

## Characterization of Minimum Error Linear Coding with Sensory and Neural Noise

**Eizaburo Doi**

*edoi@cns.nyu.edu*

*Center for Neural Science, New York University, New York, NY 10003, U.S.A.*

**Michael S. Lewicki**

*michael.lewicki@case.edu*

*Electrical Engineering and Computer Science Department, Case Western Reserve University, Cleveland, OH 44106, U.S.A.*

Robust coding has been proposed as a solution to the problem of minimizing decoding error in the presence of neural noise. Many real-world problems, however, have degradation in the input signal, not just in neural representations. This generalized problem is more relevant to biological sensory coding where internal noise arises from limited neural precision and external noise from distortion of sensory signal such as blurring and phototransduction noise. In this note, we show that the optimal linear encoder for this problem can be decomposed exactly into two serial processes that can be optimized separately. One is Wiener filtering, which optimally compensates for input degradation. The other is robust coding, which best uses the available representational capacity for signal transmission with a noisy population of linear neurons. We also present spectral analysis of the decomposition that characterizes how the reconstruction error is minimized under different input signal spectra, types and amounts of degradation, degrees of neural precision, and neural population sizes.

### 1 Introduction ---

We address the problem of how to linearly transform  $N$ -dimensional inputs into  $M$ -dimensional representations in order to best transmit signal information through noisy gaussian channels when the input signal is degraded. Here the degradation is modeled by additive white gaussian noise as well as a linear distortion such as blurring. Most of the earlier studies have addressed the problem without such degradations (for a review, see Palomar & Jiang, 2007), although they are inevitable in almost any real system. It is therefore important to characterize and understand its impact on the optimal encoder. The main result of this note is that under a minimum mean squared error (MMSE) objective, the optimal linear transform can be

decomposed exactly into optimal denoising and debarring (Wiener filtering), and optimal coding for a noisy gaussian channel (robust coding; Lee, 1975; Lee and Petersen, 1976; Doi, Balcan, & Lewicki, 2007). Such decomposition was shown earlier in the context of source coding and quantization (Dobrushin & Tsybakov, 1962; Sakrison, 1968) and proven in a general case (Wolf & Ziv, 1970). Although it has been extensively studied (for a review, see Gray & Neuhoff, 1998), a unified treatment in the context of MMSE linear coding and its characterization has not been provided. We also offer an alternative proof of the decomposition in the course of characterizing the solution.

A special case of this problem was examined previously in which the linear transform was assumed to be convolutional, implying that the channel (or neural) dimension is equal to the input (or sensory) dimension and that individual coding units are shifted versions of each other with the identical filter shape (Ruderman, 1994). Those simplifying assumptions have commonly been made in the study of optimal sensory coding, because it matches reasonably well in a certain sensory system (e.g., foveal retina) and it is analytically more tractable (Atick & Redlich, 1990, 1992; Atick, Li, & Redlich, 1990; van Hateren, 1992). Also, an approximate decomposition along similar lines has been used under an information maximization objective (Atick & Redlich, 1992; Atick, 1992; Dayan & Abbott, 2001). Here, we assume no such restrictions and provide a general solution with the exact decomposition, in which the channel dimension may be smaller than, equal to, or larger than the input dimension (undercomplete, complete, or overcomplete representations, respectively), and individual coding units are not restricted to having the same filter shape.

## 2 Problem Formulation

---

The model system consists of three processes (see Figure 1a): generation of the observed signal (see equation 2.1), encoding (see equation 2.2), and decoding (see equation 2.3). The observed signal  $\mathbf{x} \in \mathbb{R}^N$  is a degraded version of the original signal  $\mathbf{s} \in \mathbb{R}^N$  via a fixed linear transform followed by the additive white gaussian noise (AWGN),

$$\mathbf{x} = \mathbf{H}\mathbf{s} + \mathbf{v}, \quad (2.1)$$

where  $\mathbf{s}$  is zero mean with covariance  $\Sigma_{\mathbf{s}}$ ,  $\mathbf{H} \in \mathbb{R}^{N \times N}$  is a fixed linear distortion, and  $\mathbf{v} \sim \mathcal{N}(\mathbf{0}_N, \sigma_v^2 \mathbf{I}_N)$  is sensory (or input) noise. Note that signal  $\mathbf{s}$  may not necessarily be gaussian distributed. Encoding is assumed to be a linear transform into an arbitrary dimension, and the resulting representations may be undercomplete, complete, or overcomplete,

$$\mathbf{r} = \mathbf{W}\mathbf{x} + \delta, \quad (2.2)$$

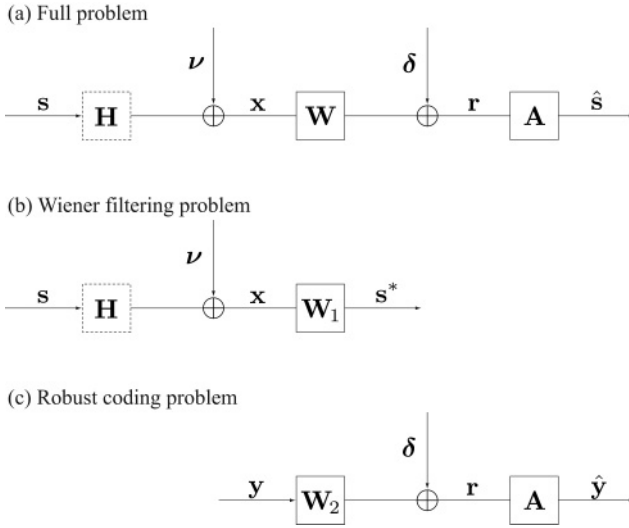


Figure 1: Model diagrams. (a) The generalized problem. Both encoder  $\mathbf{W}$  and decoder  $\mathbf{A}$  are optimized. (b) In the first subproblem, signal degradation is caused by a fixed linear transform and AWGN. The encoder  $\mathbf{W}_1$  is optimized. The solution is given by Wiener filtering. (c) In the second subproblem, signal transmission is limited by both AWGN in the channel and by the limited channel dimension, and both encoder  $\mathbf{W}_2$  and decoder  $\mathbf{A}$  are optimized. The solution is given by robust coding.

with  $\mathbf{W} \in \mathbb{R}^{M \times N}$  the linear encoder,  $\delta \sim \mathcal{N}(\mathbf{0}_M, \sigma_\delta^2 \mathbf{I}_M)$  neural (channel, or output) noise, and  $\mathbf{r}$  the representation. Decoding is assumed to be linear:

$$\hat{\mathbf{s}} = \mathbf{A}\mathbf{r}, \quad (2.3)$$

where  $\mathbf{A} \in \mathbb{R}^{N \times M}$  is the linear decoder and  $\hat{\mathbf{s}}$  is the estimate of the original signal.

The decoding error  $\boldsymbol{\epsilon}$  is

$$\boldsymbol{\epsilon} = \mathbf{s} - \hat{\mathbf{s}}, \quad (2.4)$$

and the mean squared error (MSE)  $\mathcal{E}$  is

$$\mathcal{E} = \langle \boldsymbol{\epsilon}^T \boldsymbol{\epsilon} \rangle \quad (2.5)$$

$$= \text{tr} \langle \boldsymbol{\epsilon} \boldsymbol{\epsilon}^T \rangle \quad (2.6)$$

$$= \text{tr}(\Sigma_s) - 2\text{tr}(\mathbf{A}\mathbf{W}\mathbf{H}\Sigma_s) + \text{tr}(\mathbf{A}\mathbf{W}\Sigma_x\mathbf{W}^T\mathbf{A}^T) + \sigma_\delta^2\text{tr}(\mathbf{A}\mathbf{A}^T), \quad (2.7)$$

where  $\langle \cdot \rangle$  is the sample average and  $\Sigma_x = \mathbf{H}\Sigma_s\mathbf{H}^T + \sigma_v^2\mathbf{I}_N$  is the covariance of the observations  $\mathbf{x}$ .

We are interested in minimizing the MSE function subject to one of the following power constraints:

$$\text{tr}(\mathbf{W}\Sigma_x\mathbf{W}^T) = \mathcal{P}, \quad (2.8)$$

$$\text{diag}(\mathbf{W}\Sigma_x\mathbf{W}^T) = \frac{\mathcal{P}}{M}, \quad (2.9)$$

where equation 2.8 is to constrain the sum of the variances of filter outputs (referred to as the total power constraint), while equation 2.9 is to constrain the individual variance (the individual power constraint; this is a sufficient condition to satisfy the total power constraint; Lee, 1975; Lee & Petersen, 1976).

The problem is to find  $\mathbf{W}$  and  $\mathbf{A}$  that minimize  $\mathcal{E}$  subject to one of the power constraints above.

### 3 Results

---

**3.1 Optimal Linear Decoder.** The linear decoder  $\mathbf{A}$  can be expressed in terms of  $\mathbf{W}$ , and the optimization can be reduced to finding solely the optimal  $\mathbf{W}$ . This is derived from the necessary condition of minimum MSE with respect to  $\mathbf{A}$ :

$$\frac{\partial \mathcal{E}}{\partial \mathbf{A}} = 0, \quad (3.1)$$

$$\Leftrightarrow \mathbf{A}_{opt} = \Sigma_s \mathbf{H}^T \mathbf{W}^T (\mathbf{W}\Sigma_x\mathbf{W}^T + \sigma_\delta^2 \mathbf{I}_M)^{-1}. \quad (3.2)$$

This result was shown previously for a special case in which  $\mathbf{W}$  is convolutional (Ruderman, 1994).

**3.2 Exact Decomposition.** Our main result is based on the observation that the minimum MSE can be decomposed as follows (see Wolf & Ziv, 1970, for a general case):

$$\mathcal{E}_{opt} = (\text{Wiener filtering error}) + (\text{robust coding error}). \quad (3.3)$$

The Wiener filtering error is the theoretical error bound for a linear method that best counteracts input degradation caused by blurring and AWGN and is achieved by the Wiener filtering (see Figure 1b). The robust coding error is the theoretical bound for a linear method that best utilizes noisy gaussian channels of a limited dimension (neural population size) and is achieved by robust coding (Lee, 1975; Lee & Petersen, 1976; Doi et al., 2007; see Figure 1c). If robust coding is solved in the setting under which the input signal  $\mathbf{y}$  is the Wiener estimate  $\mathbf{s}^*$  (and hence the reconstruction by robust coding  $\hat{\mathbf{y}} = \hat{\mathbf{s}}^*$ ), then it can be shown that the product of Wiener filtering and robust coding,  $\mathbf{W}_2\mathbf{W}_1$ , is the optimal linear transform for the generalized problem (see Figure 1).

**Proof of Equation 3.3.** We analyze the problem in the whitened signal space because this simplifies the formulas. We rewrite the covariance of the observed signal  $\mathbf{x}$  using the eigenvalue decomposition,

$$\Sigma_{\mathbf{x}} = \mathbf{E} \mathbf{D}_{\mathbf{x}} \mathbf{E}^T, \quad (3.4)$$

and also rewrite the linear encoder as

$$\mathbf{W} = \sqrt{\frac{\mathcal{P}}{M}} \mathbf{V} \mathbf{D}_{\mathbf{x}}^{-\frac{1}{2}} \mathbf{E}^T, \quad (3.5)$$

where  $\mathbf{D}_{\mathbf{x}}^{-\frac{1}{2}} \mathbf{E}^T$  whitens the observed signal  $\mathbf{x}$ , and  $\mathbf{V}$  is the linear encoder with respect to the whitened signal. Note that the inverse of  $\mathbf{D}_{\mathbf{x}}$  always exists because of the nonzero input noise  $\mathbf{v}$ .

The first benefit of this parameterization can be seen in the constraint functions equations 2.8 and 2.9, that are simplified as

$$\text{tr}(\mathbf{V} \mathbf{V}^T) = M, \quad (3.6)$$

$$\text{diag}(\mathbf{V} \mathbf{V}^T) = \mathbf{1}_M. \quad (3.7)$$

Also, the decoder, equation 3.2, is simplified as:

$$\mathbf{A} = \sqrt{\frac{\mathcal{P}}{M}} \Sigma_s \mathbf{H}^T \mathbf{E} \mathbf{D}_{\mathbf{x}}^{-\frac{1}{2}} \mathbf{V}^T \left[ \frac{\mathcal{P}}{M} \mathbf{V} \mathbf{V}^T + \sigma_\delta^2 \mathbf{I}_M \right]^{-1} \quad (3.8)$$

$$= \sqrt{\frac{\mathcal{P}}{M}} \Sigma_s \mathbf{H}^T \mathbf{E} \mathbf{D}_{\mathbf{x}}^{-\frac{1}{2}} \mathbf{V}^T \left[ \sigma_\delta^{-2} \mathbf{I}_M - \frac{\mathcal{P}}{M} \sigma_\delta^{-2} \mathbf{V} \mathbf{C} \mathbf{V}^T \sigma_\delta^{-2} \right] \quad (3.9)$$

$$= \sqrt{\frac{\mathcal{P}}{M}} \Sigma_s \mathbf{H}^T \mathbf{E} \mathbf{D}_{\mathbf{x}}^{-\frac{1}{2}} \left[ \mathbf{I}_M - \frac{\mathcal{P}}{M} \sigma_\delta^{-2} \mathbf{V}^T \mathbf{V} \mathbf{C} \right] \sigma_\delta^{-2} \mathbf{V}^T \quad (3.10)$$

$$= \sqrt{\frac{\mathcal{P}}{M}} \Sigma_s \mathbf{H}^T \mathbf{E} \mathbf{D}_{\mathbf{x}}^{-\frac{1}{2}} \left[ \mathbf{I}_M - (-\mathbf{I}_M + \mathbf{C}^{-1}) \mathbf{C} \right] \sigma_\delta^{-2} \mathbf{V}^T \quad (3.11)$$

$$= \sqrt{\frac{\mathcal{P}}{M}} \Sigma_s \mathbf{H}^T \mathbf{E} \mathbf{D}_{\mathbf{x}}^{-\frac{1}{2}} \mathbf{C} \sigma_\delta^{-2} \mathbf{V}^T, \quad (3.12)$$

where

$$\mathbf{C} = \left( \mathbf{I}_N + \frac{\mathcal{P}}{M} \sigma_\delta^{-2} \mathbf{V}^T \mathbf{V} \right)^{-1}, \quad (3.13)$$

and we used the Woodbury matrix identity in equation 3.9.

Under a minor assumption that the signal covariance  $\Sigma_s$  and the fixed linear distortion  $\mathbf{H}$  share the same eigenvectors,<sup>1</sup> we rewrite  $\Sigma_s = \mathbf{E}\Lambda\mathbf{E}^T$  and  $\mathbf{H} = \mathbf{E}\Delta\mathbf{E}^T$ , and equation 3.12 is further simplified as

$$\mathbf{A} = \sqrt{\frac{\mathcal{P}}{M}}\sigma_\delta^{-2}\mathbf{E}\Lambda\Delta\mathbf{D}_x^{-\frac{1}{2}}\mathbf{C}\mathbf{V}^T. \quad (3.14)$$

By substituting equations 3.5 and 3.12 in equation 2.7, the MSE can be expressed in terms of  $\mathbf{V}$ :

$$\begin{aligned} \mathcal{E} &= \text{tr}(\Lambda) - 2[\text{tr}(\Lambda^2\Delta^2\mathbf{D}_x^{-1}) - \text{tr}\{\Lambda^2\Delta^2\mathbf{D}_x^{-1}\mathbf{C}\}] \\ &\quad + \text{tr}(\Lambda^2\Delta^2\mathbf{D}_x^{-1}) - 2\text{tr}\{\Lambda^2\Delta^2\mathbf{D}_x^{-1}\mathbf{C}\} + \text{tr}\{\Lambda^2\Delta^2\mathbf{D}_x^{-1}\mathbf{C}^2\} \\ &\quad + \text{tr}\{\Lambda^2\Delta^2\mathbf{D}_x^{-1}\mathbf{C}\} - \text{tr}\{\Lambda^2\Delta^2\mathbf{D}_x^{-1}\mathbf{C}^2\} \end{aligned} \quad (3.15)$$

$$= \text{tr}(\Lambda) - \text{tr}(\Lambda^2\Delta^2\mathbf{D}_x^{-1}) + \text{tr}\{\Lambda^2\Delta^2\mathbf{D}_x^{-1}\mathbf{C}\}. \quad (3.16)$$

Thus we have arrived at the exact decomposition given in equation 3.3. The first term in equation 3.16 is the signal variance, and the second term is the (negative) variance of the best linear reconstruction subject to the input degradation (or the variance of the Wiener estimate) (Gonzalez & Woods, 2002). Therefore, the first two terms collectively correspond to the MMSE subject to the input degradation. The third term corresponds to the MSE function with  $\mathbf{V}$  subject to AWGN in  $M$ -dimensional channel whose  $N$ -dimensional input spectrum  $\tilde{\mathbf{D}}_x = \Lambda^2\Delta^2\mathbf{D}_x^{-1}$  (Lee, 1975; Lee & Petersen, 1976; Doi et al., 2007).<sup>2</sup> The optimal  $\mathbf{V}$  that satisfies the total or the individual power constraint is ready to be derived from these earlier studies.  $\square$

<sup>1</sup>This holds in the application of image coding. Specifically, the image signal  $\mathbf{s}$  is shift invariant, and the linear distortion is optical blur, and hence  $\mathbf{H}$  is convolutional. In such a case, the eigenvectors are given by the Fourier basis functions. More precisely, we further assume the periodic boundary condition for both  $\mathbf{s}$  and  $\mathbf{H}$ , and then  $\Sigma_s$  and  $\mathbf{H}$  are circulant in addition to Toeplitz. The eigenvector matrix of circulant Toeplitz matrix is the DFT matrix. Alternatively, we can employ an approximation for a noncirculant Toeplitz matrix without assuming periodic boundary condition, which is called circulant approximation (Gray, 2006).

<sup>2</sup>The corresponding MMSE problem in the original signal space (before whitening) is to find a linear transform  $\tilde{\mathbf{W}} = \sqrt{\frac{\mathcal{P}}{M}}\mathbf{V}\mathbf{D}_x^{-\frac{1}{2}}\mathbf{E}^T$  where  $\tilde{\Sigma}_x = \mathbf{E}\mathbf{D}_x\mathbf{E}^T$  is the covariance of input signal. The power constraints are, respectively,

$$\text{tr}(\tilde{\mathbf{W}}\tilde{\Sigma}_x\tilde{\mathbf{W}}^T) = \mathcal{P} \Rightarrow \text{tr}(\mathbf{V}\mathbf{V}^T) = M, \quad (3.17)$$

$$\text{diag}(\tilde{\mathbf{W}}\tilde{\Sigma}_x\tilde{\mathbf{W}}^T) = \frac{\mathcal{P}}{M}\mathbf{1}_M \Rightarrow \text{diag}(\mathbf{V}\mathbf{V}^T) = \mathbf{1}_M, \quad (3.18)$$

namely, those robust coding solutions satisfy the same power constraints in the generalized problem.

This decomposition implies that the optimal filtering in the MSE sense is given by denoising and deblurring of the input signal, followed by the optimal coding of this cleaned signal where the signal transmission is restricted by channel noise and the limited channel dimension. The first filtering is separable from the second, implying that input degradation cannot be compensated by increasing the representational capacity of the channel, for example, by decreasing channel noise or increasing the channel dimension.

In summary, the MMSE linear transform can be obtained by the following steps:

1. Find the Wiener filter given the input signal spectrum, the linear distortion, and the amplitude of input noise:  $\mathbf{W}_{wiener}$ .
2. Find the robust coding solution given the Wiener estimate spectrum, the amplitude of channel noise, and the channel dimension:  $\mathbf{W}_{robust}$ .
3. The optimal linear transform for the generalized problem is  $\mathbf{W}_{mmse} = \mathbf{W}_{robust} \mathbf{W}_{wiener}$ .

**3.3 Spectral Characterization.** The exact decomposition into Wiener filtering and robust coding allows us to gain insight into how the reconstruction error is minimized by analyzing how the power spectrum (eigenvalues) of a generic signal is transformed through the two serial processes.

The first process, Wiener filtering, can be expressed by

$$\mathbf{W}_{wiener} = \mathbf{E} \mathbf{\Delta} \mathbf{\Lambda} \mathbf{D}_x^{-1} \mathbf{E}^T, \quad (3.19)$$

where  $\mathbf{D}_x = \mathbf{\Delta}^2 \mathbf{\Lambda} + \sigma_v^2 \mathbf{I}_N$ . Recall the spectrum of the Wiener estimate  $\mathbf{s}^*$  is  $\mathbf{\Delta}^2 \mathbf{\Lambda}^2 \mathbf{D}_x^{-1} \equiv \mathbf{\Lambda}^* = \text{diag}(\lambda_1^*, \dots, \lambda_N^*)$  and how the original signal is restored against linear distortion and input noise is well understood.

The second process, robust coding, is given by

$$\mathbf{W}_{robust} = \mathbf{P} \left[ \sqrt{\frac{\mathcal{P}}{M}} \mathbf{\Omega} \mathbf{\Lambda}^{*-\frac{1}{2}} \right] \mathbf{E}^T, \quad (3.20)$$

where  $\mathbf{P}$  is some  $M$ -dimensional orthogonal matrix<sup>3</sup> and  $\left[ \sqrt{\frac{\mathcal{P}}{M}} \mathbf{\Omega} \mathbf{\Lambda}^{*-\frac{1}{2}} \right]$  is the gain of input signal in the eigenspace (note the input signal of this second process,  $\mathbf{y} = \mathbf{s}^*$ , is represented in the signal's eigenspace by  $\mathbf{E}^T$  in

<sup>3</sup>Under the total power constraint, the robust coding solution is unique up to an orthogonal matrix  $\mathbf{P}$  (see equation 3.20). Note that  $\mathbf{P}$  is cancelled out from equation 2.8. Under the individual power constraint,  $\mathbf{P}$  acts to distribute the total power evenly over the coding units. This problem is known as the inverse eigenvalue problem (Chu & Golub, 2005), and the existence of such an orthogonal matrix and an algorithm to find it were shown in Lee (1975) and Lee and Petersen (1976).

equation 3.20). The  $i$ th element of this gain (squared, for clarity) is given by

$$\frac{\mathcal{P}}{M} \omega_i^2 \frac{1}{\lambda_i^*} = \begin{cases} \left( \frac{1}{\sqrt{l_0}} - \frac{1}{\sqrt{\lambda_i^*}} \right) \frac{\sigma_\delta^2}{\sqrt{\lambda_i^*}}, & \text{if } i = 1 \sim K, \\ 0, & \text{otherwise,} \end{cases} \quad (3.21a)$$

$$(3.21b)$$

where

$$l_0 = \left[ \frac{1}{\frac{\mathcal{P}}{K\sigma_\delta^2} + 1} \frac{1}{K} \sum_{i=1}^K \sqrt{\lambda_i^*} \right]^2 \quad (3.22)$$

is the threshold and only signal components whose eigenvalues are greater than this value are encoded with robust coding, and  $K$  is the total number of such components (without loss of generality, we assume  $\lambda_i^*$  is sorted in descending order; see Lee, 1975; Lee & Petersen, 1976, for derivation).  $K$  needs to be computed numerically.

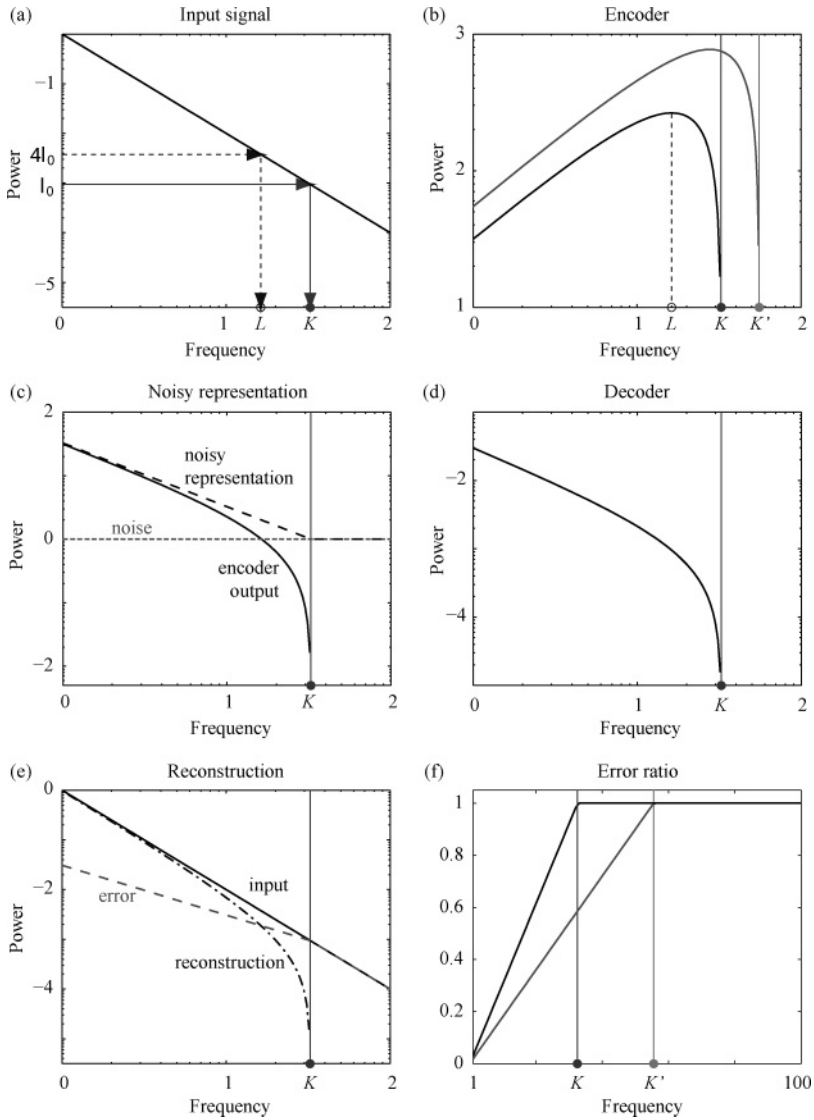
Our analysis is reduced to that in Ruderman (1994) when the linear encoder  $\mathbf{W}$  is restricted to be convolutional, namely, when the eigenvector matrix  $\mathbf{E}$  is the Fourier basis and the  $M$ -dimensional orthogonal matrix  $\mathbf{P}$  is also given by the same Fourier basis  $\mathbf{E}$  (note this is  $N$ -dimensional).

Unlike Wiener filtering, robust coding is much less widely known, and its characteristics have not been fully examined except for one- or two-dimensional signal problems (Doi et al., 2007). Next we illustrate four major characteristics of robust coding for a general  $N$ -dimensional signal.

First, the threshold  $l_0$  defines which input signal dimensions are encoded. Figure 2a shows the signal's eigenvalues (power spectrum). Components that exceed  $l_0$  are encoded, and the rest are discarded in robust coding. The corresponding "critical" index (or frequency),  $K$ , is indicated with the gray vertical line in all panels in Figure 2. From equation 3.22, we observe the following:

- The threshold  $l_0$  gets lower (and hence the critical frequency  $K$  goes higher; see Figure 2a) if a larger total power  $\mathcal{P}$  (see equation 2.8) is available in the neural population. The total power can be increased by increasing the power available for individual neurons or by increasing the neural population size while the individual neural power is fixed, or both. Note that overcompleteness in our model is useful to minimize MSE because it could increase the total power. If the overcompleteness is increased while the total power is fixed, the MMSE will not change. (This could still be beneficial for the other reason, for example, if a large number of low-power-coding units is





more economical than a small number of high-power-coding units.) The changes of encoder and the critical frequency by doubling the population size at a fixed individual neural power are illustrated in Figure 2b, and the resulting change of the error ratio is shown in Figure 2f.

- The threshold gets lower if channel noise  $\sigma_\delta^2$  is smaller. Note that  $\frac{P}{K\sigma_\delta^2}$  in equation 3.22 is the effective SNR where  $K$  is the dimension of subspace represented by the neural population; it is different from the apparent average SNR of the neural population, which is given by  $\frac{P}{M\sigma_\delta^2}$  where  $M$  is the neural population size.
- The thresholding depends on the anisotropy (or distribution) of input spectrum  $\lambda_i^*$ , and there is no thresholding when it is isotropic. This was thoroughly analyzed for the two-dimensional signal (Doi et al., 2007).

Second, the squared gain of the input signal (see equation 3.21 illustrated in Figure 2b) has the maximum at the frequency  $L$  whose eigenvalue is  $\lambda_L^* = 4I_0$  (see Figure 2a). (This can be shown by the first- and the second-order derivatives of equation 3.21a with respect to  $\lambda_i^*$ .) It is interesting that the bandpass characteristic emerges under the MMSE objective, with the  $1/f^2$  signal. In contrast, under the information maximization objective, the optimal filtering is whitening (Atick, 1992; Cover & Thomas, 2006), and hence it is high-pass filtering with  $1/f^2$  signal (its squared gain is proportional to  $f^2$ ).

Third, the power spectrum of the noisy representation  $\mathbf{r}$  (given by the product of the squared gain of robust coding, equation 3.21a, and the power spectrum of signal,  $\lambda_i^*$ , plus channel noise spectrum,  $\sigma_\delta^2$ ) is proportional to the square root of the signal spectrum if the signal is beyond the threshold;

---

Figure 2: Spectral analysis of robust coding. (a) Input signal  $\mathbf{y}$  (see Figure 1 for notation). In this example, its spectrum  $\lambda_i^*$  is given by  $1/f^2$  as in photographic images of natural scenes (i.e.,  $\lambda_i^* = 1/i^2$  by equating the index  $i$  with the frequency  $f$ ). The input and output dimensions are  $N = 100$  and  $M = 100$ , respectively (hence, a complete representation), and the variance of channel noise,  $\sigma_\delta^2$ , is set so that the average SNR of the channel  $\frac{P}{M\sigma_\delta^2} = 1$ . Two points on the horizontal axis indicate the critical frequency,  $K$  (closed circle), and the maximizer of the encoder spectrum,  $L$  (open circle), respectively (see the text). (b) Encoder  $\mathbf{W}_{robust}$ . We show the encoder spectra with  $M = 100$  (black curve) and also with  $M = 200$  (gray curve; this is a  $2\times$  overcomplete representation), with the corresponding critical frequency,  $K$  and  $K'$ , respectively. Maximizer  $L$  is shown only for the complete representation. (c) Noisy representation  $\mathbf{r}$ . The encoder output  $\mathbf{W}_{robust} \mathbf{s}$  and channel noise  $\delta$  are also shown. (d) Decoder  $\mathbf{A}$ . (e) Reconstruction  $\hat{\mathbf{y}}$ . This is given by the multiplication of noisy representation (c) and decoder (d). (f) The error ratio for each frequency, as defined in equation 3.25. With the  $1/f^2$  power spectrum of input signal, the ratio before thresholding (see equation 3.25a) is given by  $\sqrt{I_0} f$  (note that this plot is in the linear axes to clarify the linearity). As in (b) we illustrate two cases: complete (black) and  $2\times$  overcomplete (gray) representations.

otherwise it is identical to the noise spectrum, that is, those components are all noise (see Figure 2c):

$$\frac{\mathcal{P}}{M}\omega_i^2 + \sigma_\delta^2 = \begin{cases} \frac{\sigma_\delta^2}{\sqrt{I_0}}\sqrt{\lambda_i^*} & \text{if } i = 1 \sim K, \\ \sigma_\delta^2 & \text{otherwise.} \end{cases} \quad (3.23a)$$

$$(3.23b)$$

This may be seen as an intermediate between the original signal  $\lambda_i^*$  and a flat spectrum generated by whitening. This “half-whitening” is a distinct feature of robust coding.

Finally, reconstruction is more precise for components whose power is larger. The multiplication of the noisy representation (see Figure 2c) and the decoder (2d) yields the reconstructed signal (2e). The reconstruction error (also shown in Figure 2e) is

$$\begin{cases} \sqrt{I_0\lambda_i^*} & \text{if } i = 1 \sim K, \\ \lambda_i^* & \text{otherwise,} \end{cases} \quad (3.24a)$$

$$(3.24b)$$

implying that the ratio of the error at each component  $i$  is (see Figure 2f):

$$\frac{(\text{Error variance})}{(\text{Data variance})} = \begin{cases} \sqrt{\frac{I_0}{\lambda_i^*}} & \text{if } i = 1 \sim K, \\ 1 & \text{otherwise.} \end{cases} \quad (3.25a)$$

$$(3.25b)$$

This results from the optimal allocation of a limited representational resource: more resources are allocated to signal components whose power is higher. In contrast, the information maximization objective leads to whitening instead of robust coding, in which the reconstruction spectrum is proportional to the input spectrum and the error ratio is constant for any signal component, even if the power spectrum of the signal is not uniform (Doi & Lewicki, 2005). This may be interpreted as allocating the representational resources evenly over the signal components even if their signal strengths are different. The suboptimality of whitening in the MMSE sense has been observed in the literature (Bethge, 2006; Doi et al., 2007; Eichhorn, Sinz, & Bethge, 2009), and our analysis provides an explanation as to why that is the case.

#### 4 Conclusion

---

We investigated how to optimally organize a population of noisy neurons in order to represent noisy input signals. Our analysis provides an exact solution and its characterization under an idealized setting, which is applicable in a wide range of conditions. For example, the degradation of input

signals can model optical blur and additive noise in image formation and could be set from measurements in a specific system. Similarly, the neural (channel) noise and population size can also be determined for the system of interest. This provides a way to predict the optimal code for a wide variety of biological systems. The application need not be restricted to vision or even to biological systems, and our results could be used to design optimal signal processors.

### Acknowledgments

---

We thank a reviewer for pointing out the prior study on the decomposition in a general case (Wolf & Ziv, 1970).

### References

---

- Atick, J. J. (1992). Could information theory provide an ecological theory of sensory processing? *Network*, 3, 213–251.
- Atick, J. J., Li, Z., & Redlich, A. N. (1990). *Color coding and its interaction with spatiotemporal processing in the retina* (Tech. Rep. IASSNS-HEP-90/75). Princeton, NJ: Institute for Advanced Study.
- Atick, J. J., & Redlich, A. N. (1990). Towards a theory of early visual processing. *Neural Computation*, 2, 308–320.
- Atick, J. J., & Redlich, A. N. (1992). What does the retina know about natural scenes? *Neural Computation*, 4, 196–210.
- Bethge, M. (2006). Factorial coding of natural images: How effective are linear models in removing higher-order dependencies? *J. Opt. Soc. Am. A*, 23(6), 1253–1268.
- Chu, M. T., & Golub, G. H. (2005). *Inverse eigenvalue problems*. New York: Oxford University Press.
- Cover, T. M., & Thomas, J. A. (2006). *Elements of information theory* (2nd ed.). Hoboken, NJ: Wiley.
- Dayan, P., & Abbott, L. F. (2001). *Theoretical neuroscience: Computational and mathematical modeling of neural systems*. Cambridge, MA: MIT Press.
- Dobrushin, R. L., & Tsybakov, B. S. (1962). Information transmission with additional noise. *IRE Transactions on Information Theory*, 8, 293–304.
- Doi, E., Balcan, D. C., & Lewicki, M. S. (2007). Robust coding over noisy overcomplete channels. *IEEE Transactions on Image Processing*, 16, 442–452.
- Doi, E., & Lewicki, M. S. (2005). Sparse coding of natural images using an overcomplete set of limited capacity units. In L. K. Saul, Y. Weiss, & L. Bottou (Eds.), *Advances in neural information processing systems*, 17 (pp. 377–384). Cambridge, MA: MIT Press.
- Eichhorn, J., Sinz, F., & Bethge, M. (2009). Natural image coding in V1: How much use is orientation selectivity? *PLoS Computational Biology*, 5, 1–16.
- Gonzalez, R. C., & Woods, R. E. (2002). *Digital image processing* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

- Gray, R. M. (2006). Toeplitz and circulant matrices: A review. *Foundations and Trends in Communications and Information Theory*, 2, 155–239.
- Gray, R. M., & Neuhoff, D. L. (1998). Quantization. *IEEE Transactions on Information Theory*, 44, 2325–2383.
- Lee, K.-H. (1975). *Optimal linear coding for a multichannel system*. Unpublished doctoral dissertation, University of New Mexico, Albuquerque.
- Lee, K.-H., & Petersen, D. P. (1976). Optimal linear coding for vector channels. *IEEE Transactions on Communications*, COM-24, 1283–1290.
- Palomar, D. P., & Jiang, Y. (2007). MIMO transceiver design via majorization theory. *Foundations and Trends in Communications and Information Theory*, 3, 331–551.
- Ruderman, D. L. (1994). Designing receptive fields for highest fidelity. *Network: Comput. Neural Syst.*, 5, 147–155.
- Sakrison, D. J. (1968). Source encoding in the presence of random disturbance. *IEEE Transactions on Information Theory*, IT-14, 165–167.
- van Hateren, J. H. (1992). A theory of maximizing sensory information. *Biological Cybernetics*, 68, 23–29.
- Wolf, J. K., & Ziv, J. (1970). Transmission of noisy information to a noisy receiver with minimum distortion. *IEEE Transactions on Information Theory*, IT-16, 406–411.

---

Received July 30, 2010; accepted March 13, 2011.