1 MAP Estimation

We know that $\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x} + \boldsymbol{\eta}$ where $\boldsymbol{x} \in \mathbb{R}^n, \boldsymbol{y}, \boldsymbol{\eta} \in \mathbb{R}^m, \boldsymbol{\Phi} \in \mathbb{R}^{m\times n}$. Also, $\boldsymbol{x} \sim \mathcal{N}(0, \boldsymbol{\Sigma}_{\boldsymbol{x}}), \boldsymbol{\eta} \sim \mathcal{N}(0, \sigma^2 \boldsymbol{I}_{m\times m})$

Now, the MAP estimate for $\boldsymbol{x}$ is given by

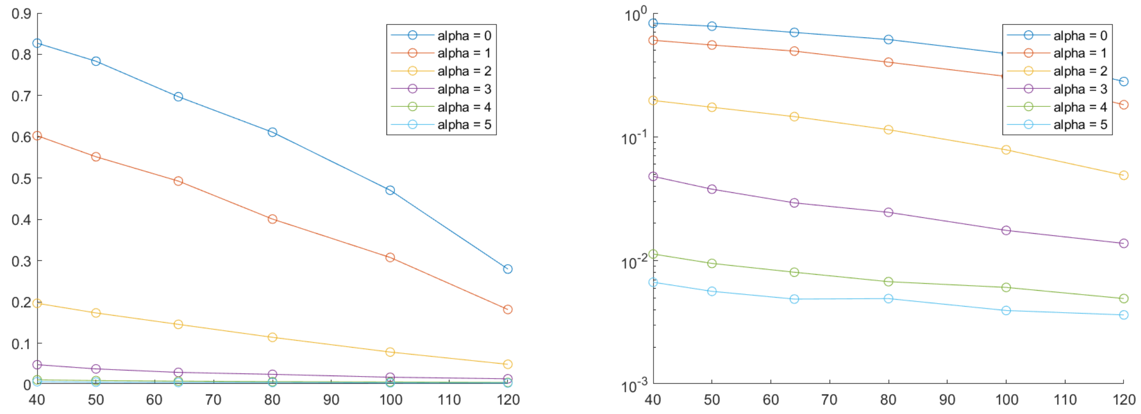
$$
\begin{aligned}
\hat{\boldsymbol{x}} &= \arg\max_{\boldsymbol{x}} \mathsf{Prob}(\boldsymbol{x}|\boldsymbol{y}) \\
&= \arg\max_{\boldsymbol{x}} \frac{\mathsf{Prob}(\boldsymbol{y}|\boldsymbol{x})\mathsf{Prob}(\boldsymbol{x})}{\mathsf{Prob}(\boldsymbol{y})} && \text{(Bayes' rule)} \\
&= \arg\max_{\boldsymbol{x}} \mathsf{Prob}(\boldsymbol{y}|\boldsymbol{x})\mathsf{Prob}(\boldsymbol{x}) && \text{(as } \mathsf{Prob}(\boldsymbol{y}) \text{ is a constant)} \\
&= \arg\max_{\boldsymbol{x}} \underbrace{\frac{1}{(2\pi\sigma^2)^{m/2}} \exp -\frac{\|\boldsymbol{y}-\boldsymbol{\Phi}\boldsymbol{x}\|_2^2}{2\sigma^2}}_{\mathsf{Prob}(\boldsymbol{y}|\boldsymbol{x})} \underbrace{\frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}_{\boldsymbol{x}}|^{1/2}} \exp -\frac{\boldsymbol{x}^T\boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1}\boldsymbol{x}}{2}}_{\mathsf{Prob}(\boldsymbol{x})} \\
&= \arg\max_{\boldsymbol{x}} \exp -\frac{\|\boldsymbol{y}-\boldsymbol{\Phi}\boldsymbol{x}\|_2^2}{2\sigma^2} \exp -\frac{\boldsymbol{x}^T\boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1}\boldsymbol{x}}{2} && \text{(ignoring scaling constants)} \\
&= \arg\min_{\boldsymbol{x}} \frac{\|\boldsymbol{y}-\boldsymbol{\Phi}\boldsymbol{x}\|_2^2}{2\sigma^2} + \frac{\boldsymbol{x}^T\boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1}\boldsymbol{x}}{2} && \text{(taking negative logarithm)} \\
&= \arg\min_{\boldsymbol{x}} \frac{(\boldsymbol{y}-\boldsymbol{\Phi}\boldsymbol{x})^T(\boldsymbol{y}-\boldsymbol{\Phi}\boldsymbol{x})}{2\sigma^2} + \frac{\boldsymbol{x}^T\boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1}\boldsymbol{x}}{2} \\
&= \arg\min_{\boldsymbol{x}} \frac{\boldsymbol{y}^T\boldsymbol{y} - (\boldsymbol{\Phi}\boldsymbol{x})^T\boldsymbol{y} - \boldsymbol{y}^T\boldsymbol{\Phi}\boldsymbol{x} + (\boldsymbol{\Phi}\boldsymbol{x})^T\boldsymbol{\Phi}\boldsymbol{x}}{2\sigma^2} + \frac{\boldsymbol{x}^T\boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1}\boldsymbol{x}}{2} \\
&= \arg\min_{\boldsymbol{x}} \underbrace{\frac{-\boldsymbol{x}^T\boldsymbol{\Phi}^T\boldsymbol{y} - \boldsymbol{y}^T\boldsymbol{\Phi}\boldsymbol{x} + \boldsymbol{x}^T\boldsymbol{\Phi}^T\boldsymbol{\Phi}\boldsymbol{x}}{2\sigma^2} + \frac{\boldsymbol{x}^T\boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1}\boldsymbol{x}}{2}}_{=\boldsymbol{F}\text{ (let)}} && \text{(ignoring additive constants)}
\end{aligned}
$$

Now, to optimise $\boldsymbol{F}$, set $\dfrac{\mathrm{d}\boldsymbol{F}}{\mathrm{d}\boldsymbol{x}} = 0$

$$
\begin{aligned}
\frac{\mathrm{d}\boldsymbol{F}}{\mathrm{d}\boldsymbol{x}} &= \frac{\mathrm{d}}{\mathrm{d}\boldsymbol{x}}\left(\frac{-\boldsymbol{x}^T\boldsymbol{\Phi}^T\boldsymbol{y} - \boldsymbol{y}^T\boldsymbol{\Phi}\boldsymbol{x} + \boldsymbol{x}^T\boldsymbol{\Phi}^T\boldsymbol{\Phi}\boldsymbol{x}}{2\sigma^2} + \frac{\boldsymbol{x}^T\boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1}\boldsymbol{x}}{2}\right) \\
&= \frac{\mathrm{d}}{\mathrm{d}\boldsymbol{x}}\left(\frac{-\boldsymbol{x}^T\boldsymbol{\Phi}^T\boldsymbol{y}}{2\sigma^2}\right) + \frac{\mathrm{d}}{\mathrm{d}\boldsymbol{x}}\left(\frac{-\boldsymbol{y}^T\boldsymbol{\Phi}\boldsymbol{x}}{2\sigma^2}\right) + \frac{\mathrm{d}}{\mathrm{d}\boldsymbol{x}}\left(\frac{\boldsymbol{x}^T\boldsymbol{\Phi}^T\boldsymbol{\Phi}\boldsymbol{x}}{2\sigma^2}\right) + \frac{\mathrm{d}}{\mathrm{d}\boldsymbol{x}}\left(\frac{\boldsymbol{x}^T\boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1}\boldsymbol{x}}{2}\right) \\
&= \left(\frac{-\boldsymbol{\Phi}^T\boldsymbol{y}}{2\sigma^2}\right) + \left(\frac{-(\boldsymbol{y}^T\boldsymbol{\Phi})^T}{2\sigma^2}\right) + \left(\frac{\left(\boldsymbol{\Phi}^T\boldsymbol{\Phi} + (\boldsymbol{\Phi}^T\boldsymbol{\Phi})^T\right)\boldsymbol{x}}{2\sigma^2}\right) + \left(\frac{\left(\boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1} + (\boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1})^T\right)\boldsymbol{x}}{2}\right) \\
&= \left(\frac{-2\boldsymbol{\Phi}^T\boldsymbol{y}}{2\sigma^2}\right) + \left(\frac{2\boldsymbol{\Phi}^T\boldsymbol{\Phi}\boldsymbol{x}}{2\sigma^2}\right) + \left(\frac{2\boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1}\boldsymbol{x}}{2}\right) \\
&\Rightarrow \left(\frac{-\boldsymbol{\Phi}^T\boldsymbol{y} + \boldsymbol{\Phi}^T\boldsymbol{\Phi}\boldsymbol{x}}{\sigma^2}\right) + \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1}\boldsymbol{x} = 0 \\
&\Rightarrow -\boldsymbol{\Phi}^T\boldsymbol{y} + \boldsymbol{\Phi}^T\boldsymbol{\Phi}\boldsymbol{x} + \sigma^2\boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1}\boldsymbol{x} = 0 \\
&\Rightarrow \left(\boldsymbol{\Phi}^T\boldsymbol{\Phi} + \sigma^2\boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1}\right)\boldsymbol{x} = \boldsymbol{\Phi}^T\boldsymbol{y} \\
&\Rightarrow \boldsymbol{x} = \left(\boldsymbol{\Phi}^T\boldsymbol{\Phi} + \sigma^2\boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1}\right)^{-1}\boldsymbol{\Phi}^T\boldsymbol{y}
\end{aligned}
$$

Hence, the MAP estimate of $\boldsymbol{x}$ is $\hat{\boldsymbol{x}} = \left(\boldsymbol{\Phi}^T\boldsymbol{\Phi} + \sigma^2\boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1}\right)^{-1}\boldsymbol{\Phi}^T\boldsymbol{y}$.

2 Observations



(a)

(b) $y-$axis on log scale

Figure 1: RMSE vs m with varying α

- RMSE is lower for higher α (more apparant on the log-scale)
- RMSE is lower (or very close) as m increases for a fixed α
- RMSE with $\alpha = 3$ is < 0.05 for all m whereas RMSE ranges from 0.8271 to 0.2797 with $\alpha = 0$

The reconstruction is better (lower RMSE) as α increases as the case of the convariance matrix with decaying eigenvalues is equivalent to signal sparsity in some orthonormal basis *(slide 84 of Statistics of Natural Images)*