

A Counterexample in Distributed Optimal Sensing and Control

Serdar Yüksel, *Member, IEEE*, and Sekhar Tatikonda, *Member, IEEE*

Abstract—This note exhibits that in a distributed multi-sensor, single-controller scenario, for the minimization of a quadratic cost function, linear sensing policies over Gaussian channels are, in general, not optimal. This is in contrast with the corresponding single-sensor problem, which does admit an optimal linear solution.

Index Terms—Control over communications channels, decentralized control, linear time invariant (LTI).

I. INTRODUCTION

This note studies a distributed sensing problem. Consider a two-sensor, single controller, linear time invariant (LTI) system

$$\begin{aligned} s(t+1) &= as(t) + r(t) + d(t), \\ y_i(t) &= s(t) + v_i(t), \quad i = 1, 2. \end{aligned} \quad (1)$$

Here, $s(t) \in \mathbb{R}$ is the state of the system with the initial state $s(0)$ a zero-mean Gaussian random variable with variance σ_s^2 , $r(t) \in \mathbb{R}$ is the control signal, and $y_i(t) \in \mathbb{R}$ is the observation available at sensor station i at time t . Here $\{d(t), v_1(t), v_2(t)\}$ are i.i.d. zero-mean Gaussian disturbance processes with variances $\{\sigma_d^2, \sigma_{v_1}^2, \sigma_{v_2}^2\}$.

The sensors transmit their signals at time t , which we denote by $\{u_1(t), u_2(t)\}$, over two noisy Gaussian channels where the channel outputs

$$z_i(t) = u_i(t) + w_i(t), \quad i = 1, 2$$

are received by the controller. Here $w_i(t)$ are zero-mean Gaussian random variables with variances $\{\sigma_{w_i}^2, i = 1, 2\}$.

Upon observing the channel outputs, the controller generates its control $r(t)$.

In the following, we discuss the information structures under which the sensor signals $\{u_1(t), u_2(t)\}$ and the control signal $r(t)$ are generated. A pictorial description for the sequence of events for a single-stage is presented in Fig. 1.

Before proceeding further, we provide a discussion on the notation of this note. Bold-face letters denote the ensemble of vectors such that $\mathbf{y}(t) = \{y_1(t), y_2(t)\}$. The notation $y([i, j])$ represents the sequence $\{y(k), i \leq k \leq j\}$. $E[\cdot]$ denotes the expectation operator, that is $E[x] = \int x\mu(dx)$, with $\mu(\cdot)$ denoting the distribution of the random variable X . When the distribution under which the expectation is performed is not clear, we write $E_x[x] = \int x\mu(dx)$, denoting explicitly that $\mu(dx)$ corresponds to the random variable X . $Pr(\cdot)$ denotes the probability of an event in the Borel sigma-algebra generated by the

product Euclidean spaces for the random variables in (1) over a finite horizon.

We now describe the information structure:

One-Step Delayed Information Structure IS: The sensor action $u_i(t)$ ($i = 1, 2$) is the output of a mapping $\gamma_i(\cdot)$ measurable with respect to the sigma-field generated by

$$I_i^s(t) = \{y_i(t), Z(t)\}$$

and is a mapping to \mathbb{R} . Here

$$Z(t) = \{\mathbf{y}([0, t-1]), \mathbf{z}([0, t-1])\}$$

is the common past information, where $z_1(t), z_2(t)$ are the information received by the controller. The controller policy $\psi(\cdot)$ is measurable with respect to the sigma-algebra generated by

$$I^c(t) = \{\mathbf{z}([0, t-1])\}$$

and is a mapping to \mathbb{R} .

We also have a power constraint on the sensors, such that

$$E[u_i(t)^2] \leq P_i, \quad i = 1, 2.$$

We now state the objective functional and formalize the problem statement:

Problem P1: The objective is to minimize the following cost function:

$$J(\gamma_1, \gamma_2, \psi) = E \left[\sum_{t=0}^T s^2(t) + qr^2(t) \right] \quad (2)$$

for some $q > 0$, over all admissible sensor and control policies under the information structure *IS* presented above. \diamond

We now revisit a result due to Bansal and Başar [1]:

Theorem 1.1: Consider Fig. 1, with only one sensor, one channel and with $v_1 = 0$. Then the optimal sensor policy is to apply

$$u(t) = \alpha(t)(s(t) - E[s(t)|\mathbf{z}([0, t-1])])$$

where $\alpha(t) = \sqrt{P/E[(s(t) - E[s(t)|\mathbf{z}([0, t-1])])^2]}$, with P denoting the power constraint on the sensor signal.

We note that, for a one-stage problem with $T = 1$, the optimality of a linear sensing policy still holds even when $v_1 \neq 0$. This was also shown by Bansal and Başar in another publication [2].

The following is the result of this note.

Theorem 1.2: For the multi-sensor setting, there exist non-linear sensing schemes which outperform linear sensing policies under the Information Structure *IS*, such that a lower cost value is attained for Problem P1.

II. PROOF OF THEOREM 1.2

In the following we provide an example where this is satisfied. We take $T = 1$, and seek to minimize $E[\sum_{t=0}^1 s^2(t) + qr^2(t)]$, under the aforementioned information structure.

A. Optimal Controller

The cost can be written as

$$E[qr^2(1) + (as(0) + r(0) + d(0))^2 + qr^2(0) + s^2(0)].$$

Manuscript received May 11, 2008; revised September 19, 2008 and October 20, 2008. Current version published April 08, 2009. This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) and in part by the National Science Foundation under Grant ECCS-0547199. This technical note appeared in part at the Annual Allerton Conference on Communication, Control, and Computing, 2007. Recommended by Associate Editor I. Paschalidis.

S. Yüksel is with the Department of Mathematics and Statistics, Queen's University, Kingston, ON K7L 3N6, Canada (e-mail: yuksel@mast.queensu.ca).

S. Tatikonda is with the Department of Electrical Engineering, Yale University, New Haven, CT 06511 USA (e-mail: sekhar.tatikonda@yale.edu).

Digital Object Identifier 10.1109/TAC.2008.2009680

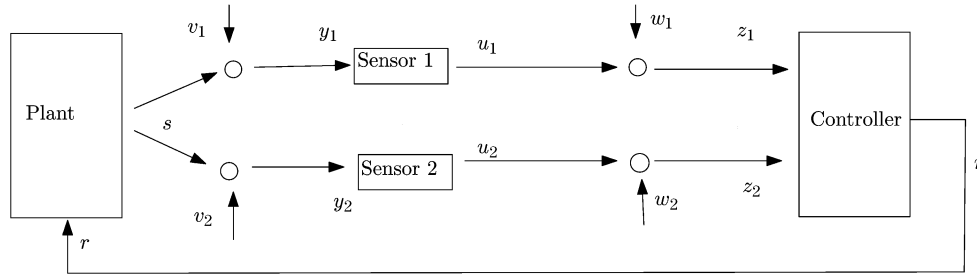


Fig. 1. Controller has access to (z_1, z_2) .

Using the completion of squares method [2], we obtain

$$J(\gamma_1, \gamma_2, \psi) = E \left[qr^2(1) + \left(\frac{a}{\sqrt{1+q}}s(0) + \sqrt{1+q}r(0) \right)^2 + \left(a^2 \left(1 - \frac{1}{1+q} \right) + 1 \right) s^2(0) + d^2(0) \right]. \quad (3)$$

Observing the fact that the estimation error is orthogonal to the best estimate at the controller, the optimal control at time $t = 0$ can be evaluated as

$$r(0) = -\frac{a}{1+q} E[s(0)|z_1(0), z_2(0)]$$

with

$$r(1) = 0.$$

Hence, the total cost for $T = 1$ can be written as

$$J(\gamma_1, \gamma_2, \psi) = \frac{a^2}{1+q} E_{\mathbf{z}(0)} [(s(0) - E[s(0)|\mathbf{z}(0)])^2] + E[s^2(0)] \left(a^2 \left(1 - \frac{1}{1+q} \right) + 1 \right) + E[d^2(0)]. \quad (4)$$

As such, the remaining issue becomes the minimization of the estimation error variance $E_{\mathbf{z}(0)} [(s(0) - E[s(0)|\mathbf{z}(0)])^2]$.

In the following, we present the performance of the best linear sensing policies, and introduce an alternative sensing scheme. We conclude the proof by a comparison of the performances under the two schemes.

B. Performance of the Best Linear Sensing Policies

Under linear policies, if one writes the joint channel from the input to the output, the controller, at time 0, has access to (we drop the time-index 0 to make the notation easier to present)

$$z_1 = \alpha_1 s + \alpha_1 v_1 + w_1$$

and

$$z_2 = \alpha_2 s + \alpha_2 v_2 + w_2.$$

The observed variables can be scaled (this does not affect the information content of the received data, since there is only a scaling involved) so as to obtain induced channels

$$\tilde{z}_1 := z_1/\alpha_1 = s + v_1 + w_1/\alpha_1$$

$$\tilde{z}_2 := z_2/\alpha_2 = s + v_2 + w_2/\alpha_2.$$

Here

$$\alpha_1^2 = \frac{P_1}{E[s^2] + \sigma_{v_1}^2}, \quad \alpha_2^2 = \frac{P_2}{E[s^2] + \sigma_{v_2}^2}.$$

The optimal Minimum Mean Square Estimate Decoder policy is linear [4], with the output \hat{s}

$$\hat{s} = \beta_1 \tilde{z}_1 + \beta_2 \tilde{z}_2$$

with the coefficients β_1, β_2 satisfying

$$E[s\tilde{z}_1] = \beta_1 E[\tilde{z}_1^2] + \beta_2 E[\tilde{z}_2\tilde{z}_1]$$

$$E[s\tilde{z}_2] = \beta_2 E[\tilde{z}_2^2] + \beta_1 E[\tilde{z}_2\tilde{z}_1].$$

The performance of the optimal linear coding and decoding can be computed from the performance of a Minimum Mean Square Estimate Decoder ([4]): In particular, the estimation error reduces to

$$E[s^2] - [E[s^2]E[s^2]] A^{-1} [E[s^2]E[s^2]]^T \quad (5)$$

where

$$A = \begin{bmatrix} E[s^2] + \sigma_{w_1'}^2 & E[s^2] \\ E[s^2] & E[s^2] + \sigma_{w_2'}^2 \end{bmatrix}$$

with the induced channel noise variances

$$\sigma_{w_1'}^2 = \sigma_{v_1}^2 + \frac{\sigma_{w_1}^2}{\frac{P_1}{E[s^2] + \sigma_{v_1}^2}}$$

$$\sigma_{w_2'}^2 = \sigma_{v_2}^2 + \frac{\sigma_{w_2}^2}{\frac{P_2}{E[s^2] + \sigma_{v_2}^2}}.$$

We note that we took the powers to be equal to P_1 and P_2 respectively, as the value in (5)

$$E[s^2] \left(1 - \frac{1}{1 + \frac{\sigma_{w_1}^2 \sigma_{w_2}^2}{E[s^2] (\sigma_{w_1}^2 + \sigma_{w_2}^2)}} \right)$$

is strictly increasing in the induced channel noise variances and lower power pairs would have lead to a higher estimation error variance.

C. Alternative Sensing Policy

We now provide an alternative sensing scheme, which we will show to perform better than the optimal linear policy. Toward obtaining the sensing scheme, we first revisit a relevant result from the information theory literature to guide the construction of the alternative coding scheme.

For a distributed joint source-channel code minimizing a mean-square distortion to be optimal, the following two conditions are sufficient (for example see [6]):

- 1) All channels send independent information.
- 2) All channels utilize the capacity (source-channel needs to be matched in that the rate-distortion achieving test channel is the

channel itself and the the source maximizing the capacity is the distribution of the source itself).

In the linear case discussed in the previous subsection, the transmitted signals are inevitably correlated.

Accordingly, the encoding policy will try to send as useful information as possible: The information to be sent needs to be non-redundant; and hence independent. In the scheme that will be provided, one of the coders will transmit the magnitude of the signals she received at time 0 (we drop the time-index hereafter to make the notation easier to follow), and the other will transmit the sign of the signal.

Lemma 2.1: Let $1_{(\cdot)}$ be the indicator function. The random variables $1_{(s+v_1>0)}$ and $|s+v_2|$ are independent.

Proof: Let $\kappa \geq 0$ and let $Pr(\cdot)$ be replaced by $P(\cdot)$ in the following. It follows by the fact that a Gaussian random variable has a symmetric density function:

$$\begin{aligned}
 P(|s+v_2| = \kappa | 1_{(s+v_1>0)} = 1) &= \frac{P(s = \kappa - v_2, s+v_1 > 0)}{P(s+v_1 > 0)} \\
 &+ \frac{P(s = -\kappa - v_2, s+v_1 > 0)}{P(s+v_1 > 0)} \\
 &= 2 \{P(s = \kappa - v_2, v_1 > -s) \\
 &\quad + P(s = -\kappa - v_2, v_1 > -s)\} \\
 &= 2 \{P(s = \kappa - v_2, v_1 > -s, v_2 \geq 0) \\
 &\quad + P(s = -\kappa - v_2, v_1 > -s, v_2 \geq 0) \\
 &\quad + P(s = \kappa - v_2, v_1 > -s, v_2 < 0) \\
 &\quad + P(s = -\kappa - v_2, v_1 > -s, v_2 < 0)\} \\
 &= 2 \{P(s = \kappa - v_2, v_1 > -s, v_2 \geq 0) \\
 &\quad + P(s = -\kappa - v_2, v_1 > -s, v_2 \geq 0) \\
 &\quad + P(s = \kappa - v_2, v_1 > s, v_2 \geq 0)\} \\
 &= 2 \{P(s = \kappa - v_2, v_2 \geq 0) + P(s = -\kappa - v_2, v_2 \geq 0)\} \\
 &= \{P(s = \kappa - v_2, v_2 \geq 0) + P(s = -\kappa - v_2, v_2 < 0) \\
 &\quad + P(s = \kappa - v_2, v_2 < 0) + P(s = -\kappa - v_2, v_2 \geq 0)\} \\
 &= \{P(s = \kappa - v_2) + P(s = -\kappa - v_2)\} \\
 &= P(|s+v_2| = \kappa) \tag{6}
 \end{aligned}$$

Hence, the random variables are independent. \diamond

As such, the messages carried by the two channels are independent; even though they do not satisfy the matching conditions with the channel (only a Gaussian source is matched to a Gaussian channel). Hence the approach is to express the signal estimate as

$$\tilde{u}_2 u_1$$

with \tilde{u}_2 denoting the magnitude information and u_1 denoting the sign information on the random variable s . To minimize the power of the transmitted signal at the sensor, we write $\tilde{u}_2 = u_2 + \eta$, where η is $E_{s,v_1}[|E[s|s+v_1]|]$ and only transmit the information on u_2 . As such, the decoder policy $\zeta(z_1, z_2) : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ will be as follows:

$$\hat{s}(z_1, z_2) = \zeta(z_1, z_2) = (\tilde{u}_2(z_2) + \eta) \hat{u}_1(z_1)$$

with $\hat{u}_2(z_2)$ being the best linear decoder estimate of the shifted magnitude of the source, and $\hat{u}_1(z_1)$ the information regarding the sign of the source.

We can write the estimation error variance as

$$\begin{aligned}
 E[(s - \hat{s})^2] &= Pr(u_1 \neq \hat{u}_1) E[(u_2 - \hat{u}_2)^2 | u_1 \neq \hat{u}_1] \\
 &\quad + Pr(u_1 = \hat{u}_1) E[(u_2 - \hat{u}_2)^2 | u_1 = \hat{u}_1]. \tag{7}
 \end{aligned}$$

We note that, because of the symmetry in the coding, it does not matter whether the sent sign is positive or negative.

One of the sensors will transmit the magnitude of the signal by scaling it with a coefficient k

$$u_2 = k (|E[s|s+v_1]| - E_{s,v_1}[|E[s|s+v_1]|])$$

which reduces to

$$\begin{aligned}
 u_2 = k \left(\left(\frac{E[s^2]}{E[s^2] + E[v_1^2]} (s+v_1) \right) \right. \\
 \left. - E_{s,v_1} \left[\left(\frac{E[s^2]}{E[s^2] + E[v_1^2]} (s+v_1) \right) \right] \right) \tag{8}
 \end{aligned}$$

or

$$u_2 = k' (|(s+v_1)| - E[|(s+v_1)|])$$

where

$$k' = \frac{P_2}{E[|(s+v_1)| - E[|(s+v_1)|]]^2}$$

Finally, upon the linear decoding operation at the decoder, the mean-square estimation error with regard to u_2 can be written as (with $E[u_2] = \hat{u}_2$)

$$E[(u_2 - E[u_2])^2] = E[u_2^2] - \frac{E[u_2^2] E[u_2^2]}{E[u_2^2] + E[w_2^2] / \frac{P_2}{E[u_2^2]}} \tag{9}$$

We now consider the transmission of the sign of the expected signal, which is transmitted in the form of u_1 . The transmission for u_1 is a binary signal with opposite amplitudes. The maximum likelihood error is given by:

$$Pr(w_1 > \sqrt{P_1})$$

which is identical to one half of Complementary Error function for Gaussian random variables

$$\frac{1}{2} Q \left(\sqrt{\frac{P_1}{2\sigma_{w_1}^2}} \right) := \int_{\sqrt{\frac{P_1}{2\sigma_{w_1}^2}}}^{\infty} \frac{1}{\sqrt{\pi}} e^{-\xi^2} d\xi \tag{10}$$

However, we should observe that, the transmitted sign might not be the actual sign of s . For this we also need that

$$\text{sign}(s+v_1) = \text{sign}(s).$$

In the following, the error is further upper bounded by including errors in the sign of the signal. Let $p_e(s_i)$ denote the error in the expectation of the sign information such that, $p_e(s_i) := 1 - Pr(\text{sign}(s+v_1) = \text{sign}(s))$, and let $p'_e(s_i)$ denote the error probability in the transmission of the sign information given by (10): The overall estimation error can be upper bounded by

$$\begin{aligned}
 (1 - p_e(s_i)) (1 - p'_e(s_i)) \{E[(u_2 - E[u_2])^2]\} \\
 + \{1 - (1 - p_e(s_i)) (1 - p'_e(s_i))\} \\
 \times \{E[(u_2 - E[u_2])^2] + 4\eta^2\}. \tag{11}
 \end{aligned}$$

We first compute

$$E[(u_2 - E[u_2])^2] = E[u_2^2] - \frac{E[u_2^2] E[u_2^2]}{E[u_2^2] + E[w_2^2] / \frac{P_2}{E[u_2^2]}}$$

as a function of $E[s^2]$. We have that, for a Gaussian random variable x

$$E[|(x| - E[|x|])^2] = E[x^2] - E[x^2] \frac{2}{\pi} = E[x^2] \left(1 - \frac{2}{\pi}\right)$$

which follows from the result that for a Gaussian distribution:

$$\int_{\mathbb{R}} |x| \mu(dx) = \sqrt{\frac{E[x^2]}{\pi}}.$$

Hence, $E[(u_2 - E[u_2])^2]$ becomes

$$E[v^2] \left(1 - \frac{2}{\pi}\right) \left(1 - \frac{E[v^2] \left(1 - \frac{2}{\pi}\right)}{E[v^2] \left(1 - \frac{2}{\pi}\right) + E[w_2^2] / \frac{P_2}{E[v^2] \left(1 - \frac{2}{\pi}\right)}}\right)$$

with $E[v^2] = E[s^2] + E[v_2^2]$.

The error in the sign information is equivalent to $Pr(v_1 \geq -s)$ if $s < 0$ and $Pr(v_1 \leq -s)$ if $s \geq 0$, which is equal to

$$p_e(s) = 2 \int_{s=0}^{\infty} \frac{1}{\sqrt{2\pi\sigma_s^2}} \frac{1}{2} Q\left(\frac{s}{\sqrt{2}\sigma_{v_1}}\right) e^{-\frac{s^2}{2\sigma_s^2}} ds.$$

Hence, we have the upper bound on the estimation error variance as

$$\begin{aligned} & (1 - p_e(s)) \left(1 - \frac{1}{2} Q\left(\sqrt{\frac{P_1}{2\sigma_{w_1}^2}}\right)\right) \{E[(u_2 - E[u_2])^2]\} \\ & + \left(1 - (1 - p_e(s)) \left(1 - \frac{1}{2} Q\left(\sqrt{\frac{P_1}{2\sigma_{w_1}^2}}\right)\right)\right) \\ & \cdot (E[(u_2 - E[u_2])^2] + 4\eta^2) \end{aligned} \quad (12)$$

with $E[(u_2 - E[u_2])^2]$

$$E[v^2] \left(1 - \frac{2}{\pi}\right) \left(1 - \frac{E[v^2] \left(1 - \frac{2}{\pi}\right)}{E[v^2] \left(1 - \frac{2}{\pi}\right) + E[w_2^2] / \frac{P_2}{E[v^2] \left(1 - \frac{2}{\pi}\right)}}\right) \quad (13)$$

and

$$E[v^2] = E[s^2] + E[v_2^2].$$

D. Comparison

We now compare (5) with (12) in the overall cost (4).

Let us pick the values as $a = 1.2$, $q = 0.005$, $P_1 = 10$, $P_2 = 40$, $E[s^2] = 5$, $E[v_1^2] = 0$, $E[v_2^2] = 0$, $E[w_2^2] = E[w_1^2] = 2$, $E[d^2] = 0.2$.

One checks that with these values an upper bound on the cost $J(\cdot)$ with the alternative coding scheme is 5.36, whereas the optimal linear scheme leads to a cost of 5.51.

Hence, there exists a scheme in which linear policies are outperformed.

III. DISCUSSION

For a scalar, single-sensor controller system, even when the information structure is not partially nested [5] (the sensor action affects the controller, but the controller does not have access to the sensor information), the optimal sensing policy for the system and cost studied had been shown to be linear in [2].

This note exhibited that the optimality of linear policies does not extend to the multi-sensor case.

The optimality of linear policies occur infrequently in distributed quadratic optimization problems: On one hand, we have the discussion regarding information structures (nestedness and weakened versions of nestedness), on another hand we require the channel noise to be

Gaussian, the source to be Gaussian, and the cost to be quadratic. The reason in our case linear policies are suboptimal is primarily because the Gaussian-source channel matching does not occur here: the optimal source distribution for a parallel Gaussian channel is not realizable by a real-time code which encodes a single realization of a Gaussian random variable. If it were possible to extract two Gaussian random variables $a(t)$, $b(t)$ deterministically such that, $s(t) = a(t) + b(t)$ and $E[a(t)b(t)] = 0$, and $(a(t), b(t))$ jointly Gaussian, linear policies could have been optimal. However, such a scheme is not possible with a linear transformation. This is possible, however, in an infinite-dimensional, block-version of the problem [7]. This is also possible if the size of the space the source lives in matches to the dimensionality of the channels by applying a linear transformation leading to independent variables to be transmitted, for example if it were possible to send two uncorrelated Gaussian signals of each dimension one through each of the Gaussian channels.

In the literature, discussions regarding a lack of incentive for signaling (that is the controllers communicating via feedback) and nestedness conditions are presented as sufficient conditions leading to optimality of linear policies for LQG team problems [3]. The notion of partial-nestedness requires further refinement since there are problems which are not partially nested, but solvable when there is a clear distinction between the incentives on communications and control [1]. Furthermore, the objective in the system optimization is also important in the tractability of the optimization problems as discussed in [8]. Current efforts are in the direction towards understanding these issues.

REFERENCES

- [1] R. Bansal and T. Başar, "Simultaneous design of measurement and control strategies for stochastic systems with feedback," *Automatica*, vol. 25, pp. 679–694, Sep. 1989.
- [2] R. Bansal and T. Başar, "Solutions to a class of linear-quadratic-Gaussian LQG stochastic team problems with nonclassical information," *Syst. Control Lett.*, vol. 9, pp. 125–130, 1987.
- [3] A. Rantzer, "Linear quadratic team theory revisited," in *Proc. IEEE Amer. Control Conf.*, Minneapolis, MN, Jun. 2006, pp. 1637–1641.
- [4] H. V. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag, 1988.
- [5] Y. C. Ho and K. C. Chu, "Team decision theory and information structures in optimal control problems—Part I," *IEEE Trans. Automat. Control*, vol. AC-17, no. 1, pp. 15–22, Feb. 1972.
- [6] S. Shamai, S. Verdú, and R. Zamir, "Systematic lossy source/channel coding," *IEEE Trans. Inform. Theory*, vol. 44, no. 2, pp. 564–579, Mar. 1998.
- [7] A. B. Wagner, S. Tavildar, and P. Viswanath, "Rate region of the quadratic Gaussian two-encoder source-coding problem," *IEEE Trans. Inform. Theory*, vol. 54, no. 5, pp. 1938–1961, May 2008.
- [8] M. Rotkowitz, "Linear controllers are uniformly optimal for the Witsenhausen counterexample," in *Proc. IEEE Conf. Decision Control*, San Diego, CA, Dec. 2006, pp. 553–558.