Minimum Norm Estimation

Behrad Soleimani
Department of Electrical and Computer Engineering, University of Maryland, College Park
Email: Behrad@umd.edu

I. Introduction

In many applications, finding the inverse solution from the forward model might not be straight forward due to ill-posedness and rank sufficiency nature of the problem. In this case, a proper approximation of inverse solution would be the minimum norm estimation (MNE) where among the least square solutions we pick the one with the smallest norm. In this report, the closed-form solution of the MNE is obtained.

II. Preliminaries and Problem Formulation

We consider a linear system with observation vector $\mathbf{y}_t \in \mathbb{R}^{N_y}$ at time t. The linear mapping between source vectors and observations considered as follows

$$\mathbf{y}_t = \mathbf{C}\mathbf{x}_t + \mathbf{n}_t, \quad t = 1, 2, \ldots, T, \tag{1}$$

where $\mathbf{C} \in \mathbb{R}^{N_y \times N_x}$ is the linear mapping, $\mathbf{x}_t \in \mathbb{R}^{N_x}$ is the source vector at time t, and $\mathbf{n}_t \in \mathbb{R}^{N_x}$ represents the observation noise at time t. The observations noise is considered as a zero-mean white Gaussian vector with identity covariance matrix, i.e., $\mathbf{n}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{N_y})$ and is independent and identically distributed across the time (in case that the noise covariance matrix is not identity, one can whiten the data first).
The goal is to find MNE such that

$$\widehat{\mathbf{X}}_{\text{MNE}} = \underset{\mathbf{X}}{\text{argmin}} \ \|\mathbf{X}\|_F^2 \quad \text{s.t} \quad \|\mathbf{Y} - \mathbf{C}\mathbf{X}\|_F^2 \leq \eta \tag{2}$$

where $\|.\|_F$ represents the *Frobenius* norm, $\mathbf{X} = [\mathbf{x}_t, \ \forall t] \in \mathbb{R}^{N_x \times T}$ and $\mathbf{Y} = [\mathbf{y}_t, \ \forall t] \in \mathbb{R}^{N_y \times T}$ are the matrix representation of the source and observation vectors, respectively, and $\eta = \gamma \times N_y \times T$ such that γ is a threshold and can be chosen as 95% Quantile.

III. Minimum Norm Estimation

One can write down the Lagrangian of the (2) as

$$\mathcal{L}(\mathbf{X}, \lambda) = \|\mathbf{X}\|_F^2 + \lambda \left(\|\mathbf{Y} - \mathbf{C}\mathbf{X}\|_F^2 - \eta \right), \tag{3}$$

for some $\lambda \geq 0$. Solving the equation $\frac{\partial}{\partial \mathbf{X}} \mathcal{L}(\mathbf{X}, \lambda) = 0$ results in the following solution

$$\widehat{\mathbf{X}}_{\text{MNE}} = \left(\mathbf{C}^\top \mathbf{C} + \frac{1}{\lambda} \mathbf{I} \right)^{-1} \mathbf{C}^\top \mathbf{Y}. \tag{4}$$

In order to find the coefficient λ, we start from the constraint in (2). By substituting $\widehat{\mathbf{X}}_{\text{MNE}}$ in the constraint, λ should satisfy the following inequality

$$\left\| \left(\mathbf{I} - \mathbf{C} \left(\mathbf{C}^\top \mathbf{C} + \frac{1}{\lambda} \mathbf{I} \right)^{-1} \mathbf{C}^\top \right) \mathbf{Y} \right\|_F^2 \leq \eta. \tag{5}$$

Based on the Woodbury matrix identity [2] which states that

$$(\mathbf{A} + \mathbf{U}\mathbf{G}\mathbf{V})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U} \left(\mathbf{G}^{-1} + \mathbf{V}\mathbf{A}^{-1}\mathbf{U} \right)^{-1} \mathbf{V}\mathbf{A}^{-1}, \tag{6}$$

one can rewrite the left hand side of the inequality in (5) as

$$\left\| \left(\mathbf{I} - \mathbf{C} \left(\mathbf{C}^\top \mathbf{C} + \frac{1}{\lambda} \mathbf{I} \right)^{-1} \mathbf{C}^\top \right) \mathbf{Y} \right\|_F^2 = \left\| \left(\mathbf{I} + \lambda \mathbf{C}\mathbf{C}^\top \right)^{-1} \mathbf{Y} \right\|_F^2 = \left\| \mathbf{Q} \left(\mathbf{I} + \lambda \mathbf{D} \right)^{-1} \mathbf{Q}^\top \mathbf{Y} \right\|_F^2. \tag{7}$$

The MATLAB implementation is uploaded on [1].

where in the second equality we used the eigenvalue decomposition $\mathbf{C}\mathbf{C}^\top = \mathbf{Q}\mathbf{D}\mathbf{Q}^\top$ such that $\mathbf{D}$ and $\mathbf{Q}$ are the diagonal eigenvalues and unitary matrices, respectively. Defining $\mathbf{Z} = \mathbf{Q}^\top\mathbf{Y}$ and utilizing norm preservation property of the unitary matrix, we have

$$\left\|\mathbf{Q}\left(\mathbf{I}+\lambda\mathbf{D}\right)^{-1}\mathbf{Q}^\top\mathbf{Y}\right\|_F^2 = \overbrace{\sum_{i=1}^{N_y}\frac{\|\mathbf{Z}_i\|_2^2}{\left(1+\lambda\mathbf{D}_{ii}\right)^2}}^{:=f(\lambda)}, \tag{8}$$

where $\mathbf{Z}_i$ is the i-th row of the matrix $\mathbf{Z} = \mathbf{Q}^\top\mathbf{Y}$ and $\mathbf{D}_{ii}$ represents the i-th diagonal elements of the matrix $\mathbf{D}$. Thus, the inequality in (5) can be rewritten as $f(\lambda) \leq \eta$. It can be easily seen that the function $f(\lambda) \geq 0$ is monotonically decreasing with respect to λ. If $f(\lambda)|_{\lambda=0} \leq \eta$, we conclude that $\lambda = 0$. Otherwise, in order to find the smallest λ satisfying the inequality $f(\lambda) \leq \eta$, one can utilize the Newton's method [3] to solve the equation $f(\lambda) = \eta$. Starting with $\lambda^{(o)} = 0$, we can update λ as

$$\lambda^{(l+1)} = \lambda^{(l)} - \frac{f(\lambda^{(l)}) - \eta}{f'(\lambda^{(l)})},\ l = 1, 2, \ldots, \quad (\lambda^{(0)} = 0) \tag{9}$$

where $f'(\lambda) = -2\sum\limits_{i=1}^{N_y}\frac{\mathbf{D}_{ii}\|\mathbf{Z}_i\|_2^2}{(1+\lambda\mathbf{D}_{ii})^3}$ is the first derivative of $f(\lambda)$.

References

[1] B. Soleimani, “Mne,” https://github.com/BehradSol/MNE, 2020.

[2] K. Petersen and M. Pedersen, “The matrix cookbook, version 20121115,” *Technical Univ. Denmark, Kongens Lyngby, Denmark, Tech. Rep*, vol. 3274, 2012.

[3] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.