CHOOSING THE DIAGONAL LOADING FACTOR FOR LINEAR SIGNAL ESTIMATION USING CROSS VALIDATION

Jun Tong, Qinghua Guo, Jiangtao Xi, Yanguang Yu

SECTE
University of Wollongong
Wollongong, NSW 2522, Australia
{jtong, qguo, jiangtao, yanguang}@uow.edu.au

Peter J. Schreier

SST Group
University of Paderborn
Paderborn, Germany
peter.schreier@sst.upb.de

ABSTRACT

Linear signal estimation based on sample covariance matrices (SCMs) can perform poorly if the training data are limited and the SCMs are ill-conditioned. Diagonal loading (DL) may be used to improve robustness in the face of limited training data. This paper introduces two leave-one-out cross-validation schemes for choosing the DL factor. One scheme repeatedly splits the training data with respect to time, while the other repeatedly splits the out-of-training data with respect to space. We derive computationally efficient implementations and compare them with the oracle choice in terms of the mean squared error.

Index Terms— Cross validation, diagonal loading, sample covariance matrix

1. INTRODUCTION

Consider M-input N-output systems with zero-mean observation $\mathbf{y} \in \mathbb{C}^{N\times 1}$ and zero-mean input $\mathbf{x} \in \mathbb{C}^{M\times 1}$. A typical example is the multi-antenna wireless communication system [1]. The widely applied linear minimum mean squared error (LMMSE) signal estimator [2] can be constructed using estimated covariance matrices as

$$\widehat{\mathbf{x}} = \widehat{\mathbf{C}}_{\mathbf{yx}}^{\dagger}\widehat{\mathbf{C}}_{\mathbf{yy}}^{-1}\mathbf{y}, \tag{1}$$

where $\mathbf{C}_{\mathbf{ab}} \triangleq \mathrm{E}[\mathbf{a}\mathbf{b}^{\dagger}]$ denotes the cross-covariance matrix between $\mathbf{a}$ and $\mathbf{b}$, $\mathrm{E}[\cdot]$ expectation, $(\cdot)^{\dagger}$ conjugate transpose, and $\widehat{a}$ an estimate of a. This estimator minimizes the mean squared error (MSE)

$$\mathrm{MSE}_{\mathbf{x}} \triangleq \mathrm{E}_{\mathbf{x}}[||\mathbf{x} - \widehat{\mathbf{x}}||]^2, \tag{2}$$

where $||\cdot||$ denotes the Frobenius norm, and $\mathrm{E}_{\mathbf{x}}[\cdot]$ is the expectation with respect to $\mathbf{x}$.

The LMMSE estimator in (1) can perform poorly [3]-[5] if $(\widehat{\mathbf{C}}_{\mathbf{yy}}, \widehat{\mathbf{C}}_{\mathbf{yx}})$ are estimated from limited training data using the sample covariance matrices (SCMs):

$$\widehat{\mathbf{C}}_{\mathbf{yy}} = \frac{1}{T}\mathbf{Y}\mathbf{Y}^{\dagger}, \quad \widehat{\mathbf{C}}_{\mathbf{yx}} = \frac{1}{T}\mathbf{Y}\mathbf{X}^{\dagger}, \tag{3}$$

where $\mathbf{X} \in \mathbb{C}^{M\times T}$ and $\mathbf{Y} \in \mathbb{C}^{N\times T}$ consist of the length-T training data. Diagonal loading (DL) is a widely used approach to improve robustness, which leads to the regularized estimate

$$\widehat{\mathbf{x}} = \widehat{\mathbf{C}}_{\mathbf{yx}}^{\dagger}\left(\widehat{\mathbf{C}}_{\mathbf{yy}} + \gamma\mathbf{I}_N\right)^{-1}\mathbf{y}. \tag{4}$$

The diagonal loading factor (DLF) γ can significantly affect the achievable $\mathrm{MSE}_{\mathbf{x}}$ and must be carefully chosen. Ad hoc choices, e.g., [3, p. 748], can result in very conservative performance. More systematic methods have been proposed recently. In particular, [6] circumvents this problem by solving a covariance matrix estimation problem. Both [5] and [7] aim to maximize the output signal-to-interference-plus-noise ratio of the minimum variance distortionless response beamformer. The choice of the DLF that minimizes the $\mathrm{MSE}_{\mathbf{x}}$ are studied in [8] and [9]. Both schemes assume large systems and apply random matrix theory. Furthermore, they assume independent, identically distributed (i.i.d.) observations.

This paper introduces an alternative, automatic choice of the DLF γ for (4) in training-based applications. In contrast to [5, 6, 7], we aim to minimize directly the MSE of signal estimation. Furthermore, instead of evoking the large system assumption and random matrix theory as in [5, 8, 9], cross validation (CV) is applied to choose the DLF. For applications where the training and out-of-training data are identically distributed, we derive a CV scheme for determining the DLF that reuses the training data, which were originally deployed for producing $(\widehat{\mathbf{C}}_{\mathbf{yy}}, \widehat{\mathbf{C}}_{\mathbf{yx}})$. We repeatedly split the training data into two sets, one for estimating $(\widehat{\mathbf{C}}_{\mathbf{yy}}, \widehat{\mathbf{C}}_{\mathbf{yx}})$ and one for choosing γ. For applications where the training and out-of-training data are not identically distributed, an alternative CV scheme will be derived, which exploits the out-of-training symbols. This scheme makes use of the spatial correlation among the out-of-training outputs and chooses the DLF that optimizes the prediction of the output symbols. Computationally efficient schemes are obtained for both cases, which addresses the complexity concern of the standard CV scheme. It is shown by simulation examples that the DLF chosen can approach the oracle choice that minimizes $\mathrm{MSE}_{\mathbf{x}}$.

2. CV USING TRAINING DATA

We wish to choose the DLF γ for (4) that minimizes the MSE of signal estimation defined in (2). Given $\widehat{\mathbf{C}}_{\mathbf{yy}}$ and $\widehat{\mathbf{C}}_{\mathbf{yx}}$, the MSE of estimating $\mathbf{x}$ using (4) can be written as

$$\mathrm{MSE}_{\mathbf{x}} = \mathrm{tr}\left(\mathbf{C}_{\mathbf{xx}} - \mathbf{C}_{\mathbf{yx}}^{\dagger}\mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{C}_{\mathbf{yx}} + \mathbf{E}_{\gamma}^{\dagger}\mathbf{C}_{\mathbf{yy}}\mathbf{E}_{\gamma}\right), \quad (5)$$

where

$$\mathbf{E}_{\gamma} \triangleq \left(\widehat{\mathbf{C}}_{\mathbf{yy}} + \gamma\mathbf{I}_N\right)^{-1}\widehat{\mathbf{C}}_{\mathbf{yx}} - \mathbf{C}_{\mathbf{yy}}^{-1}\mathbf{C}_{\mathbf{yx}}. \quad (6)$$

If the actual covariance matrices $\mathbf{C}_{\mathbf{yy}}$ and $\mathbf{C}_{\mathbf{yx}}$ are known, the optimal γ that minimizes $\mathrm{MSE}_{\mathbf{x}}$ can be found, which will be referred to as the *oracle choice*. This can be achieved by solving the unconstrained, univariate optimization problem

$$\gamma_{\mathrm{oracle}}^{*} = \arg\min_{\gamma} \mathrm{tr}\left(\mathbf{E}_{\gamma}^{\dagger}\mathbf{C}_{\mathbf{yy}}\mathbf{E}_{\gamma}\right) \quad (7)$$

using standard optimization methods. The oracle choice, which is not realizable in practice, will be used as a benchmark for practical choices.

We apply the leave-one-out CV (LOOCV) principles [10, 11] to derive practical schemes whose performance approaches the oracle choice. We first consider i.i.d. training and out-of-training data, where the DLF can be determined using only the training data. The training data $(\mathbf{X}, \mathbf{Y})$ are repeatedly split into two sets. As illustrated in Fig. 1(a), for the i-th split, $T-1$ pairs of training symbols in $(\mathbf{X}_{\sim i}, \mathbf{Y}_{\sim i})$ (with $(\mathbf{x}_i, \mathbf{y}_i)$ omitted from $(\mathbf{X}, \mathbf{Y})$) are used for generating SCMs and the remaining one pair $(\mathbf{x}_i, \mathbf{y}_i)$ is spared for estimating the MSE of signal estimation for a different DLF. In total, T different splits are obtained and the DLF that optimizes the estimated performance will be chosen. The LOOCV method minimizes the average squared error of estimating $\mathbf{x}_i$ from $\mathbf{y}_i$ using an estimator $\mathbf{W}_{\sim i}$ derived from $(\mathbf{X}_{\sim i}, \mathbf{Y}_{\sim i})$:

$$\gamma^{*} = \arg\min_{\gamma} \frac{1}{T}\sum_{i=1}^{T} ||\mathbf{x}_i - \mathbf{W}_{\sim i,\gamma}^{\dagger}\mathbf{y}_i||^2, \quad (8)$$

where the estimator constructed from $(\mathbf{X}_{\sim i}, \mathbf{Y}_{\sim i})$ is given by

$$\mathbf{W}_{\sim i,\gamma} = \left(\mathbf{Y}_{\sim i}\mathbf{Y}_{\sim i}^{\dagger} + \gamma\mathbf{I}_N\right)^{-1}\mathbf{Y}_{\sim i}\mathbf{X}_{\sim i}^{\dagger}. \quad (9)$$

Then applying the Woodbury matrix identity, we can show that the error of predicting $\mathbf{x}_i$ using $\mathbf{W}_{\sim i}^{\dagger}\mathbf{y}_i$ can be written as

$$\mathbf{x}_i - \mathbf{W}_{\sim i,\gamma}^{\dagger}\mathbf{y}_i = \mathbf{x}_i - \left[\frac{\mathbf{X}\mathbf{Y}^{\dagger}\left(\mathbf{Y}\mathbf{Y}^{\dagger} + \gamma\mathbf{I}_N\right)^{-1}\mathbf{y}_i}{1 - \mathbf{y}_i\left(\mathbf{Y}\mathbf{Y}^{\dagger} + \gamma\mathbf{I}_N\right)^{-1}\mathbf{y}_i} - \mathbf{x}_i\frac{\mathbf{y}_i\left(\mathbf{Y}\mathbf{Y}^{\dagger} + \gamma\mathbf{I}_N\right)^{-1}\mathbf{y}_i}{1 - \mathbf{y}_i\left(\mathbf{Y}\mathbf{Y}^{\dagger} + \gamma\mathbf{I}_N\right)^{-1}\mathbf{y}_i}\right]. \quad (10)$$

By plugging this into (8) and after some manipulations, the optimal DLF is given by

$$\gamma^{*} = \arg\min_{\gamma} \left|\left|\mathbf{X} - \mathbf{X}\left(\mathbf{B}_{\gamma} - \mathbf{D}_{\mathbf{B}_{\gamma}}\right)\left(\mathbf{I} - \mathbf{D}_{\mathbf{B}_{\gamma}}\right)^{-1}\right|\right|^2 \quad (11)$$

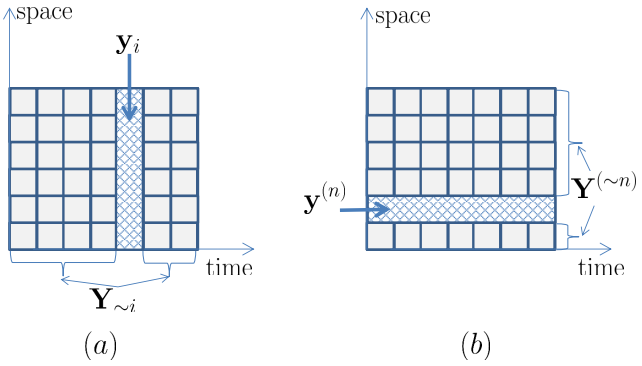

Fig. 1. Split of (a) training data and (b) out-of-training data for the LOOCV schemes in Section 2 and 3, respectively, where $\mathbf{y}_i$ and $\mathbf{y}^{(n)} \triangleq [y_1^{(n)}, y_2^{(n)}, \cdots, y_D^{(n)}]$ denote the symbols spared for DLF validation in the i-th and n-th split, respectively.

where

$$\mathbf{B}_{\gamma} \triangleq \mathbf{Y}^{\dagger}\left(\mathbf{Y}\mathbf{Y}^{\dagger} + \gamma\mathbf{I}_N\right)^{-1}\mathbf{Y}. \quad (12)$$

Computing the SVD of $\mathbf{Y}$ can facilitate the calculation of $\mathbf{B}_{\gamma}$ and the cost function in (11) for different candidates of γ.

3. CV USING OUT-OF-TRAINING DATA

The LOOCV scheme in Section 2 chooses the DLF using only the training data. It cannot be applied if the covariance matrices are not estimated as SCMs. Furthermore, the DLF choice given by (11) can perform poorly when the training symbols are distributed differently than the out-of-training symbols. This is the case, for example, when orthogonal signaling or higher power is applied to the training symbols to obtain high-quality estimates of the covariance matrices. This may be encountered in applications like wireless communications where the training symbols can be optimized [1].

This section introduces an alternative LOOCV scheme to address the above limitations of training-based LOOCV schemes. A block of out-of-training symbols, denoted by $\mathbf{y}_d, d = 1, 2, \cdots, D$, are exploited. Note that the input symbols $\mathbf{x}_d$ that lead to $\mathbf{y}_d$ are unknown and need to be estimated. We assume that the estimated covariance matrices $\widehat{\mathbf{C}}_{\mathbf{yx}}$ and $\widehat{\mathbf{C}}_{\mathbf{yy}}$ are given. Similarly to Section 2 where the T training symbols are split with respect to time, each out-of-training symbol $\mathbf{y}_d$ here is also split repeatedly, as illustrated in Fig. 1(b). Let $y_d^{(n)}$ be the n-th entry of $\mathbf{y}_d$ and $\mathbf{y}_d^{(\sim n)}$ the vector obtained by excluding $y_d^{(n)}$ from $\mathbf{y}_d$. We now choose γ to minimize the average squared error of predicting $y_d^{(n)}$ from $\mathbf{y}_d^{(\sim n)}$ using the estimated covariance matrices $(\widehat{\mathbf{C}}_{\mathbf{yx}}, \widehat{\mathbf{C}}_{\mathbf{yy}})$, i.e.,

$$\gamma^{*} = \arg\min_{\gamma} \frac{1}{ND}\sum_{d=1}^{D}\sum_{n=1}^{N}\left|\left|y_d^{(n)} - \widehat{y}_{d,\gamma}^{(n)}\right|\right|^2. \quad (13)$$

The intuition is that the out-of-training observations can be spatially correlated and this correlation can be exploited to provide an LMMSE prediction of $y_d^{(n)}$ from $\mathbf{y}_d^{(\sim n)}$. If a DLF γ improves such a prediction, $\widehat{\mathbf{C}}_{\mathbf{yy}} + \gamma \mathbf{I}_N$ may outperform $\widehat{\mathbf{C}}_{\mathbf{yy}}$ in estimating $\mathbf{C}_{\mathbf{yy}}$, which may improve the estimation of $\mathbf{x}$.

Define $\widehat{\mathbf{c}}_n$ as the estimated cross-covariance of $\mathbf{y}_d^{(\sim n)}$ and $y_d^{(n)}$, obtained as the n-th column of $\widehat{\mathbf{C}}_{\mathbf{yy}}$ with the n-th entry excluded, and $\widehat{\mathbf{C}}_{\sim n}$ the estimated auto-covariance matrix of $\mathbf{y}_d^{(\sim n)}$, constructed by excluding the n-th row and n-th column of $\widehat{\mathbf{C}}_{\mathbf{yy}}$. The LMMSE predictor of $y_d^{(n)}$ from $\mathbf{y}_d^{(\sim n)}$ with DLF γ is then given by

$$\begin{aligned} \widehat{y}_{d,\gamma}^{(n)} &= \widehat{\mathbf{c}}_n^{\dagger} \left(\widehat{\mathbf{C}}_{\sim n} + \gamma \mathbf{I}_{N-1} \right)^{-1} \mathbf{y}_d^{(\sim n)} \\ &= \mathbf{w}_{n,\gamma}^{\dagger} \mathbf{y}_d^{(\sim n)}, \end{aligned} \tag{14}$$

where

$$\mathbf{w}_{n,\gamma} = \left(\widehat{\mathbf{C}}_{\sim n} + \gamma \mathbf{I}_{N-1} \right)^{-1} \widehat{\mathbf{c}}_n. \tag{15}$$

The direct implementation of the LOOCV method in (13) involves the inversion of an $(N-1) \times (N-1)$ matrix for each n and γ. This can be avoided. To see this, take $n = N$ as an example. From the Woodbury matrix identity,

$$\begin{aligned} \mathbf{\Phi}_\gamma &\triangleq \left(\widehat{\mathbf{C}}_{\mathbf{yy}} + \gamma \mathbf{I}_N \right)^{-1} \\ &\triangleq \begin{bmatrix} \widehat{\mathbf{C}}_{\sim N} + \gamma \mathbf{I}_{N-1} & \widehat{\mathbf{c}}_N \\ \widehat{\mathbf{c}}_N^{\dagger} & \widehat{c}_N + \gamma \end{bmatrix}^{-1} \\ &= \begin{bmatrix} \times & -\frac{(\widehat{\mathbf{C}}_{\sim N} + \gamma \mathbf{I}_{N-1})^{-1} \widehat{\mathbf{c}}_N}{\widehat{c}_N + \gamma - \widehat{\mathbf{c}}_N^{\dagger} (\widehat{\mathbf{C}}_{\sim N} + \gamma \mathbf{I}_{N-1})^{-1} \widehat{\mathbf{c}}_N} \\ \times & \frac{1}{\widehat{c}_N + \gamma - \widehat{\mathbf{c}}_N^{\dagger} (\widehat{\mathbf{C}}_{\sim N} + \gamma \mathbf{I}_{N-1})^{-1} \widehat{\mathbf{c}}_N} \end{bmatrix}, \end{aligned} \tag{16}$$

where $\widehat{c}_N$ denotes the n-th diagonal entry of $\widehat{\mathbf{C}}_{\mathbf{yy}}$ and $\times$ represents entries which are not of interests here. Comparing this with (15) we can find

$$\mathbf{w}_{N,\gamma} = \frac{-1}{[\mathbf{\Phi}_\gamma]_{N,N}} [\mathbf{\Phi}_\gamma]_{1:N-1,N}, \tag{17}$$

where $[\mathbf{\Phi}_\gamma]_{N,N}$ denotes the (N,N)-th entry of $[\mathbf{\Phi}_\gamma]$ and $[\mathbf{\Phi}_\gamma]_{1:N-1,N}$ the vector consisting of the first $N-1$ entries of the N-th column of $[\mathbf{\Phi}_\gamma]$. As such, we can write the prediction error as

$$\begin{aligned} y_d^{(N)} - \widehat{y}_{d,\gamma}^{(N)} &= - \begin{bmatrix} \mathbf{w}_{N,\gamma} \\ -1 \end{bmatrix}^{\dagger} \begin{bmatrix} \mathbf{y}_d^{(\sim N)} \\ y_d^{(N)} \end{bmatrix} \\ &= \left(\frac{[\mathbf{\Phi}_\gamma]_{:,N}}{[\mathbf{\Phi}_\gamma]_{N,N}} \right)^{\dagger} \mathbf{y}_d. \end{aligned} \tag{18}$$

We can find a similar relationship for $y_d^{(n)}$, $n = 1, 2, \cdots, N-1$, as

$$y_d^{(n)} - \widehat{y}_{d,\gamma}^{(n)} = \left(\frac{[\mathbf{\Phi}_\gamma]_{:,n}}{[\mathbf{\Phi}_\gamma]_{n,n}} \right)^{\dagger} \mathbf{y}_d. \tag{19}$$

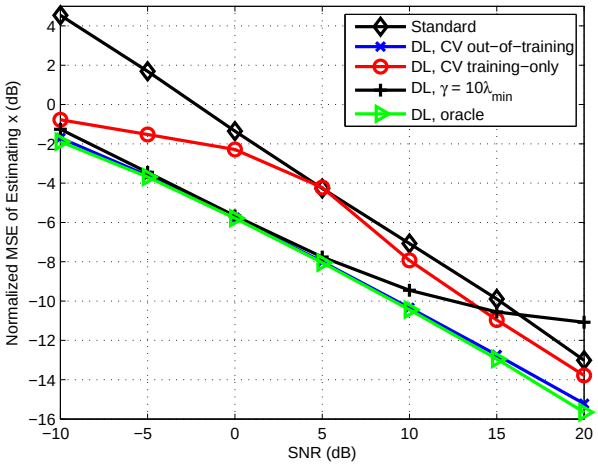

Fig. 2. Normalized MSE versus SNR with orthogonal training of length $T = 24$ constructed from the discrete Fourier transform (DFT) matrix. $D = 24$ for the CV using out-of-training data.

Summarizing,

$$\sum_{n=1}^{N} |y_d^{(n)} - \widehat{y}_{d,\gamma}^{(n)}|^2 = \left|\left| \left[\mathbf{\Phi}_\gamma \mathbf{D}_{\mathbf{\Phi}_\gamma}^{-1} \right]^{\dagger} \mathbf{y}_d \right|\right|^2, \tag{20}$$

where $\mathbf{D}_{\mathbf{\Phi}_\gamma}$ denotes the diagonal matrix that shares the diagonal entries of $\mathbf{\Phi}_\gamma$. We can now write the optimal DLF as

$$\gamma^* = \arg\min_{\gamma} \frac{1}{ND} \left|\left| \mathbf{D}_{\mathbf{\Phi}_\gamma}^{-1} \mathbf{\Phi}_\gamma \mathbf{Y} \right|\right|^2, \tag{21}$$

where

$$\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_D]$$

collects the D out-of-training symbols used for choosing the DLF. From (21) only one matrix inversion is needed for each γ. The calculations of $\mathbf{\Phi}_\gamma$ for different γ can be implemented by reusing the eigenvalue decomposition (EVD) of $\widehat{\mathbf{C}}_{\mathbf{yy}}$.

4. EXAMPLES

We consider an example of MIMO systems modelled by

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z}, \tag{22}$$

where $\mathbf{H} \in \mathbb{C}^{N \times M}$ is the channel matrix, and $\mathbf{z} \in \mathbb{C}^{N \times 1}$ is the white noise, which is uncorrelated with $\mathbf{x}$. We assume that $M = N = 20$, $\mathbf{H}$ has i.i.d., complex Gaussian entries with unit variance, and $\mathbf{z}$ has i.i.d., complex Gaussian entries with variance ρ. The signal to noise ratio (SNR) is $\mathrm{SNR} = 1/\rho$.

We first consider a case where the training block $\mathbf{X}$ is orthogonal, while the out-of-training data are i.i.d., complex Gaussian. Fig. 2 shows the normalized MSE

$$\frac{\mathrm{E}[||\widehat{\mathbf{x}} - \mathbf{x}||]^2}{\mathrm{E}[||\mathbf{x}||]^2}$$

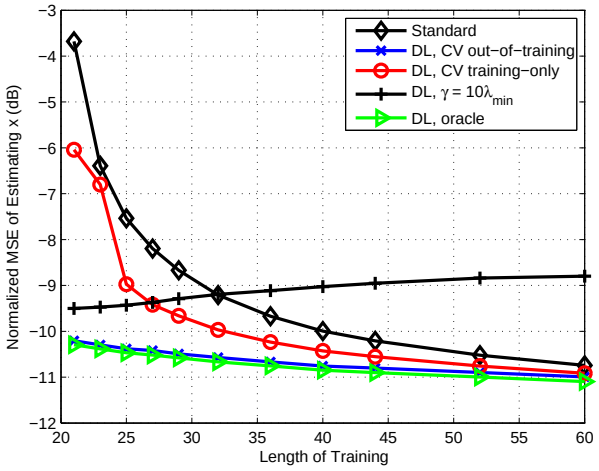

Fig. 3. Normalized MSE versus the length of orthogonal training T for $\mathrm{SNR} = 10$ dB. $D = 24$ for the CV scheme using out-of-training data.

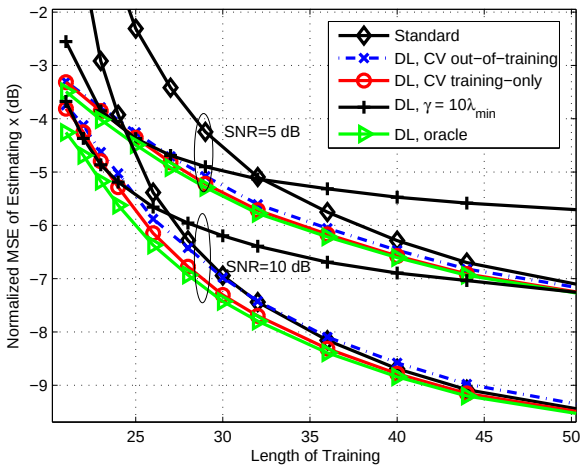

Fig. 4. Normalized MSE versus the length of nonorthogonal training T. $D = 24$ for the CV scheme using out-of-training data.

of estimating $\mathbf{x}$ using different choices of DLFs when the length of training $T = 24$. Fig. 3 shows the comparisons when $\mathrm{SNR} = 10$ dB and T varies. We can see that compared to the standard scheme based on SCMs (with $\gamma = 0$), the oracle choice of the DLF can provide significant performance improvements, demonstrating the potential of diagonal loading. Among the two CV schemes proposed, the one that uses the out-of-training data significantly outperforms that using only the training data, and it approaches the oracle choice in most cases shown. This is because the training and out-of-training data are not identically distributed. We also show the performance of the DL scheme using the empirical choice [3]

$$\gamma = 10\lambda_{\min},$$

where $\lambda_{\min}$ is the smallest eigenvalue of $\widehat{\mathbf{C}}_{\mathbf{yy}}$. It is seen that this empirical choice performs well when the SNR is low or the length of training is small. However, when the SNR is high or length of training is large, its performance degrades significantly as $\gamma = 10\lambda_{\min}$ becomes too large.

Fig. 4 shows the performance when the training and out-of-training data are identically distributed. It is seen that both CV schemes perform closely to the oracle choice. Note that the CV scheme using only the training data performs slightly better in this example and does not need to exploit the out-of-training data. Comparing Figs. 3 and 4, we also see that using orthogonal training may significantly improve the performance of signal estimation.

5. CONCLUSIONS

We have introduced two CV approaches for choosing the DLF for linear signal estimation based on SCMs. We derived closed-form expressions for the cost function of CV based on training and out-of-training data, respectively. The latter approach can be noticeably better when specially tailored training data are used, while the former has the advantage that it does not need to exploit any out-of-training data.

6. REFERENCES

[1] M. Biguesh and A. B. Gershman, "Training-based MIMO channel estimation: a study of estimator tradeoffs and optimal training signals," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 884–893, 2006.

[2] L. L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*, Addison–Wesley, Bosten, 1991.

[3] H. L. Van Trees, *Optimum Array Processing.* NewYork, NY,USA: Wiley, 2002.

[4] J. Tong and P. J. Schreier, "A unified framework for regularized linear estimation in communication systems," *(Elsevier) Signal Process.*, vol. 93, no. 9, pp. 2671-2686, 2013.

[5] X. Mestre and M. A. Lagunas, "Finite sample size effect on minimum variance beamformers: Optimum diagonal loading factor for large arrays," *IEEE Trans. Sig. Process.*, vol. 54, no. 1, pp. 69-82, 2006.

[6] P. Stoica , J. Li , X. Zhu and J. R. Guerci, "On using a priori knowledge in space-time adaptive processing," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2598-2602, 2008.

[7] L. Du, J. Li, and P. Stoica, "Fully automatic computation of diagonal loading levels for robust adaptive beamforming," *IEEE Trans. Aero. Elect. Sys.*, vol. 46, no. 1, pp. 449-458, Jan. 2010.

[8] M. Zhang, F. Rubio, D. Palomar, and X. Mestre, "Finite-sample linear filter optimization in wireless communications and financial systems," *IEEE Trans. Sig. Process.,* vol. 61, no. 20, pp. 5014-5025, 2013.

[9] J. Serra and M. Najar, "Asymptotically optimal linear shrinkage of sample LMMSE and MVDR filters," *IEEE Trans. Sig. Process.*, vol. 62, no. 14, pp. 3552-3564, 2014.

[10] S. Arlot and A. Celisse, "A survey of cross-validation procedures for model selection," *Statist. Surv.*, vol. 4, pp. 40-79, 2010.

[11] R. D. Nowak, "Optimal signal estimation using cross-validation," *IEEE Sig. Process. Letters,* vol. 4, no. 1, pp. 23-25, 1997.