Note: This is a supplementary file for paper "Invisible geo-location signature in a single image," published in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing* by Chau-Wai Wong, Adi Hajj-Ahmad, and Min Wu.

APPENDIX

Using single column to estimate parameters

Let us define a column of smooth native image content as $\mathbf{x}^{(k)} \in \mathbb{R}^{N\times 1}$, and an embedding ENF signal as $\mathbf{e}$, where

$$e(i) = A_{\text{g}} \cos\left(2\pi \frac{f_{\text{g}}}{f_{\text{row}}} \cdot i + \phi_{\text{g}}\right). \tag{2}$$

We define $\mathbf{p}_{\text{g}} = [A_{\text{g}}, f_{\text{g}}, \phi_{\text{g}}]^T$ to be a vector of the ground truth parameters. The corrupted sensing signal is therefore $\mathbf{y} = \mathbf{x}^{(k)} + \mathbf{e}$. We set up a cost function to estimate $\mathbf{p}$ as follows:

$$J_k(\mathbf{p}) = \text{entropy}\left(\text{hist}\left[\left\{x^{(k)}(i) + r(i;\mathbf{p}_{\text{g}},\mathbf{p})\right\}_{i=1}^{N}\right]\right) \tag{3a}$$

$$= \text{entropy}\left(\text{hist}\left[\mathbf{x}^{(k)} + \mathbf{r}(\mathbf{p}_{\text{g}},\mathbf{p})\right]\right), \tag{3b}$$

where $r(i;\mathbf{p}_{\text{g}},\mathbf{p}) = e(i) - A\cos\left(2\pi\frac{f}{f_{\text{row}}}\cdot i + \phi\right)$. We denote $\hat{\mathbf{p}}(\mathbf{x}^{(k)})$ as a numerical suboptimal solution to the optimization problem

$$\min_{\mathbf{p}} J_k(\mathbf{p}). \tag{4}$$

By taking into consideration the randomness of the image, the suboptimality of the numerical algorithm for finding the optimal solution, we model $\hat{\mathbf{p}}(\mathbf{x}^{(k)})$ as a random vector with a small bias $\mathbf{b}$ and a positive definite variance-covariance matrix $\mathbf{\Sigma}$, namely,

$$\mathbb{E}\left[\hat{\mathbf{p}}(\mathbf{x}^{(k)})\right] = \mathbf{p}_{\text{g}} + \mathbf{b}, \quad \text{VarCov}\left(\hat{\mathbf{p}}(\mathbf{x}^{(k)})\right) = \mathbf{\Sigma}. \tag{5}$$

Using multiple columns to improve the accuracy of the estimation

We use a random sample $(\mathbf{x}^{(1)},\cdots,\mathbf{x}^{(K)})$ to improve the accuracy of the estimation. We use $\hat{\mathbf{p}}\left(\{\mathbf{x}^{(k)}\}_{k=1}^{K}\right)$ to denote a numerical suboptimal solution to the multi-column optimization problem

$$\min_{\mathbf{p}} \sum_{k=1}^{K} J_k(\mathbf{p}). \tag{6}$$

We conduct Taylor-series expansion around $\hat{\mathbf{p}}(\mathbf{x}^{(k)})$ for $J_k(p)$, and obtain

$$J_k(\mathbf{p}) = J_k\left(\hat{\mathbf{p}}(\mathbf{x}^{(k)})\right) + \left[\mathbf{p} - \hat{\mathbf{p}}(\mathbf{x}^{(k)})\right]^T \nabla J_k\left(\hat{\mathbf{p}}(\mathbf{x}^{(k)})\right) + \frac{1}{2}\left[\mathbf{p} - \hat{\mathbf{p}}(\mathbf{x}^{(k)})\right]^T \nabla^2 J_k \left[\mathbf{p} - \hat{\mathbf{p}}(\mathbf{x}^{(k)})\right], \tag{7}$$

where $\nabla^2 J_k$ is a Hessian matrix whose individual components are 2nd-order partial derivatives evaluated near $\hat{\mathbf{p}}(\mathbf{x}^{(k)})$. We further assume $J_k(\cdot)$ is continuous hence $\nabla^2 J_k$ is a symmetric matrix.

Substituting the Taylor expanded cost function for each column in (7) into the multi-column cost function in (6), and by setting the gradient with respect to $\mathbf{p}$ to zero, we obtain the optimal analytical solution to (6) as follows:

$$\hat{\mathbf{p}}\left(\{\mathbf{x}^{(k)}\}_{k=1}^{K}\right) = \left(\sum_{k=1}^{K}\nabla^2 J_k\right)^{-1}\left\{\sum_{k=1}^{K}\left[\nabla^2 J_k \hat{\mathbf{p}}(\mathbf{x}^{(k)}) - \nabla J_k\left(\hat{\mathbf{p}}(\mathbf{x}^{(k)})\right)\right]\right\} \tag{8a}$$

$$= \sum_{k=1}^{K}\mathbf{W}_k\,\hat{\mathbf{p}}(\mathbf{x}^{(k)}) - \left(\sum_{k=1}^{K}\nabla^2 J_k\right)^{-1}\left(\sum_{k=1}^{K}\nabla J_k\right) \tag{8b}$$

$$\approx \sum_{k=1}^{K}\mathbf{W}_k\,\hat{\mathbf{p}}(\mathbf{x}^{(k)}), \tag{8c}$$

where $\mathbf{W}_k = \left(\sum_{k=1}^{K}\nabla^2 J_k\right)^{-1}\nabla^2 J_k$. The sample average of gradients in (8b) converges to $\mathbf{0}$ in distribution as K increases. Hence,

$$\mathbb{E}\left[\hat{\mathbf{p}}\left(\{\mathbf{x}^{(k)}\}_{k=1}^{K}\right)\right] \approx \mathbf{p}_{\text{g}} + \mathbf{b}, \tag{9}$$

$$\text{VarCov}\left(\hat{\mathbf{p}}\left(\{\mathbf{x}^{(k)}\}_{k=1}^{K}\right)\right) \approx \sum_{k=1}^{K}\mathbf{W}_k\,\mathbf{\Sigma}\,\mathbf{W}_k^T. \tag{10}$$

The above results reveal that by using K columns, the estimator has a significantly reduced variance and a similar bias as the single column case.

In the 1-d case, the variance formula is degenerated to $\text{Var}(\hat{p}) \approx \left(\sum_{k=1}^{K} w_k^2\right)\sigma^2$. When the weight vector is uniform, i.e., $w_i = \frac{1}{K}$, the best variance reduction is achived, i.e., $\text{Var}(\hat{p}) \approx \sigma^2/K$.