RVs $x, \hat{x}$ with covariance $D_{k,l}$ as in (53). We also write

$$
\rho_{N}(x_{k,l}) = E_{\hat{x}} \left\{ g_{N}(x_{k,l} \mid \hat{x}; \hat{q}_{k,l}^{-1} \mathbf{I}_{N}) \right\}, \quad (57)
$$

so that

$$
\langle \hat{x} \rangle_{p}^{(k,l)} = \frac{1}{\rho(x_{k,l})} E_{\hat{x}} \left\{ \mathbf{E} \left\{ \frac{1}{\rho(x_{k,l})} g_{N}(x_{k,l} \mid \hat{x}; \hat{q}_{k,l}^{-1} \mathbf{I}_{N}) \right\} \right\}. \quad (58)
$$

Note that (58) is the MMSE estimator of the Gaussian channel $z_{k,l} = x + \eta_{k,l} \in \mathbb{R}^{N}$, when the receiver knows the correct distributions of $\eta_{k,l} \sim g_{N}(\eta \mid 0; \hat{q}_{k,l}^{-1} \mathbf{I}_{N})$ and $x \sim g_{N}(x \mid 0; \mathbf{D}_{k,l})$. Furthermore, from (55) and (56) we get

$$
\text{mse}_{\text{eq}}(p_{k,l})(\hat{q}_{k,l}) = \lim_{N \rightarrow \infty} \frac{1}{N} E^{(k,l)} \left\{ ||x||^2 \right\} - E^{(k,l)} \left\{ ||\hat{x}||^2 \right\} \quad (59)
$$

$$
\text{mse}_{\text{eq}}(p_{k,l})(\hat{q}_{k,l}) = p_{k,l} - q_{k,l}, \quad (60)
$$

where $\text{mse}_{\text{eq}}(p_{k,l})$ is the MMSE of the Gaussian channel (52).

The free energy density under the RS assumption reads

$$
f_{\text{free}} = \lim_{u \rightarrow 0} \frac{\partial}{\partial u} \lim_{N \rightarrow \infty} \frac{1}{N} \log \left( \sum_{k=1}^{K} \sum_{l=1}^{L_k} \omega_{k,l} e^{N T^{(u)}_{k,l}(\lambda)} \right) \quad (61)
$$

Switching the order of the limits once more yields

$$
f_{\text{free}} = \lim_{N \rightarrow \infty} \frac{1}{N} \lim_{u \rightarrow 0} \left\{ \sum_{k=1}^{K} \sum_{l=1}^{L_k} \omega_{k,l} e^{N T^{(u)}_{k,l}(\lambda)} \right\}^{-1} \quad \times \left[ \sum_{k=1}^{K} \sum_{l=1}^{L_k} \omega_{k,l} \frac{\partial}{\partial u} e^{N T^{(u)}_{k,l}(\lambda)} \right]. \quad (62)
$$
by (5) the denominator becomes just unity and can be omitted. For the latter part,
\[ \frac{\partial}{\partial u} N T_k (\lambda) = N e^{N T_k (\lambda)} \left( \frac{\partial}{\partial u} T_k (\lambda) \right), \]
where the derivative is assessed
\[ \lim_{u \to 0} \frac{\partial}{\partial u} T_k (\lambda) = - \frac{1}{2} \log [\sigma^2 + \beta (p_{k,l} - q_{k,l})] - q_{k,l} \tilde{q}_{k,l} + \frac{1}{2} \left[ 1 + \log (2\pi \tilde{q}_{k,l}^2) \right] - \frac{1}{2\beta} \log (2\pi \sigma^2) + \lim_{N \to \infty} \frac{1}{N} \int p_N (z_{k,l}) h_N (z_{k,l}; \lambda, \tilde{q}_{k,l}) dz_{k,l}. \]
Note that we defined above the function
\[ h_N (z_{k,l}; \lambda, \tilde{q}_{k,l}) = \log \left( E_x \{ e^{X \lambda} \tilde{g}_N (z_{k,l} | \tilde{x}; \tilde{q}_{k,l} I N) \} \right), \]
for notational convenience. Recalling (63), we finally have the RS free energy density
\[ f_{\text{free}} = \frac{1}{2} \left[ 1 - \beta^{-1} \log (2\pi \sigma^2) \right] + \frac{1}{2} \sum_{k=1}^K \sum_{l=1}^{L_k} \omega_{k,l} \left[ \log (2\pi \tilde{q}_{k,l}) + \beta^{-1} \log \tilde{q}_{k,l} - 2q_{k,l} \tilde{q}_{k,l} \right] + \lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^K \sum_{l=1}^{L_k} \omega_{k,l} \int p_N (z_{k,l}) h_N (z_{k,l}; \lambda, \tilde{q}_{k,l}) dz_{k,l}. \]
where only the last term depends on \( \lambda \) and is relevant for the assessment of the MSE, as given in (60).

The final task is to compute \( \nabla^2 h_N (z_{k,l}, \lambda, \xi^2) \big|_{\xi=0} \). First,
\[ \nabla^2 h_N (z_{k,l}, \lambda, \tilde{q}_{k,l}) = \left. \frac{\partial^2}{\partial \lambda^2} h_N (z_{k,l}, \lambda, \xi^2) \right|_{\lambda=0}, \]
so that the estimator (58) can also be written as
\[ \left( \tilde{x}_{p}^{(k,l)} \right) = \nabla h_N (z_{k,l}; \lambda, \tilde{q}_{k,l}) \big|_{\lambda=0}. \]
Proceeding similarly, after a bit of algebra we obtain the conditional covariance matrix of the error
\[ E_N (z_{k,l}) = \nabla^2 h_N (z_{k,l}; \lambda, \tilde{q}_{k,l}) \big|_{\lambda=0} = \left( \tilde{x} \tilde{x}^T \right)_p - \left( \tilde{x} \right)_p \left( \tilde{x} \right)_p^T, \]
which is also the error covariance of the estimator (58). Thus, by (50), the per-component MSE of the original MMSE estimator given in Proposition 1 reads
\[ \text{mse} = \sum_{k=1}^K \sum_{l=1}^{L_k} \omega_{k,l} \lim_{N \to \infty} \frac{1}{N} \text{tr} \left( \int p_N (z_{k,l}) E_N (z_{k,l}) dz_{k,l} \right) \]
which can be written due to (55), (56) and (60) as
\[ \text{mse} = \sum_{k=1}^K \sum_{l=1}^{L_k} \omega_{k,l} \text{mse}_{eq}^{(k,l)} (\tilde{q}_{k,l}), \]
completing the proof.