

Precise Single Letter Characterization

Dror Baron

This supplement provides more details about a precise single letter characterization, which was published by the instructor with Dongning Guo and Shlomo Shamai in 2009. The presentation here is greatly simplified and intended only to provide intuition.

Problem setting: Recall that we are considering

$$y = \sqrt{\gamma}\Phi x + z,$$

where the input $x \in \mathbb{R}^N$ has sparsity ϵ and is generated by distribution P_U when nonzero; the matrix $\Phi \in \mathbb{R}^{M \times N}$ has unit norm columns, on average; $z \in \mathbb{R}^M$ is zero mean unit norm additive white Gaussian noise; γ is the signal to noise ratio (SNR); and $y \in \mathbb{R}^M$ are the noisy measurements. Recall that we consider the *large system limit*, where M, N are both growing to infinity, and their ratio satisfies

$$\delta = \lim_{N \rightarrow \infty} \frac{M}{N} > 0.$$

Hypothetical case: Suppose that we are interested in estimating x_n . Recall that the signal, matrix, and noise are all independent and identically distributed (i.i.d.). They are also independent of one another, and so any x_n could be considered analogous with other entries of the input x , in the sense that they all have the same impact on the overall estimation task at hand.

We will hypothetically assume that the rest of x is known *precisely*, and denote the rest of x by $x_{\setminus n}$, where $\setminus n$ denotes all indices except for n . An actual recovery algorithm is unlikely to know $x_{\setminus n}$ precisely, and so any estimation theoretic result we can derive for x_n in this hypothetical scenario is an optimistic upper bound.

We now have

$$y = \sqrt{\gamma}[\Phi_n \Phi_{\setminus n}][x_n x_{\setminus n}]^T + z,$$

where Φ_n is column n of the matrix, $\Phi_{\setminus n}$ denotes all columns in the matrix except for column n , and \cdot^T denotes matrix transpose, where we note in passing that in our case $[x_n x_{\setminus n}]$ is a row vector, which is a matrix of dimension $1 \times N$.

Because we know $x_{\setminus n}$, we can subtract off its effect on y , yielding modified measurements,

$$\begin{aligned} \tilde{y}_n &:= y - \sqrt{\gamma}\Phi_{\setminus n}x_{\setminus n} \\ &= [\sqrt{\gamma}[\Phi_n \Phi_{\setminus n}][x_n x_{\setminus n}]^T + z] - [\sqrt{\gamma}\Phi_{\setminus n}x_{\setminus n}] \\ &= \sqrt{\gamma}\Phi_n x_n + z. \end{aligned}$$

We now have a *vector estimation problem*. That is, $\tilde{y}_n \in \mathbb{R}^M$ is a vector of noisy measurements, and we want to estimate the scalar x_n from this vector. To do so, it is well known in communication and signal processing that the optimal approach involves a *matched filter*, which multiplies \tilde{y}_n by $(\Phi_n)^T$,

$$\begin{aligned}\bar{y}_n &:= (\Phi_n)^T \tilde{y}_n \\ &= (\Phi_n)^T [\sqrt{\gamma} \Phi_n x_n + z] \\ &= \sqrt{\gamma} (\Phi_n)^T \Phi_n x_n + (\Phi_n)^T z.\end{aligned}$$

Note that we now have a scalar estimation problem, instead of a vector estimation problem.

Recall that the matrix Φ has unit norm columns, on average. Therefore, $(\Phi_n)^T \Phi_n \approx 1$, and so \bar{y}_n is well approximated as follows,

$$\bar{y}_n \approx \sqrt{\gamma} x_n + (\Phi_n)^T z.$$

This form is often called a *scalar channel* estimation problem. That is, we have a scalar x_n , and we observe it through a channel, which is a noise-adding mechanism. It is convenient to think of this scalar channel as follows: (i) begin with the scalar input x_n ; (ii) modulate it by $\sqrt{\gamma}$, the square root of the SNR; (iii) add noise, $(\Phi_n)^T z$. We remind the reader that $(\Phi_n)^T \Phi_n$ is not exactly 1, and one could view this as another noise mechanism, but this imprecision vanishes in the large system limit.

Next, we analyze the distribution of

$$\bar{z}_n = (\Phi_n)^T z.$$

Recall that z is comprised of M i.i.d. zero mean unit norm Gaussian random variables. To compute the distribution of \bar{z}_n , note that any linear transformation of a Gaussian random vector (GRVec) is also a GRVec, which is precisely characterized by its mean vector and covariance matrix. Next, $\bar{z}_n = (\Phi_n)^T z \in \mathbb{R}$ is a scalar, because we have an inner product between Φ_n and z . Therefore, \bar{z}_n is a Gaussian random variable, and its distribution is completely determined by its expected value and variance.

To establish the distribution of \bar{z}_n , we will compute its expected value and variance. The expected value is relatively simple to compute,

$$E[\bar{Z}_n] = (\Phi_n)^T E[Z] = 0,$$

where we note in passing that random variables and vectors are denoted by capital letters. Because the expected value is zero, the variance is $E[\bar{Z}_n^2]$,

$$\begin{aligned}\text{Var}(\bar{Z}_n) &= E[\bar{Z}_n^2] \\ &= E[(\Phi_n^T Z)^2] \\ &= E\left[\left(\sum_{m=1}^M \Phi_{mn} Z_m\right)^2\right]\end{aligned}$$

$$\begin{aligned}
&= E \left[\sum_{m_1, m_2=1}^M \Phi_{m_1 n} \Phi_{m_2 n} Z_{m_1} Z_{m_2} \right] \\
&= \sum_{m_1, m_2=1}^M \Phi_{m_1 n} \Phi_{m_2 n} E[Z_{m_1} Z_{m_2}].
\end{aligned}$$

The Rvec Z is i.i.d. with zero mean and unit variance, and so $E[Z_{m_1} Z_{m_2}]$ is 0 when $m_1 \neq m_2$, else 1. Therefore,

$$\text{Var}(\bar{Z}_n) = \sum_{m=1}^M (\Phi_{mn})^2 = \|\Phi_n\|_2^2,$$

which is approximately 1. Therefore, \bar{Z}_n is a zero mean unit norm Gaussian random variable, i.e., $\bar{Z}_n \sim \mathcal{N}(0, 1)$. We conclude that

$$\bar{y} \approx \sqrt{\gamma} x_n + \mathcal{N}(0, 1).$$

Again, the imprecision in this approximation vanishes in the large system limit.

How well can we estimate x_n from \bar{y} ? To keep things simple, suppose that we are interested in minimizing the mean squared error (MSE) between \hat{x}_n , our estimate for x_n , and the original x_n . It is well known that the MSE is minimized by computing the conditional expectation, $E[X_n | \bar{Y}_n]$. The minimum mean squared error (MMSE) is the best-possible MSE, obtained using conditional expectation,

$$\text{MMSE}(P_U, \gamma, \epsilon) = E[(X_n - E[X_n | \bar{Y}_n])^2].$$

We highlight that the MMSE is a function of (i) P_U , the distribution of nonzero elements of x ; (ii) ϵ is the sparsity rate, i.e., $\epsilon = \Pr(X_n \neq 0)$; and (iii) γ is the SNR.

Realistic case: The hypothetical case assumed that $x_{\setminus n}$ is known precisely. However, in a realistic setting it is only known imprecisely. We have an estimate, $\hat{x}_{\setminus n}$, which is comprised of the true signal and additive estimation error,

$$\hat{x}_{\setminus n} = x_{\setminus n} + e_{\setminus n}.$$

We can now revisit our modified measurements,

$$\begin{aligned}
\tilde{y}_n &= y - \sqrt{\gamma} \Phi_{\setminus n} \hat{x}_{\setminus n} \\
&= y - \sqrt{\gamma} \Phi_{\setminus n} [x_{\setminus n} + e_{\setminus n}] \\
&= \sqrt{\gamma} \Phi_n x_n - \sqrt{\gamma} \Phi_{\setminus n} e_{\setminus n} + z.
\end{aligned}$$

Applying the matched filter yields

$$\begin{aligned}
\bar{y}_n &:= (\Phi_n)^T \tilde{y}_n \\
&= (\Phi_n)^T [\sqrt{\gamma} \Phi_n x_n - \sqrt{\gamma} \Phi_{\setminus n} e_{\setminus n} + z] \\
&= \sqrt{\gamma} (\Phi_n)^T \Phi_n x_n - \sqrt{\gamma} (\Phi_n)^T \Phi_{\setminus n} e_{\setminus n} + \bar{z}_n.
\end{aligned}$$

We have already discussed how $\sqrt{\gamma}(\Phi_n)^T \Phi_n x_n \approx \sqrt{\gamma}x_n$ and $\bar{z}_n \sim \mathcal{N}(0, 1)$. What about the new term,

$$\psi = (\Phi_n)^T \Phi_{\setminus n} e_{\setminus n}?$$

Because $\Phi_{\setminus n} \in \mathbb{R}^{M \times (N-1)}$ and $e_{\setminus n} \in \mathbb{R}^{N-1}$, $\Phi_{\setminus n} e_{\setminus n} \in \mathbb{R}^M$. Moreover, $(\Phi_n)^T \in \mathbb{R}^{1 \times M}$, implying that ψ is a scalar,

$$\psi = (\Phi_n)^T \Phi_{\setminus n} e_{\setminus n} = \sum_{m=1}^N \sum_{n' \neq n} \Phi_{mn} \Phi_{mn'} e_{n'}.$$

Recall that Φ is i.i.d., and its elements have zero mean and variance $\frac{1}{M}$. Therefore, the central limit theorem (CLT) can be used to show that the above summation multiplies each $e_{n'}$ by a Gaussian random variable whose mean is zero and variance is M times the product of the variances of Φ_{mn} and $\Phi_{mn'}$, and $M \frac{1}{M} \frac{1}{M} = \frac{1}{M}$. To summarize, the CLT yields

$$\sum_{m=1}^N \Phi_{mn} \Phi_{mn'} \sim \mathcal{N}\left(0, \frac{1}{M}\right).$$

As mentioned earlier, the i.i.d. nature of our estimation problem implies that $e_{n'}$ has the same distribution for all n . This distribution must have zero mean (else the estimator is easily improved by subtracting off this nonzero bias term), and it has some MSE, which is *equal to the MMSE of the CS problem*. We have $N - 1$ products of $e_{n'}$ with terms of the form $\sum_{m=1}^M \Phi_{mn} \Phi_{mn'} \sim \mathcal{N}\left(0, \frac{1}{M}\right)$. Because $N - 1$ is large in the large system limit, the CLT shows that

$$\psi \sim \mathcal{N}\left(0, \frac{N}{M} MMSE\right) = \mathcal{N}\left(0, \frac{1}{\delta} MMSE\right).$$

We now have

$$\begin{aligned} \bar{y}_n &= \sqrt{\gamma}(\Phi_n)^T \Phi_n x_n - \sqrt{\gamma}(\Phi_n)^T \Phi_{\setminus n} e_{\setminus n} + \bar{z}_n \\ &= \sqrt{\gamma}x_n - \sqrt{\gamma}\psi + \mathcal{N}(0, 1) \\ &= \sqrt{\gamma}x_n + \mathcal{N}\left(0, \frac{\gamma}{\delta} MMSE\right) + \mathcal{N}(0, 1) \\ &= \sqrt{\gamma}x_n + \mathcal{N}\left(0, 1 + \frac{\gamma}{\delta} MMSE\right), \end{aligned}$$

where we note in passing that multiplication by $\sqrt{\gamma}$ multiplies the standard deviation of ψ by $\sqrt{\gamma}$, which corresponds to multiplying the variance by $(\sqrt{\gamma})^2 = \gamma$. Instead of unit norm Gaussian noise, our more realistic and less hypothetical scalar channel has lower SNR. Instead of γ , the new SNR is $\gamma/(1 + \frac{\gamma}{\delta} MMSE)$, which we denote by $\gamma\eta$. That is,

$$\eta = \frac{1}{1 + \frac{\gamma}{\delta} MMSE} \in (0, 1)$$

is the *degradation* of the scalar channel.

Finally, earlier we expressed the MMSE as a function of P_U , ϵ , and γ . The MMSE is now a function of P_U , ϵ , and $\gamma\eta$, and η follows Tanaka's celebrated fixed point equation,

$$\frac{1}{\eta} = 1 + \frac{\gamma}{\delta} MMSE(P_U, \gamma\eta, \epsilon).$$