[11] B. Picinbono and P. Bondon, "Second-order statistics of complex signals," *IEEE Trans. Signal Process.*, vol. 45, no. 2, pp. 411–420, Feb. 1997.
[12] R. A. Hedges and B. W. Suter, "Numerical spread: Quantifying local stationarity," *Digital Signal Process.*, vol. 12, pp. 628–643, 2002.
[13] F. D. Neeser and J. L. Massey, "Proper complex random processes with applications to information theory," *IEEE Trans. Inf. Theory*, vol. 39, no. 7, pp. 1293–1302, Jul. 1993.
[14] H. Hindberg and S. C. Olhede, Estimation of Ambiguity Functions With Limited Spread Dept. Statist. Sci., Univ. College London, London, U.K., 2007 [Online]. Available: http://arxiv.org/abs/0804.1038
[15] M. Peligrad, "On the asymptotic normality of sequences of weak dependent random variables," *J. Theo. Prob.*, vol. 9, pp. 703–715, 1996.
[16] T. S. Ferguson*, A Course in Large Sample Theory*. London, U.K.: Chapman & Hall, 1996.
[17] A. N. Kolmogorov and Y. A. Rozanov, "On strong mixing conditions for stationary Gaussian processes," *Theory. Prob. A*, vol. 5, pp. 204–208, 1960.
[18] L. Wasserman*, All of Nonparametric Statistics* . New York: Springer Verlag, 2007.
[19] S. Olhede, "Hyperanalytic denoising," *IEEE Trans. Image Process.*, vol. 16, no. 6, pp. 1522–1537, Jun. 2007.
[20] R. R. Coifman and D. L. Donoho, "Translation-invariant denoising," in *Wavelets and Statistics*, A. Antoniadis and G. Oppenheim, Eds. New York: Springer-Verlag, 1995.
[21] I. M. Johnstone and B. W. Silverman, "Wavelet threshold estimators for data with correlated noise," *J. Roy. Statist. Soc. B*, vol. 59, pp. 319–351, 1997.

Competitive Linear Estimation Under Model Uncertainties

Suleyman S. Kozat and Alper T. Erdogan

***Abstract*—We investigate a linear estimation problem under model uncertainties using a competitive algorithm framework under mean square error (MSE) criteria. Here, the performance of a linear estimator is defined relative to the performance of the linear minimum MSE estimator tuned to the underlying unknown system model. We then find the linear estimator that minimizes this relative performance measure, i.e., the regret, for the worst possible system model. Two definitions of regret are given: first as a difference of MSEs and second as a ratio of MSEs. We demonstrate that finding the linear estimators that minimize these regret definitions can be cast as a Semidefinite Programming (SDP) problem and provide numerical examples.**

***Index Terms*—Competitive, convex optimization, linear estimation, regret, uncertainties.**

I. Introduction

In this correspondence, a basic linear estimation problem is investigated from a competitive algorithm framework under mean square error (MSE) criteria. Here, a desired unknown data vector with a known correlation matrix is observed through an unknown linear system, where the output of the system is corrupted by additive noise with a known correlation matrix. Although, the underlying linear system is unknown, an estimate of it is given (or produced), which may contain possible uncertainties (or inaccuracies). Based on the observations, an estimate of the desired data vector is produced using a linear estimator. However, since the underlying system is not accurately known, it may not be possible to directly choose this linear estimator as the linear minimum MSE (MMSE) estimator.

A common approach to solve such estimation problems under model uncertainties is to pose the estimation problem in a worst-case performance-optimization-based framework [1], [2]. Especially, in [2] and in the references therein, the H^{∞} criterion has been applied to the linear estimation model. According to this criterion, the signals in the setup are modeled as deterministic (unknown) disturbances and the maximum energy gain from the input signals to the output estimation errors are minimized. However, in this correspondence, we refrain from such a cost function and the deterministic formulation of the input sequences. Instead, we investigate a competitive approach inspired by [3], where the overall performance is defined based on relative MSEs. In this competitive framework, the performance of a linear estimator is defined relative to the performance of the linear MMSE estimator tuned to the underlying unknown channel. We then seek for the linear estimator that minimizes this relative performance, i.e., the regret for committing to a linear estimator that is not the linear MMSE estimator.

In this correspondence, we investigate two different "regret" formulations. The first formulation defines the regret committing to a particular linear estimator as the difference between the MSE of this linear estimator and the MSE of the linear MMSE estimator tuned to the underlying model. In the second formulation, the regret is defined as the "ratio" between the MSE of this linear estimator divided by the MSE of the linear MMSE estimator tuned to the underlying model. We emphasize that although defining the regret as a difference between MSEs is well studied in the literature [3]–[5], defining regret as the ratio of MSEs is introduced in here to our knowledge.

The linear estimation framework investigated in this correspondence could be used to model certain digital communication scenarios, where the underlying channel coefficients are not known accurately. In such applications, the statistics of the desired signal, e.g., the transmitted data, and the noise process, which can be readily estimated from the observed data for independent noise processes, are usually assumed to be known. The underlying unknown channel may be estimated using either blind or supervised estimation algorithms. However, inaccuracies may exist due to the limited training data, the presence of noise, and/or the time variations in the channel. The intended linear estimator is then the linear equalizer that optimizes the MSE performance. We note that when the underlying channel has a finite-impulse response (FIR), then the linear system model has a convolution matrix structure constructed using the FIR channel coefficients. However, even in this case the algorithms introduced in here can be used directly.

The problem studied in this correspondence within the competitive algorithm framework is investigated in [3], where uncertainties were present in the correlation matrices of the input and noise processes, but the underlying system model was assumed to be known. However, in this correspondence, we cover the complementary case where the uncertainty is present in the underlying system model and the correlation matrices of the input and noise process are known. Furthermore, we solve the underlying problem under model uncertainties for two different regret definitions, i.e., regret as the difference of MSEs and regret as the ratio of MSEs. In both cases, we demonstrate that finding the linear estimators that minimize the worst case regrets, can be cast as semidefinite programming (SDP) problems. We should emphasize

Manuscript received April 14, 2009; accepted September 29, 2009. First published November 20, 2009; current version published March 10, 2010. This work is supported in part by TUBITAK Career Award under Contract 104E073 and Contract 108E195 and in part by the Turkish Academy of Sciences GEBIP Program. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Isao Yamada.

The authors are with the Electrical and Electronics Engineering Department, Koc University, 34450 Istanbul, Turkey (e-mail: skozat@ku.edu.tr; alperdogan@ku.edu.tr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2009.2037066

that SDP problems are convex optimization problems, where efficient algorithms such as the interior point methods, are available for their solutions [6]. In this sense, the desired linear estimators can be found efficiently using the introduced methods.

We note that although the well-known H^∞ framework [2] also uses similar minimax formulation as an optimization tool, there are important differences in the competitive framework studied here. In the H^∞ estimation framework, the cost function that is optimized is the maximum energy gain from input disturbances to the output estimation errors. The uncertainty in H^∞ case is in the signals, where H^∞ approach treats these disturbances as completely deterministic signals. Therefore, this criterion amounts to minimizing the ratio of the error signal energy to the energy of the disturbances for all possible signals with nonzero energy. However, in the competitive framework, the cost function is completely different and the signals in the estimation setup are not taken as deterministic sequences but taken as stochastic signals. The uncertainty is not in the signals, but in the linear mapping of the observation setup describing the channel.

This correspondence paper is organized as follows. We first introduce the basic problem setup in Section II. The regret formulations and corresponding results of this correspondence paper as theorems follow in Section III. We produce numerical results in Section IV. The correspondence concludes with couple of remarks.

II. System Description

In this correspondence, all vectors are column vectors and represented by boldface lowercase letters. Matrices are represented by boldface uppercase letters. Given a vector $\mathbf{x}$, $\|\mathbf{x}\| = \mathbf{x}^H\mathbf{x}$ is the l_2-norm, where $\mathbf{x}^H$ is the conjugate transpose and $\mathbf{x}^T$ is the ordinary transpose. For a matrix $\mathbf{R}$, $\mathrm{Tr}[\mathbf{R}]$ is the trace, $\|\mathbf{R}\|$ is the spectral norm, $\mathbf{R} > 0$ represents a positive define matrix and $\mathbf{R} \geq 0$ represents a positive-semidefinite matrix.

In this problem, an unknown desired vector $\mathbf{x} \in \mathbb{C}^N$ is observed through an unknown linear system $\mathbf{H}$, where the output of the system is corrupted by additive noise, i.e.,

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$$

$\mathbf{H} \in \mathbb{C}^{M\times N}$ and $\mathbf{w} \in \mathbb{C}^M$. Here, $\mathbf{x}$ is zero mean with known correlation matrix $\mathbf{R}_x > 0$, $\mathbf{w}$ is the corrupting noise vector, independent from $\mathbf{x}$, with zero mean and known correlation matrix $\mathbf{R}_w > 0$. Although, $\mathbf{H}$ is unknown, an estimate of $\mathbf{H}$ is assumed to be available, which is denoted as $\mathbf{H}_0$. This estimate is usually imperfect and contains an uncertainty

$$\Delta\mathbf{H} = \mathbf{H} - \mathbf{H}_0$$

which is assumed to be bounded, i.e., $\|\Delta\mathbf{H}\| \leq \alpha$, $\alpha > 0$. We assume that α or a bound on α is known.

After observing $\mathbf{y}$, an estimate of the data vector $\mathbf{x}$ using a linear estimator $\mathbf{G}$ is constructed as

$$\begin{aligned}\hat{\mathbf{x}} &= \mathbf{G}\mathbf{y} = \mathbf{G}(\mathbf{H}\mathbf{x} + \mathbf{w}),\\ &= \mathbf{G}\left[(\mathbf{H}_0 + \Delta\mathbf{H})\mathbf{x} + \mathbf{w}\right]\end{aligned}$$

where $\mathbf{G} \in \mathbb{C}^{N\times M}$. Given this linear model and estimator, the estimation error is defined as

$$\begin{aligned}E\left[\|\mathbf{x} - \hat{\mathbf{x}}\|^2\right] &= \mathrm{Tr}\left[(\mathbf{I} - \mathbf{G}\mathbf{H})\mathbf{R}_x(\mathbf{I} - \mathbf{G}\mathbf{H})^H\right]\\ &\quad + \mathrm{Tr}[\mathbf{G}\mathbf{R}_w\mathbf{G}^H]\\ &= \mathrm{Tr}\left[\mathbf{R}_x(\mathbf{I} - \mathbf{G}\mathbf{H})^H(\mathbf{I} - \mathbf{G}\mathbf{H})\right] + \mathrm{Tr}[\mathbf{G}\mathbf{R}_w\mathbf{G}^H]\\ &= \mathrm{Tr}\Big[\mathbf{R}_x^{\frac{1}{2}}\left[\mathbf{I} - \mathbf{G}(\mathbf{H}_0 + \Delta\mathbf{H})\right]^H\\ &\qquad \times \left[\mathbf{I} - \mathbf{G}(\mathbf{H}_0 + \Delta\mathbf{H})\right]\mathbf{R}_x^{\frac{1}{2}}\Big] + \mathrm{Tr}[\mathbf{G}\mathbf{R}_w\mathbf{G}^H].\end{aligned} \tag{1}$$

For any $\mathbf{H}$, the optimal linear estimator in MSE sense tuned to $\mathbf{H}$ is given by [7]

$$\begin{aligned}\mathbf{G}_{\mathrm{MMSE}}(\mathbf{H}) &\triangleq \arg\min_{\mathbf{G}} E\left[\|\mathbf{x} - \mathbf{G}(\mathbf{H}\mathbf{x} + \mathbf{w})\|^2\right]\\ &= \left(\mathbf{R}_x^{-1} + \mathbf{H}^H\mathbf{R}_w^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^H\mathbf{R}_w^{-1}\end{aligned}$$

with the corresponding MMSE

$$J_{\mathrm{MMSE}}(\mathbf{H}) \triangleq \min_{\mathbf{G}} E\left[\|\mathbf{x} - \hat{\mathbf{x}}\|^2\right] \tag{2}$$

$$= \mathrm{Tr}\left[\left(\mathbf{R}_x^{-1} + \mathbf{H}^H\mathbf{R}_w^{-1}\mathbf{H}\right)^{-1}\right]. \tag{3}$$

Since we assume that only $\mathbf{H}_o$ an erroneous estimate of the linear system $\mathbf{H}$ is available, the direct use of MMSE formulation provided above based on this estimate would not have a reliable performance. For this purpose, we propose the use of competitive approach, where the information about the uncertainty in the estimate enters into the estimation formulation. In the next section, we define two different minimax regret formulations, i.e., certain relative performance measures as a part of this competitive framework. These regret formulations target to optimize the variation in the MSE performance relative to the linear MMSE estimator with the exact knowledge of the underlying channel. In evaluating this relative performance, we use both the difference and the ratio as two alternative approaches leading to the corresponding alternative regret-based estimator formulations.

III. Regret Formulations

A. Regret in Additive Form

For an unknown linear system $\mathbf{H}$, we define our regret for committing to a particular linear estimator $\mathbf{G}$ [not to $\mathbf{G}_{\mathrm{MMSE}}(\mathbf{H})$] as the difference between the MSE of $\mathbf{G}$ and the scaled MMSE that can be achievable by the linear MMSE estimator $\mathbf{G}_{\mathrm{MMSE}}(\mathbf{H})$ tuned to $\mathbf{H}$ as

$$E\left[\|\mathbf{x} - \mathbf{G}(\mathbf{H}\mathbf{x} + \mathbf{w})\|^2\right] - \rho J_{\mathrm{MMSE}}(\mathbf{H}) \tag{4}$$

for any arbitrary $\rho > 0$. Note that ρ is usually selected as $\rho = 1$, such as in [3] and [4]. Furthermore, selecting $\rho = 0$ also yields the minimax MSE framework investigated in the first part of [3]. We then seek for the linear estimator that minimizes the worst case regret, i.e.,

$$\mathbf{G}^* = \arg\min_{\mathbf{G}}\max_{\mathbf{H}=\mathbf{H}_0+\Delta\mathbf{H}}\left\{E\left[\|\mathbf{x} - \mathbf{G}(\mathbf{H}\mathbf{x}+\mathbf{w})\|^2\right] - \rho J_{\mathrm{MMSE}}(\mathbf{H})\right\} \tag{5}$$

with a norm constraint on $\Delta\mathbf{H}$ such that $\|\Delta\mathbf{H}\| \leq \alpha$, $\alpha \in \mathbb{R}^+$.

For the regret definition, using (1) and (3) in (4) yields

$$\begin{aligned}&E\left[\|\mathbf{x} - \mathbf{G}(\mathbf{H}\mathbf{x} + \mathbf{w})\|^2\right] - \rho J_{\mathrm{MMSE}}(\mathbf{H})\\ &\quad = \mathrm{Tr}[\mathbf{G}\mathbf{R}_w\mathbf{G}^H] + \mathrm{Tr}\left[\mathbf{R}_x^{\frac{1}{2}}(\mathbf{I} - \mathbf{G}\mathbf{H})^H(\mathbf{I} - \mathbf{G}\mathbf{H})\mathbf{R}_x^{\frac{1}{2}}\right]\\ &\qquad - \rho\,\mathrm{Tr}\left[\left(\mathbf{R}_x^{-1} + \mathbf{H}^H\mathbf{R}_w^{-1}\mathbf{H}\right)^{-1}\right].\end{aligned} \tag{6}$$

The regret expression obtained above can be simplified to a more tractable form, by replacing the $J_{\mathrm{MMSE}}(\mathbf{H})$ term with its first-order (linear) approximation around the estimate $\mathbf{H}_o$ that is provided in Lemma 1 in the Appendix. Based on this approximation, the regret cost function can be rewritten as

$$\begin{aligned}&E\left[\|\mathbf{x} - \mathbf{G}(\mathbf{H}\mathbf{x}+\mathbf{w})\|^2\right] - \rho J_{\mathrm{MMSE}}(\mathbf{H})\\ &\quad \approx \mathrm{Tr}[\mathbf{G}\mathbf{R}_w\mathbf{G}^H]\\ &\qquad + \mathrm{Tr}\left[\mathbf{R}_x^{\frac{1}{2}}\left(\mathbf{I} - \mathbf{G}[\mathbf{H}_0+\Delta\mathbf{H}]\right)^H\left(\mathbf{I} - \mathbf{G}[\mathbf{H}_0+\Delta\mathbf{H}]\right)\mathbf{R}_x^{\frac{1}{2}}\right]\\ &\qquad - \rho a - \rho 2\Re e\left\{\mathrm{Tr}[\Delta\mathbf{H}^H\mathbf{B}]\right\}\end{aligned} \tag{7}$$

$$= \mathrm{Tr}[\mathbf{G}\mathbf{R}_w\mathbf{G}^H] + \mathrm{Tr}\left[\mathbf{R}_x^{\frac{1}{2}}\left(\mathbf{I}-\mathbf{G}[\mathbf{H}_0+\Delta\mathbf{H}]\right)^H\left(\mathbf{I}-\mathbf{G}[\mathbf{H}_0+\Delta\mathbf{H}]\right)\mathbf{R}_x^{\frac{1}{2}}\right] - \rho a - \rho 2\Re e\left\{\mathrm{Tr}\left[\mathbf{R}_x^{\frac{1}{2}}\Delta\mathbf{H}^H\mathbf{B}_x^{-\frac{1}{2}}\right]\right\} \quad (8)$$

where we omitted the $o(\cdot)$ term in (7) and use $\mathrm{Tr}[\Delta\mathbf{H}^H\mathbf{B}] = \mathrm{Tr}[\mathbf{R}_x^{1/2}\Delta\mathbf{H}^H\mathbf{B}\mathbf{R}_x^{-(1/2)}]$ in (8). Note that the first-order approximation is introduced in order to make the solution of (4) in a minimax setting tractable. Clearly, the effect of this approximation diminishes as $\|\Delta\mathbf{H}\|$ gets smaller. For distortions with larger $\|\Delta\mathbf{H}\|$, one can use the higher order approximations instead. However, we have observed through our simulations that the solution using the first-order approximation yields successful results even for fairly large $\|\Delta\mathbf{H}\|$ (when compared to $\|\mathbf{H}\|$). In order to obtain the linear estimator to minimize the worst case regret in (8), we have the following theorem, which formulates the underlying problem as an SDP problem.

Theorem 1: Suppose a desired unknown data vector $\mathbf{x}$ with zero mean and known correlation matrix $\mathbf{R}_x > 0$ is observed through an unknown linear system $\mathbf{H}$ as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$$

where $\mathbf{H} \in \mathbb{C}^{M\times N}$ and the corrupting noise vector $\mathbf{w} \in \mathbb{C}^M$, independent of $\mathbf{x}$, is zero mean with known correlation matrix $\mathbf{R}_w > 0$. Given an estimate of $\mathbf{H}$ as $\mathbf{H}_0$, then the problem

$$\underset{\mathbf{G}}{\text{minimize}}\ \underset{\mathbf{H}=\mathbf{H}_0+\Delta\mathbf{H}}{\text{maximize}} \Big\{\|\mathrm{Tr}[\mathbf{G}\mathbf{R}_w\mathbf{G}^H] + \mathrm{Tr}\left[\mathbf{R}_x^{\frac{1}{2}}(\mathbf{I}-\mathbf{G}\mathbf{H})^H(\mathbf{I}-\mathbf{G}\mathbf{H})\mathbf{R}_x^{\frac{1}{2}}\right] - \rho a - \rho 2\Re e\left(\mathrm{Tr}\left[\mathbf{R}_x^{\frac{1}{2}}\Delta\mathbf{H}^H\mathbf{B}\mathbf{R}_x^{-\frac{1}{2}}\right]\right)\Big\}, \quad \|\Delta\mathbf{H}\| \le \alpha \quad (9)$$

is equivalent to the SDP problem (see (10) and (11), shown at the bottom of page) where $\alpha, \rho > 0$, $\mathbf{M} = \mathbf{M}^H$, a and $\mathbf{B}$ are given in (28) and (29), respectively.

Proof of Theorem 1: We first observe that the problem in (9) is equivalent to the minimization problem

$$\min_{\mathbf{G}}\max_{\|\Delta\mathbf{H}\|\le\alpha} \mathrm{Tr}[\mathbf{G}\mathbf{R}_w\mathbf{G}^H] + \mathrm{Tr}\left[\mathbf{R}_x^{\frac{1}{2}}(\mathbf{I}-\mathbf{G}\mathbf{H})^H(\mathbf{I}-\mathbf{G}\mathbf{H})\mathbf{R}_x^{\frac{1}{2}}\right] - \rho a - \rho 2\Re e\left\{\mathrm{Tr}\left[\mathbf{R}_x^{\frac{1}{2}}\Delta\mathbf{H}^H\mathbf{B}\mathbf{R}_x^{-\frac{1}{2}}\right]\right\} = \min_{t,\mathbf{G}} t$$

such that

$$\mathrm{Tr}\Big\{\mathbf{G}\mathbf{R}_w\mathbf{G}^H + \mathbf{R}_x^{\frac{1}{2}}(\mathbf{I}-\mathbf{G}\mathbf{H})^H(\mathbf{I}-\mathbf{G}\mathbf{H})\mathbf{R}_x^{\frac{1}{2}} - (\rho a/N)\mathbf{I} - \rho\mathbf{R}_x^{\frac{1}{2}}\Delta\mathbf{H}^H\mathbf{B}\mathbf{R}_x^{-\frac{1}{2}} - \rho\mathbf{R}_x^{-\frac{1}{2}}\mathbf{B}^H\Delta\mathbf{H}\mathbf{R}_x^{\frac{1}{2}}\Big\} \le t, \quad \|\Delta\mathbf{H}\| \le \alpha \quad (12)$$

and $\mathbf{H} = \mathbf{H}_0 + \Delta\mathbf{H}$. Defining an intermediate matrix $\mathbf{M} = \mathbf{M}^H \in \mathbb{C}^{N\times N}$, the inequality in (12) can be written as

$$\mathrm{Tr}[\mathbf{M}] \le t \quad (13)$$

$$\Big\{\mathbf{G}\mathbf{R}_w\mathbf{G}^H + \mathbf{R}_x^{\frac{1}{2}}(\mathbf{I}-\mathbf{G}\mathbf{H})^H(\mathbf{I}-\mathbf{G}\mathbf{H})\mathbf{R}_x^{\frac{1}{2}} - (\rho a/N)\mathbf{I} - \rho\mathbf{R}_x^{\frac{1}{2}}\Delta\mathbf{H}^H\mathbf{B}\mathbf{R}_x^{-\frac{1}{2}} - \rho\mathbf{R}_x^{-\frac{1}{2}}\mathbf{B}^H\Delta\mathbf{H}\mathbf{R}_x^{\frac{1}{2}}\Big\} \le \mathbf{M} \quad (14)$$

where $\|\Delta\mathbf{H}\| \le \alpha$. Applying Lemma 2 from the Appendix for $\mathbf{G}\mathbf{R}_w\mathbf{G}^H$ in (14) yields (15), shown at the bottom of page. Applying Lemma 2 the second time to (15) yields (16), shown at the bottom of page. After straightforward algebra, (16) can be written as

$$\begin{bmatrix} \mathbf{M}+(\rho a/N)\mathbf{I} & \mathbf{G} & \left[(\mathbf{I}-\mathbf{G}\mathbf{H}_0)\mathbf{R}_x^{\frac{1}{2}}\right]^H \\ \mathbf{G}^H & \mathbf{R}_w^{-1} & \mathbf{0} \\ (\mathbf{I}-\mathbf{G}\mathbf{H}_0)\mathbf{R}_x^{\frac{1}{2}} & \mathbf{0} & \mathbf{I} \end{bmatrix} \ge \begin{bmatrix} -\rho\mathbf{R}_x^{-\frac{1}{2}}\mathbf{B}^H \\ \mathbf{0} \\ \mathbf{G} \end{bmatrix} \Delta\mathbf{H} \begin{bmatrix} \mathbf{R}_x^{\frac{1}{2}} & \mathbf{0} & \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{R}_x^{\frac{1}{2}} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} \Delta\mathbf{H}^H \begin{bmatrix} -\rho\mathbf{B}\mathbf{R}_x^{-\frac{1}{2}} & \mathbf{0} & \mathbf{G}^H \end{bmatrix}, \quad \|\Delta\mathbf{H}\| \le \alpha. \quad (17)$$

$$\min_{t,\mathbf{G},\lambda,\mathbf{M}} t$$

subject to

$$\mathrm{Tr}\{\mathbf{M}\} \le t \quad (10)$$

$$\begin{bmatrix} \mathbf{M}+(\rho a/N)\mathbf{I}-\lambda\mathbf{R}_x & \mathbf{G} & \left[(\mathbf{I}-\mathbf{G}\mathbf{H}_0)\mathbf{R}_x^{\frac{1}{2}}\right]^H & \alpha\rho\mathbf{R}_x^{-\frac{1}{2}}\mathbf{B}^H \\ \mathbf{G}^H & \mathbf{R}_w^{-1} & \mathbf{0} & \mathbf{0} \\ (\mathbf{I}-\mathbf{G}\mathbf{H}_0)\mathbf{R}_x^{\frac{1}{2}} & \mathbf{0} & \mathbf{I} & -\alpha\mathbf{G} \\ \alpha\rho\mathbf{B}\mathbf{R}_x^{-\frac{1}{2}} & \mathbf{0} & -\alpha\mathbf{G}^H & \lambda\mathbf{I} \end{bmatrix} \ge 0 \quad (11)$$

$$\begin{bmatrix} \mathbf{M}-\mathbf{R}_x^{\frac{1}{2}}(\mathbf{I}-\mathbf{G}\mathbf{H})^H(\mathbf{I}-\mathbf{G}\mathbf{H})\mathbf{R}_x^{\frac{1}{2}}+(\rho a/N)\mathbf{I}+\rho\mathbf{R}_x^{\frac{1}{2}}\Delta\mathbf{H}^H\mathbf{B}\mathbf{R}_x^{-\frac{1}{2}}+\mathbf{R}_x^{-\frac{1}{2}}\mathbf{B}^H\Delta\mathbf{H}\mathbf{R}_x^{\frac{1}{2}} & \mathbf{G} \\ \mathbf{G}^H & \mathbf{R}_w^{-1} \end{bmatrix} \ge 0 \quad (15)$$

$$\begin{bmatrix} \mathbf{M}+(\rho a/N)\mathbf{I}+\rho\mathbf{R}_x^{\frac{1}{2}}\Delta\mathbf{H}^H\mathbf{B}\mathbf{R}_x^{-\frac{1}{2}}+\rho\mathbf{R}_x^{-\frac{1}{2}}\mathbf{B}^H\Delta\mathbf{H}\mathbf{R}_x^{\frac{1}{2}} & \mathbf{G} & \left\{(\mathbf{I}-\mathbf{G}[\mathbf{H}_0+\Delta\mathbf{H}])\mathbf{R}_x^{\frac{1}{2}}\right\}^H \\ \mathbf{G}^H & \mathbf{R}_w^{-1} & \mathbf{0} \\ (\mathbf{I}-\mathbf{G}[\mathbf{H}_0+\Delta\mathbf{H}])\mathbf{R}_x^{\frac{1}{2}} & \mathbf{0} & \mathbf{I} \end{bmatrix} \ge 0 \quad (16)$$

Applying Lemma 3 from the Appendix to (17) yields the corresponding constraint (11) of Theorem 1. Combining constraints (17) and (13) yields the result in Theorem 1. This completes the proof of Theorem 1. □

B. Regret in Ratio Form

In this section, we introduce a regret formulation using ratios of MSEs. For an unknown linear model $\mathbf{H}$, we define our regret for committing to a particular linear estimate $\mathbf{G}$ as the ratio between the MSE of $\mathbf{G}$ and the scaled MMSE of the linear MMSE estimator $\mathbf{G}_{\mathrm{MMSE}}(\mathbf{H})$ tuned to $\mathbf{H}$ as

$$\frac{E\left[\|\mathbf{x}-\mathbf{G}(\mathbf{H}\mathbf{x}+\mathbf{w})\|^2\right]}{\rho J_{\mathrm{MMSE}}(\mathbf{H})}. \tag{18}$$

We then seek for the linear estimator that minimizes the worst case regret, i.e.,

$$\mathbf{G}^* = \arg\min_{\mathbf{G}} \max_{\mathbf{H}=\mathbf{H}_0+\Delta\mathbf{H}} \left\{\frac{E\left[\|\mathbf{x}-\mathbf{G}(\mathbf{H}\mathbf{x}+\mathbf{w})\|^2\right]}{\rho J_{\mathrm{MMSE}}(\mathbf{H})}\right\} \tag{19}$$

given the system estimate $\mathbf{H}_0$, where $\mathbf{H}=\mathbf{H}_0+\Delta\mathbf{H}$, with a norm constraint on $\Delta\mathbf{H}$ such that $\|\Delta\mathbf{H}\|\le\alpha$, $\alpha\in\mathbb{R}^+$. The parameter $\rho>0$ is arbitrary. We point out that, since ρ is a positive constant, it has no affect in the minimax formulation given in (19). However, ρ is included for notational consistency. Using the first-order linear approximation given in Lemma 1 for $J_{\mathrm{MMSE}}(\mathbf{H})$ without $o(\cdot)$ term and (1) for $E[\|\mathbf{x}-\hat{\mathbf{x}}\|^2]$ in (18) yield the regret formulation as

$$\begin{aligned}
&\frac{E\left[\|\mathbf{x}-\mathbf{G}(\mathbf{H}\mathbf{x}+\mathbf{w})\|^2\right]}{\rho J_{\mathrm{MMSE}}(\mathbf{H})}\\
&= \frac{\mathrm{Tr}[\mathbf{G}\mathbf{R}_w\mathbf{G}^H]+\mathrm{Tr}\left[\mathbf{R}_x^{\frac{1}{2}}(\mathbf{I}-\mathbf{G}\mathbf{H})^H(\mathbf{I}-\mathbf{G}\mathbf{H})\mathbf{R}_x^{\frac{1}{2}}\right]}{\rho a+2\rho\Re e\left\{\mathrm{Tr}[\Delta\mathbf{H}^H\mathbf{B}]\right\}}\\
&= \frac{\mathrm{Tr}[\mathbf{G}\mathbf{R}_w\mathbf{G}^H]+\mathrm{Tr}\left[\mathbf{R}_x^{\frac{1}{2}}(\mathbf{I}-\mathbf{G}\mathbf{H})^H(\mathbf{I}-\mathbf{G}\mathbf{H})\mathbf{R}_x^{\frac{1}{2}}\right]}{\rho a+2\rho\Re e\left\{\mathrm{Tr}\left[\mathbf{R}_x^{\frac{1}{2}}\Delta\mathbf{H}^H\mathbf{B}\mathbf{R}_x^{-\frac{1}{2}}\right]\right\}}.
\end{aligned}$$

The following theorem poses obtaining the linear estimator corresponding to this regret formulation as another SDP problem.

Theorem 2: Suppose an unknown desired data vector $\mathbf{x}$ with zero mean and known correlation matrix $\mathbf{R}_x>0$ is observed through an unknown linear system $\mathbf{H}$ as

$$\mathbf{y}=\mathbf{H}\mathbf{x}+\mathbf{w}$$

where $\mathbf{H}\in\mathbb{C}^{M\times N}$ and the corrupting noise vector $\mathbf{w}\in\mathbb{C}^M$, independent of $\mathbf{x}$, is zero mean with known correlation matrix $\mathbf{R}_w>0$. Given an estimate of $\mathbf{H}$ as $\mathbf{H}_0$, then the we have the problem (see (20), shown at the bottom of page) where $\mathbf{H}=\mathbf{H}_0+\Delta\mathbf{H}$ and $\|\Delta\mathbf{H}\|\le\alpha$, $\alpha,\rho>0$, is equivalent to the SDP problem

$$\min_{t,\mathbf{G},\lambda,\mathbf{M}} t$$

subject to

$$\mathrm{Tr}\{\mathbf{M}\}\le t\rho a \tag{21}$$

$$\begin{bmatrix}
\mathbf{M}-\lambda\mathbf{R}_x & \mathbf{G} & \left[(\mathbf{I}-\mathbf{G}\mathbf{H}_0)\mathbf{R}_x^{\frac{1}{2}}\right]^H & \alpha t\rho\mathbf{R}_x^{-\frac{1}{2}}\mathbf{B}^H\\
\mathbf{G}^H & \mathbf{R}_w^{-1} & \mathbf{0} & \mathbf{0}\\
(\mathbf{I}-\mathbf{G}\mathbf{H}_0)\mathbf{R}_x^{\frac{1}{2}} & \mathbf{0} & \mathbf{I} & -\alpha\mathbf{G}\\
\alpha t\rho\mathbf{B}\mathbf{R}_x^{-\frac{1}{2}} & \mathbf{0} & -\alpha\mathbf{G}^H & \lambda\mathbf{I}
\end{bmatrix}\ge 0 \tag{22}$$

where $\mathbf{M}=\mathbf{M}^H$, a and $\mathbf{B}$ are given in (28) and (29), respectively.

Proof of Theorem 2: We first observe that the problem in (18) is equivalent to the minimization problem shown at the bottom of page such that

$$\left\{\frac{\mathrm{Tr}[\mathbf{G}\mathbf{R}_w\mathbf{G}^H]+\mathrm{Tr}\left[\mathbf{R}_x^{\frac{1}{2}}(\mathbf{I}-\mathbf{G}\mathbf{H})^H(\mathbf{I}-\mathbf{G}\mathbf{H})\mathbf{R}_x^{\frac{1}{2}}\right]}{\rho a+2\rho\Re e\left\{\mathrm{Tr}\left[\mathbf{R}_x^{\frac{1}{2}}\Delta\mathbf{H}^H\mathbf{B}\mathbf{R}_x^{-\frac{1}{2}}\right]\right\}}\right\}\le t,\qquad \|\Delta\mathbf{H}\|\le\alpha \tag{23}$$

and $\mathbf{H}=\mathbf{H}_0+\Delta\mathbf{H}$. However, the constraint in (23) is equivalent to

$$\begin{aligned}
&\mathrm{Tr}[\mathbf{G}\mathbf{R}_w\mathbf{G}^H]+\mathrm{Tr}\left[\mathbf{R}_x^{\frac{1}{2}}(\mathbf{I}-\mathbf{G}\mathbf{H})^H(\mathbf{I}-\mathbf{G}\mathbf{H})\mathbf{R}_x^{\frac{1}{2}}\right]-\\
&\quad 2t\rho\Re e\left\{\mathrm{Tr}\left[\mathbf{R}_x^{\frac{1}{2}}\Delta\mathbf{H}^H\mathbf{B}\mathbf{R}_x^{-\frac{1}{2}}\right]\right\}\le t\rho a,\quad \|\Delta\mathbf{H}\|\le\alpha.
\end{aligned} \tag{24}$$

Defining an intermediate matrix $\mathbf{M}$, (24) can be equivalently written as

$$\mathrm{Tr}\{\mathbf{M}\}\le t\rho a \tag{25}$$

$$\begin{aligned}
&\Big\{\mathbf{G}\mathbf{R}_w\mathbf{G}^H+\mathbf{R}_x^{\frac{1}{2}}(\mathbf{I}-\mathbf{G}\mathbf{H})^H(\mathbf{I}-\mathbf{G}\mathbf{H})\mathbf{R}_x^{\frac{1}{2}}-\\
&\quad t\rho\mathbf{R}_x^{\frac{1}{2}}\Delta\mathbf{H}^H\mathbf{B}\mathbf{R}_x^{-\frac{1}{2}}-t\rho\mathbf{R}_x^{-\frac{1}{2}}\mathbf{B}^H\Delta\mathbf{H}\mathbf{R}_x^{\frac{1}{2}}\Big\}\le\mathbf{M},\\
&\quad \|\Delta\mathbf{H}\|\le\alpha.
\end{aligned} \tag{26}$$

We point out that (26) is in the same form as (14), hence we proceed following the same lines. We first apply Lemma 2, two times and then use the decomposition in (16) following Lemma 3 yielding

$$\begin{bmatrix}
\mathbf{M}-\lambda\mathbf{R}_x & \mathbf{G} & \left[(\mathbf{I}-\mathbf{G}\mathbf{H}_0)\mathbf{R}_x^{\frac{1}{2}}\right]^H & \alpha t\rho\mathbf{R}_x^{-\frac{1}{2}}\mathbf{B}^H\\
\mathbf{G}^H & \mathbf{R}_w^{-1} & \mathbf{0} & \mathbf{0}\\
(\mathbf{I}-\mathbf{G}\mathbf{H}_0)\mathbf{R}_x^{\frac{1}{2}} & \mathbf{0} & \mathbf{I} & -\alpha\mathbf{G}\\
\alpha t\rho\mathbf{B}\mathbf{R}_x^{-\frac{1}{2}} & \mathbf{0} & -\alpha\mathbf{G}^H & \lambda\mathbf{I}
\end{bmatrix}\ge 0 \tag{27}$$

which is the constraint (21). Combining (27) and (25) yields the constraints in Theorem 2. This completes the proof of Theorem 2. □

IV. Simulations

In this section, we demonstrate the performance of the introduced algorithms through numerical examples. In the first set of examples, linear

$$\underset{\mathbf{G}}{\text{minimize}}\ \underset{\mathbf{H}=\mathbf{H}_0+\Delta\mathbf{H}}{\text{maximize}}\ \frac{\mathrm{Tr}[\mathbf{G}\mathbf{R}_w\mathbf{G}^H]+\mathrm{Tr}\left[\mathbf{R}_x^{\frac{1}{2}}(\mathbf{I}-\mathbf{G}\mathbf{H})^H(\mathbf{I}-\mathbf{G}\mathbf{H})\mathbf{R}_x^{\frac{1}{2}}\right]}{\rho a+2\rho\Re e\left\{\mathrm{Tr}\left[\mathbf{R}_x^{\frac{1}{2}}\Delta\mathbf{H}^H\mathbf{B}\mathbf{R}_x^{-\frac{1}{2}}\right]\right\}} \tag{20}$$

$$\min_{\mathbf{G}}\max_{\|\Delta\mathbf{H}\|\le\alpha}\frac{\mathrm{Tr}[\mathbf{G}\mathbf{R}_w\mathbf{G}^H]+\mathrm{Tr}\left[\mathbf{R}_x^{\frac{1}{2}}(\mathbf{I}-\mathbf{G}\mathbf{H})^H(\mathbf{I}-\mathbf{G}\mathbf{H})\mathbf{R}_x^{\frac{1}{2}}\right]}{\rho a+2\rho\Re e\left\{\mathrm{Tr}\left[\mathbf{R}_x^{\frac{1}{2}}\Delta\mathbf{H}^H\mathbf{B}\mathbf{R}_x^{-\frac{1}{2}}\right]\right\}}=\min_{t,\mathbf{G}} t$$

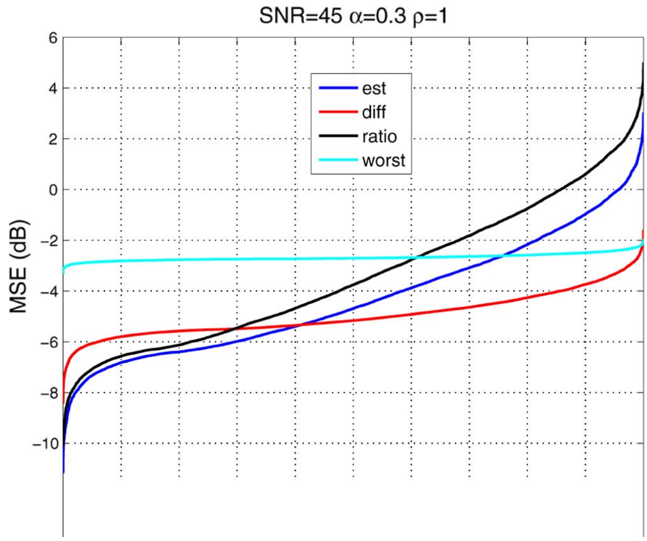

Fig. 1. Sorted MSEs for different $\mathbf{H}$, where $N = 4$, $M = 4$, $\mathrm{SNR} = 45$ (dB), $\rho = 1$, $\alpha = 0.3$. The algorithms are explained in the text along with the definitions of $\mathbf{R}_x$ and $\mathbf{R}_w$.

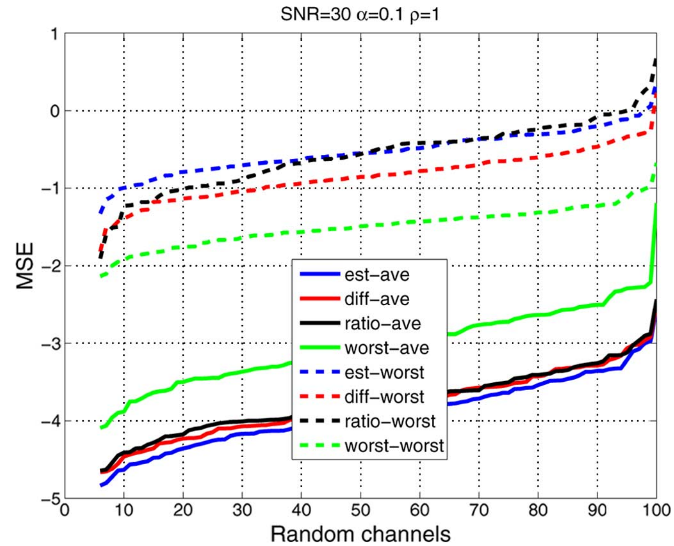

Fig. 2. Sorted average MSEs and worst case MSEs for different $\mathbf{H}_0$, where $N = 7$, $M = 7$, $\mathrm{SNR} = 30$ (dB), $\rho = 1$, $\alpha = 0.1$. The algorithms are explained in the text along with the definitions of $\mathbf{R}_x$ and $\mathbf{R}_w$.

models $\mathbf{H}$ with $N = 4$ and $M = 4$ are randomly generated, where for each such linear model, a random distortion with spectral norm less than $\alpha = 0.3$ is introduced to get $\mathbf{H}_0 = \mathbf{H} - \Delta\mathbf{H}$, i.e., $\|\Delta\mathbf{H}\| \leq 0.3$. For the transmitted data and noise processes, the correlation matrices are selected as $\mathbf{R}_x = \mathbf{I}$ and $\mathbf{R}_w = c\mathbf{I}$, where c is chosen to yield $\mathrm{SNR} = 45$ dB. In Fig. 1, we present results for the algorithm in Theorem 1 with $\rho = 1$ as "diff"; for the algorithm in Theorem 2 with $\rho = 1$ as "ratio"; for the linear MMSE estimator that is tuned to $\mathbf{H}_0$ as "est"; and finally for the minimax algorithm tuned to the worst possible model $\mathbf{H}$ in terms of MSE without the relative regret term that is introduced in the first part of [3] (which is equivalent to the algorithm introduced in Theorem 1 with $\rho = 0$) as "worst." Here, we randomly generated 10000 linear system models $\mathbf{H}$ and plot the corresponding MSEs sorted in ascended order in Fig. 1. The worst or the largest MSEs for $\alpha = 0.3$ and given the random $\mathbf{H}_0$ are: -1.2287 (dB) for the "worst" algorithm, 10.1999 (dB) for the "ratio" algorithm, 1.1661 (dB) for the "diff" algorithm and 7.7859 (dB) for the "est" algorithm. We observe that since the "worst" algorithm optimizes the MSE performance with respect to the worst possible model, it yields the smallest worst case MSE among all algorithms for these simulations. Nevertheless, due to this highly conservative design, the overall performance of the "worst" algorithm is significantly inferior to "est" and "diff" algorithms. We observe that although the "est" algorithm yields smaller average MSE, i.e., the area under the blue curve normalized with the number of trials. However, it also produces significantly larger worst case MSE than the "worst" algorithm. From Fig. 1, we observe that the "diff" algorithm provides superior average performance compared to the "worst" and "est" algorithms, and significantly superior worst case performance compared to the "est" algorithm for these simulations.

In the next set of experiments, we generate 100 random $\mathbf{H}_0$'s with $M = 7$, $N = 7$, where c is selected to yield $\mathrm{SNR} = 30$ (dB). For each linear model $\mathbf{H}_0$, we first calculate the corresponding worst case MSEs over the ball $\|\Delta\mathbf{H}\| \leq 0.1$ for each algorithm and plot the results in Fig. 2, as the contiguous lines. For each linear model, we also calculate the average MSEs over the ball $\|\Delta\mathbf{H}\| \leq 0.1$ and plot the results in Fig. 2, as the dashed lines. We observe that the "est" algorithm yields the largest worst case MSEs. For the "worst" algorithm, we have the largest average MSEs, however, the smallest worst case MSEs as expected. We observe from Fig. 2 that the regret formulations provide a fair trade off such that they provide good average MSEs with reduced worst case MSEs compared to "est" algorithm.

V. Conclusion

In this correspondence, a basic linear estimation problem is investigated in a competitive algorithm framework under MSE criteria. For this framework, two different regret formulations are studied that target to optimize the variation in the MSE performance relative to the linear MMSE estimator with the exact knowledge of the underlying linear system model. We investigated both the difference and the ratio as two alternative approaches leading to the corresponding alternative regret-based estimator formulations. We demonstrated that finding the linear estimators that minimize the worst case regret formulations can be cast as SDP problems, which can be efficiently solved using interior point methods. Numerical examples illustrate the potential merit for the proposed approaches, especially for the difference regret algorithm.

Appendix

Lemma 1: The first-order Taylor expansion of $J_{\mathrm{MMSE}}(\mathbf{H})$ around $\mathbf{H}_0$ is given by

$$\begin{aligned} J_{\mathrm{MMSE}}(\mathbf{H}) &= J_{\mathrm{MMSE}}(\mathbf{H}_0) \\ &\quad + 2\Re e\left\{\mathrm{Tr}\left[\nabla_{\mathbf{H}} J_{\mathrm{MMSE}}(\mathbf{H}_0)^H \Delta\mathbf{H}\right]\right\} + o\left(\|\Delta\mathbf{H}\|^2\right) \\ &= a + \mathrm{Tr}[\mathbf{B}^H\Delta\mathbf{H} + \Delta\mathbf{H}^H\mathbf{B}] + o\left(\|\Delta\mathbf{H}\|^2\right) \end{aligned}$$

where

$$a \triangleq J_{\mathrm{MMSE}}(\mathbf{H}_0) \tag{28}$$

$$\begin{aligned} \mathbf{B} &\triangleq \nabla_{\mathbf{H}} J_{\mathrm{MMSE}}(\mathbf{H}_0) \\ &= -\mathbf{R}_w^{-1}\mathbf{H}_0\left(\mathbf{R}_x^{-1} + \mathbf{H}_0^H\mathbf{R}_w^{-1}\mathbf{H}_0\right)^{-2}. \end{aligned} \tag{29}$$

Proof of Lemma 1: To get the first-order Taylor series expansion, we just need to derive the gradient of $J_{\mathrm{MMSE}}(\mathbf{H})$ with respect to $\mathbf{H}$

$$\begin{aligned} \nabla_{\mathbf{H}} J_{\mathrm{MMSE}}(\mathbf{H}) &= \nabla_{\mathbf{H}}\mathrm{Tr}\left[\left(\mathbf{R}_x^{-1} + \mathbf{H}^H\mathbf{R}_w^{-1}\mathbf{H}\right)^{-1}\right] \\ &= \nabla_{\mathbf{H}}\left[\sum_{i=1}^{M}\mathbf{e}_i^H\left(\mathbf{R}_x^{-1} + \mathbf{H}^H\mathbf{R}_w^{-1}\mathbf{H}\right)^{-1}\mathbf{e}_i\right] \end{aligned} \tag{30}$$

where $\mathbf{e}_i$ is the unit vector in the ith direction, i.e., all entries of $\mathbf{e}_i$ is zero, except the ith entry, which is equal to 1. Note that, the length of $\mathbf{e}_i$ is understood from the context. To get $\nabla_{\mathbf{H}}\mathbf{e}_i^H(\mathbf{R}_x^{-1} + \mathbf{H}^H\mathbf{R}_w^{-1}\mathbf{H})^{-1}\mathbf{e}_i$, we use that

$$\left(\mathbf{R}_x^{-1} + \mathbf{H}^H\mathbf{R}_w^{-1}\mathbf{H}\right)^{-1}\left(\mathbf{R}_x^{-1} + \mathbf{H}^H\mathbf{R}_w^{-1}\mathbf{H}\right) = \mathbf{I}. \tag{31}$$

Taking the derivative of (31) with respect to $H_{k,l}$, where $H_{k,l}$ is the entry of $\mathbf{H}$ located at row k and column l, based on the Wirtinger calculus [8], yields

$$\frac{\partial\left(\mathbf{R}_x^{-1}+\mathbf{H}^H\mathbf{R}_w^{-1}\mathbf{H}\right)^{-1}}{\partial H_{kl}}\left(\mathbf{R}_x^{-1}+\mathbf{H}^H\mathbf{R}_w^{-1}\mathbf{H}\right)+\left(\mathbf{R}_x^{-1}+\mathbf{H}^H\mathbf{R}_w^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^H\mathbf{R}_w^{-1}\mathbf{e}_k\mathbf{e}_l^T=\mathbf{0}. \quad (32)$$

Hence, after straightforward algebra, we have

$$\frac{\partial\left(\mathbf{R}_x^{-1}+\mathbf{H}^H\mathbf{R}_w^{-1}\mathbf{H}\right)^{-1}}{\partial H_{kl}}=-\left(\mathbf{R}_x^{-1}+\mathbf{H}^H\mathbf{R}_w^{-1}\mathbf{H}\right)^{-1}\times\mathbf{H}^H\mathbf{R}_w^{-1}\mathbf{e}_k\mathbf{e}_l^T\left(\mathbf{R}_x^{-1}+\mathbf{H}^H\mathbf{R}_w^{-1}\mathbf{H}\right)^{-1} \quad (33)$$

yielding

$$\begin{aligned}&\frac{\partial\mathbf{e}_i^T\left(\mathbf{R}_x^{-1}+\mathbf{H}^H\mathbf{R}_w^{-1}\mathbf{H}\right)^{-1}\mathbf{e}_i}{\partial H_{kl}}\\&=-\mathbf{e}_i^T\left(\mathbf{R}_x^{-1}+\mathbf{H}^H\mathbf{R}_w^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^H\mathbf{R}_w^{-1}\mathbf{e}_k\mathbf{e}_l^T\\&\quad\times\left(\mathbf{R}_x^{-1}+\mathbf{H}^H\mathbf{R}_w^{-1}\mathbf{H}\right)^{-1}\mathbf{e}_i,\\&=-\mathbf{e}_l^T\left(\mathbf{R}_x^{-1}+\mathbf{H}^H\mathbf{R}_w^{-1}\mathbf{H}\right)^{-1}\mathbf{e}_i\mathbf{e}_i^T\\&\quad\times\left(\mathbf{R}_x^{-1}+\mathbf{H}^H\mathbf{R}_w^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^H\mathbf{R}_w^{-1}\mathbf{e}_k. \qquad (34)\end{aligned}$$

Since by definition, $(\partial\mathbf{e}_i^T(\mathbf{R}_x^{-1}+\mathbf{H}^H\mathbf{R}_w^{-1}\mathbf{H})^{-1}\mathbf{e}_i)/\partial H_{kl}$ is the kth row and lth column entry of the matrix

$$-\left(\mathbf{R}_x^{-1}+\mathbf{H}^H\mathbf{R}_w^{-1}\mathbf{H}\right)^{-1}\mathbf{e}_i\mathbf{e}_i^T\left(\mathbf{R}_x^{-1}+\mathbf{H}^H\mathbf{R}_w^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^H\mathbf{R}_w^{-1}$$

then

$$\nabla_{\mathbf{H}}\mathbf{e}_i^H(\mathbf{R}_x^{-1}+\mathbf{H}^H\mathbf{R}_w^{-1}\mathbf{H})^{-1}\mathbf{e}_i=-\mathbf{R}_w^{-1}\mathbf{H}\times(\mathbf{R}_x^{-1}+\mathbf{H}^H\mathbf{R}_w^{-1}\mathbf{H})^{-1}\mathbf{e}_i\mathbf{e}_i^T(\mathbf{R}_x^{-1}+\mathbf{H}^H\mathbf{R}_w^{-1}\mathbf{H})^{-1}.$$

Using this in (30) yields

$$\begin{aligned}\nabla_{\mathbf{H}}J_{\mathrm{MMSE}}(\mathbf{H})&=-\sum_{i=1}^{M}\mathbf{R}_w^{-1}\mathbf{H}\left(\mathbf{R}_x^{-1}+\mathbf{H}^H\mathbf{R}_w^{-1}\mathbf{H}\right)^{-1}\\&\quad\times\mathbf{e}_i\mathbf{e}_i^T\left(\mathbf{R}_x^{-1}+\mathbf{H}^H\mathbf{R}_w^{-1}\mathbf{H}\right)^{-1}\\&=-\mathbf{R}_w^{-1}\mathbf{H}\left(\mathbf{R}_x^{-1}+\mathbf{H}^H\mathbf{R}_w^{-1}\mathbf{H}\right)^{-1}\mathbf{I}\\&\quad\times\left(\mathbf{R}_x^{-1}+\mathbf{H}^H\mathbf{R}_w^{-1}\mathbf{H}\right)^{-1}\\&=-\mathbf{R}_w^{-1}\mathbf{H}\left(\mathbf{R}_x^{-1}+\mathbf{H}^H\mathbf{R}_w^{-1}\mathbf{H}\right)^{-2}.\end{aligned}$$

This completes the proof of Lemma 1. □

Lemma 2 [6, Ch. 2]: The inequality

$$\begin{bmatrix}\mathbf{Q}&\mathbf{S}\\\mathbf{S}^H&\mathbf{R}\end{bmatrix}\geq 0 \quad (35)$$

where $\mathbf{Q}=\mathbf{Q}^H$, $\mathbf{R}=\mathbf{R}^H$, and $\mathbf{R}\geq 0$ is equivalent to

$$\mathbf{R}>0,\quad \mathbf{Q}-\mathbf{S}\mathbf{R}^{-1}\mathbf{S}^H\geq 0 \quad (36)$$

i.e., the set of nonlinear inequalities in (35) can be represented as (36).

Lemma 3 [3, Prop. 2]: Given matrices $\mathbf{P}$, $\mathbf{Q}$, and $\mathbf{A}$ with $\mathbf{A}=\mathbf{A}^H$

$$\mathbf{A}\geq\mathbf{P}^H\mathbf{Z}\mathbf{Q}+\mathbf{Q}^H\mathbf{Z}^H\mathbf{P},\quad\forall\|\mathbf{Z}\|\leq\alpha$$

if and only if there exists a $\lambda\geq 0$ such that

$$\begin{bmatrix}\mathbf{A}-\lambda\mathbf{Q}^H\mathbf{Q}&-\alpha\mathbf{P}^H\\-\alpha\mathbf{P}&\lambda\mathbf{I}\end{bmatrix}\geq 0.$$

A proof of Lemma 3 is given in [3].

References

[1] S. A. Kassam and H. V. Poor, "Robust signal processing for communication systems," *IEEE Commun. Mag.*, vol. 21, no. 1, pp. 20–28, Jan. 1983.

[2] A. T. Erdogan, B. Hassibi, and T. Kailath, "MIMO decision feedback equalization from an H^∞ perspective," *IEEE Trans. Signal Process.*, vol. 52, no. 3, pp. 734–745, Mar. 2004.

[3] Y. C. Eldar and N. Merhav, "A competitive minimax approach to robust estimation and random parameters," *IEEE Trans. Signal Process.*, vol. 52, no. 7, pp. 1931–1946, Jul. 2004.

[4] S. S. Kozat and A. C. Singer, "Universal switching linear least squares prediction," *IEEE Trans. Signal Process.*, vol. 56, no. 1, pp. 189–204, Jan. 2008.

[5] Y. C. Eldar, A. Ben-Tal, and A. Nemirovski, "Robust mean-squared error estimation in the presence of model uncertainties," *IEEE Trans. Signal Process.*, vol. 53, no. 1, pp. 168–181, Jan. 2005.

[6] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*, ser. Studies in Applied Mathematics. Philadelphia, PA: SIAM, 1994.

[7] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*. Upper Saddle River, NJ: Prentice-Hall, 2000.

[8] A. van den Bos, "Complex gradient and hessian," *Proc. IEE Vision Image Signal Process*, vol. 141, no. 6, pp. 380–383, Dec. 1994.

Least-Squares Design of DFT Filter-Banks Based on Allpass Transformation of Higher Order

Heinrich W. Löllmann and Peter Vary, *Fellow, IEEE*

***Abstract*—The allpass transformation of higher order is a very general concept to construct a frequency warped analysis–synthesis filter bank (AS FB) with nonuniform time-frequency resolution. In contrast to the more common allpass transformation of first order, the delay elements of the analysis filter bank are substituted by allpass filters of higher order to achieve a more flexible control over its frequency selectivity. Known analytical closed-form designs for the synthesis filter bank can ensure perfect reconstruction (PR), but the synthesis subband filters are not necessarily stable and exhibit no distinctive bandpass characteristic. These problems are addressed by a new least-squares error (LSE) filter bank design. The coefficients of the finite-impulse-response (FIR) synthesis filters are determined simply by a linear set of equations where the signal delay is an adjustable design parameter. This approach can achieve a perfect signal reconstruction with synthesis filters which are inherently stable and feature a bandpass characteristic. The proposed filter bank is of interest for various subband processing systems requiring nonuniform frequency bands.**

***Index Terms*—Allpass transformation, frequency warping, least squares, nonuniform filter banks, perfect reconstruction.**

I. Introduction

The allpass transformation is a common approach to design a filter bank with a nonuniform time-frequency resolution [1]–[3].

Manuscript received June 25, 2009; accepted December 03, 2009. First published January 08, 2010; current version published March 10, 2010. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Alper Tunga Erdogan.

The authors are with the Institute of Communication Systems and Data Processing (IND), RWTH Aachen University, 52056 Aachen, Germany, (e-mail: loellmann@ind.rwth-aachen.de; vary@ind.rwth-aachen.de.)

Digital Object Identifier 10.1109/TSP.2009.2039838