Supplementary Material

Abstract

This document is the supplementary material for the article "Knowledge and Data Dual-Driven Channel Estimation and Feedback for Ultra-Massive MIMO Systems". There is one sole part including the derivation of the minimum mean square error (MMSE)-based denoiser.

I. Derivation of the Minimum Mean Square Error Denoiser

Consider the denoising problem for the signal model given by

$$\tilde{\mathbf{x}} = \bar{\mathbf{x}} + \boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{n}, \tag{1}$$

where $\tilde{\mathbf{x}}$ and $\bar{\mathbf{x}} \in \mathbb{C}^{K\times 1}$ denote the noisy and noiseless signals, respectively, $\mathbf{n} \in \mathbb{C}^{K\times 1} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ is an AWGN vector, and $\boldsymbol{\Sigma} = \mathrm{diag}\left([\sigma^2[1], \sigma^2[2], \ldots, \sigma^2[K]]\right)$ is the diagonal matrix depicting the noise power with $\sigma^2[k] \geq 0, \forall k \in \{1, 2, \ldots, K\}$. Because the support across different subcarriers appears or disappears at the same time, $\bar{\mathbf{x}}$ follows a Bernoulli-Gaussian distribution as $(1-\gamma)\delta_0 + \gamma p_{\mathbf{h}_\epsilon}$. Here, δ_0 denotes the point mass measure at zero and $p_{\mathbf{h}_\epsilon}$ denotes the distribution of $\mathbf{h}_\epsilon \sim \mathcal{CN}(\mathbf{0}, \epsilon\mathbf{I})$.

In this way, the probability when $\tilde{\mathbf{x}} = \mathbf{x}' = \mathbf{h}_\epsilon + \boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{n}$ is γ, and the probability is $1-\gamma$ when $\tilde{\mathbf{x}} = \boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{n}$. According to standard estimation theory, defining $\boldsymbol{\Theta} = \mathrm{diag}\left(\frac{1}{\epsilon+\sigma^2[1]}, \ldots, \frac{1}{\epsilon+\sigma^2[K]}\right)$ the mean and covariance matrix of $\mathbf{h}_\epsilon$ can be computed respectively as

$$\mathbb{E}[\mathbf{h}_\epsilon | \mathbf{x}' = \mathbf{x}] = \epsilon\boldsymbol{\Theta}\mathbf{x}, \tag{2}$$

$$\mathbb{E}[\mathbf{h}_\epsilon \mathbf{h}_\epsilon^{\mathrm{H}} | \mathbf{x}' = \mathbf{x}] = \epsilon\mathbf{I} - \epsilon^2\boldsymbol{\Theta} + \epsilon^2\boldsymbol{\Theta}\mathbf{x}\mathbf{x}^{\mathrm{H}}\boldsymbol{\Theta}. \tag{3}$$

Furthermore, we can compute the mean of $\bar{\mathbf{x}}$ as

$$\begin{aligned}
\mathbb{E}[\bar{\mathbf{x}}|\hat{\mathbf{x}} = \hat{\mathbf{x}}'] &= \int \bar{\mathbf{x}} p_{\mathbf{x}|\hat{\mathbf{x}}}(\bar{\mathbf{x}} = \mathbf{x}|\hat{\mathbf{x}} = \hat{\mathbf{x}}')\mathrm{d}\mathbf{x} \\
&= \frac{1}{p_{\hat{\mathbf{x}}}} \int p_{\mathbf{x}|\hat{\mathbf{x}}}(\bar{\mathbf{x}} = \mathbf{x}|\hat{\mathbf{x}} = \hat{\mathbf{x}}')(\gamma p_{\mathbf{h}_\epsilon}(\mathbf{h}_\epsilon = \mathbf{x}) + (1-\gamma)\delta_0(\mathbf{x}))\mathrm{d}\mathbf{x} \\
&= \frac{\gamma p_{\mathbf{x}'}(\mathbf{x}' = \hat{\mathbf{x}}')}{p_{\hat{\mathbf{x}}}(\hat{\mathbf{x}} = \hat{\mathbf{x}}') p_{\mathbf{x}'}(\mathbf{x}' = \hat{\mathbf{x}}')} \mathbb{E}[\mathbf{h}_\epsilon|\mathbf{x}' = \hat{\mathbf{x}}'],
\end{aligned} \tag{4}$$

By defining $\phi(\hat{\mathbf{x}}) = \frac{1}{1+\frac{1-\gamma}{\gamma} e^{-\hat{\mathbf{x}}^{\mathrm{H}}\mathbf{P}\hat{\mathbf{x}}} \prod_{k=1}^{K}(1+\frac{\epsilon}{\sigma^2[k]})}$ and $\mathbf{P} = \mathrm{diag}\left(\frac{\epsilon}{\sigma^2[1](\sigma^2[1]+\epsilon)}, \ldots, \frac{\epsilon}{\sigma^2[K](\sigma^2[K]+\epsilon)}\right)$, the shrinkage function $\boldsymbol{\eta}_{\mathrm{CS}}(\hat{\mathbf{x}}'; \gamma, \epsilon, \boldsymbol{\Sigma})$ can be rewritten as

$$\boldsymbol{\eta}_{\mathrm{CS}}(\hat{\mathbf{x}}'; \gamma, \epsilon, \boldsymbol{\Sigma}) = \mathbb{E}[\mathbf{x}|\hat{\mathbf{x}} = \hat{\mathbf{x}}'] = \phi(\hat{\mathbf{x}}')\boldsymbol{\Theta}\hat{\mathbf{x}}'. \tag{5}$$

It should be noted that when taking the derivative of (5), we can approximate $\phi(\hat{\mathbf{x}})$ as a constant since the dimension of $\hat{\mathbf{x}}$ is quite large, and the derivative becomes $\frac{\epsilon\phi(\hat{\mathbf{x}})}{\epsilon+\sigma^2[k]}$.