Distributed signal cancelation inspired by Witsenhausen's counterexample

Pulkit Grover and Anant Sahai
Wireless Foundations, Department of EECS
University of California at Berkeley, CA-94720, USA
{pulkit, sahai} @eecs.berkeley.edu

Abstract

We consider the problem of two-stage signal cancelation based on noisy observations. This problem turns out to be an extension of the Witsenhausen counterexample — a famous open problem in distributed control. Cost is imposed on the power expended by the first controller, and the residual signal after the actions of the two controllers. Along the lines of a recent approximate solution to the Witsenhausen counterexample, we provide an approximate solution to this distributed signal cancelation problem to within a constant factor. This approximation holds uniformly over all problem parameters and for all vector lengths.

I. Introduction

Consider the distributed signal canceling system shown in Fig. 1. The first controller (encoder $\mathcal{E}$) observes a noisy version of the signal $\mathbf{S}^m$ and modifies it by adding a power-constrained input $\mathbf{U}^m$. The resulting signal $\mathbf{X}^m$ is observed noisily by a second controller (decoder $\mathcal{D}$). Based on this observation, the decoder subtracts an input $\widehat{\mathbf{X}}^m$ from $\mathbf{X}^m$. The goal is to reduce the mean square value of the resulting signal $\mathbf{X}^m - \widehat{\mathbf{X}}^m$. Intuitively, it might appear that the optimal strategy for each controller is to scale down the signal as much as possible. After all, the system is a Linear-Quadratic-Gaussian (LQG) system and it is well known that for centralized LQG systems, control laws linear in the observation are optimal [1]. Even within information theory, when the encoder has noiseless observations, if the objective is to communicate $\mathbf{S}^m$ to the decoder, then the optimal strategy is well known to be linear [2]. The related problem of state masking, where the transmitter wants to hide $\mathbf{S}^m$ from the decoder (in a mean-square error sense), also has a linear solution [3].

It may come as a surprise, therefore, that the optimal strategy for the problem of signal cancelation is nonlinear. This is because the problem is a generalization[1] of the infamous Witsenhausen counterexample [6], a *distributed* control problem for which it is known that the optimal strategy is nonlinear [6]. Witsenhausen's counterexample is still unsolved in that it is unknown what the optimal strategy or the optimal costs are. The problem formulation is quite similar to some problems in information theory. For example, in the problem of remote source coding [7], the encoders have noisy observations of the source that needs to be communicated to the decoder. The encoder can also be thought of as an agent that is relaying the source, as considered in [8], [9]. The main difference in Witsenhausen's formulation (as well as our extension here) as compared to communication problems is that it

[1] In the limit of zero observation noise at the encoder, the signal cancelation problem reduces to a vector extension of the counterexample [4]. To the best of our knowledge, the connection of Witsenhausen's counterexample with signal cancelation was first noticed by Martins [5].

is the *modified state* $\mathbf{X}^m$ (and not a message, or the source $\mathbf{S}^m$) that is to be reconstructed at the decoder. This feels a bit unnatural from a traditional information-theoretic standpoint, because it amounts to modification of the information that is meant to be communicated. However, the problem formulation is perfectly natural in a control setting because cancelation and tracking are problems of wide applicability in control systems.

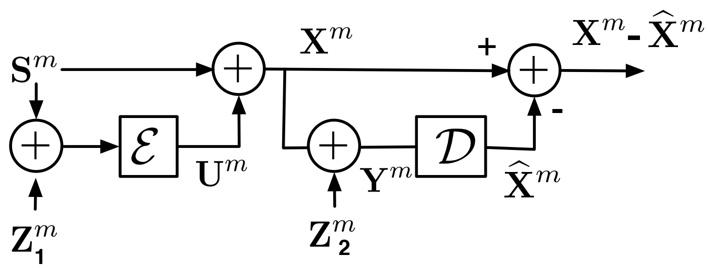

Fig. 1. The model for a noise canceling system. This model is an extension of Witsenhausen's counterexample, where the noise $\mathbf{Z}_1^m$ is not present, and $\mathbf{S}^m$ is observed perfectly by the encoder.

Recently, we formulated a vector version of the Witsenhausen counterexample, which allows the application of laws of large numbers, thus simplifying the problem. For this simplified problem, we obtained information-theoretic upper and lower bounds on the cost[2], characterizing the asymptotic (infinite-length) optimal costs to within a factor of 1.3 (calculated numerically) for all problem parameters [4], [10]. Based on this work, we then obtained similar approximate-optimality results for the finite-length extensions using lattice-based techniques and sphere-packing outer bounds [11], [12]. For example, for the original counterexample, which corresponds to the scalar case, we characterize the optimal cost to within a factor of 8 (calculated numerically) for all problem parameters.

Building on our work on the counterexample, this paper provides approximately optimal solutions to the problem of distributed signal cancelation. In Section II, we consider the problem of Fig. 1 where the difference from the counterexample is that the observation of the encoder is noisy. In Section III, we provide an equivalent problem, where there is no noise in the observation at the encoder, but there is noise in evolution of the state $\mathbf{X}^m$. Using this equivalent problem, in Section IV, we characterize the asymptotic optimal costs for the signal cancelation problem to within a factor of 80 for all problem parameters (numerical evaluation of the bounds shows that the actual factor is smaller than 10). Using more sophisticated techniques, we then derive results for finite vector-lengths, characterizing the optimal costs to within a constant factor (uniformly over all problem parameters) for any vector length. In Section V, we observe that our techniques also provide approximately optimal solutions to the problem with noises in all inputs, state evolutions, and observations. This compliments our earlier work [13] where we provided approximately optimal solutions to the extension of the counterexample with quadratic costs on all states and inputs, and suggests that the set of tools developed in this line of work are rich enough to begin addressing more sophisticated distributed control problems.

This line of work parallels (and is intimately connected to) the recent work in information theory where advances have been made on long-standing problems using similar approximation approaches. Because of space limitations,

[2]As is usual in control, the cost is defined as a weighted sum of the power and distortion costs.

we refer the reader to [4], [12] for survey of related results in information theory (including connections with deterministic models [14] and constant gap results for capacity e.g. [15]).

II. Problem statement, definitions, and notation

A. The signal cancelation problem

The "initial state" $\mathbf{S}^m$ is distributed $\mathcal{N}(0,\sigma^2\mathbb{I})$. The encoder $\mathcal{E}$ observes $\mathbf{S}^m+\mathbf{Z}_1^m$, where $\mathbf{Z}_1^m\sim\mathcal{N}(0,N_1)$ is independent of $\mathbf{S}^m$. Based on this observation, the encoder modifies state $\mathbf{S}^m$ using an input $\mathbf{U}^m$ of average power at most P, resulting in a state $\mathbf{X}^m$, i.e. $\mathbb{E}\left[\|\mathbf{S}^m-\mathbf{X}^m\|^2\right]\leq mP$. The decoder $\mathcal{D}$ observes the state $\mathbf{X}^m$ through a noisy channel with additive white Gaussian noise $\mathbf{Z}_2^m\sim\mathcal{N}(0,N_2\mathbb{I})$, which is independent of $\mathbf{X}^m$. Without loss of generality, we assume that $N_2=1$. The decoder maps its observation $\mathbf{Y}^m=\mathbf{X}^m+\mathbf{Z}_2^m$ to an estimate $\widehat{\mathbf{X}}^m$ of the modified state $\mathbf{X}^m$. The objective is to minimize the MMSE error $\mathbb{E}\left[\|\mathbf{X}^m-\widehat{\mathbf{X}}^m\|^2\right]$.

Alternatively, the control-theoretic weighted cost formulation [6] defines the total cost to be

$$J=\frac{1}{m}k^2\|\mathbf{U}^m\|^2+\frac{1}{m}\|\mathbf{X}^m-\widehat{\mathbf{X}}^m\|^2, \tag{1}$$

where $k\in\mathbb{R}^+$. The objective is to minimize the average cost, $\mathbb{E}[J]$, in an unconstrained manner. The average is taken over the realizations of the initial state and the observation noises. It is this weighted cost formulation that we address in this paper.

B. Notation and definitions

Let $\bar{J}^{(\gamma)}$ denote the average cost for a given strategy $\gamma=(\gamma_1,\gamma_2)$ of the encoder and the decoder (γ_1 is the function that maps $\mathbf{S}^m+\mathbf{Z}_1^m$ to $\mathbf{U}^m$ for the encoder, and similarly, γ_2 is the mapping function for the decoder). Where there is no confusion, we drop the superscript (γ). Let $\bar{J}_{opt}=\inf_\gamma \bar{J}^{(\gamma)}$ be the optimal cost.

Vectors are denoted in bold font, random variables in upper case, and their realizations in lower case. We use $A \perp\!\!\!\perp B$ to imply that the random variables A and B are independent. $\mathcal{B}$ is used to denote the unit ball in L_2-norm in $\mathbb{R}^m$.

Definition 1 (Lattice): An m-dimensional lattice Λ is a set of points in $\mathbb{R}^m$ such that if $\mathbf{x}^m,\mathbf{y}^m\in\Lambda$, then $\mathbf{x}^m+\mathbf{y}^m\in\Lambda$, and if $\mathbf{x}^m\in\Lambda$, then $-\mathbf{x}^m\in\Lambda$.

Definition 2 (Packing and packing radius): Given an m-dimensional lattice Λ and a radius r, the set $\Lambda+r\mathcal{B}=\{\mathbf{x}^m+r\mathbf{y}^m : x\in\Lambda,\mathbf{y}^m\in\mathcal{B}\}$ is a *packing* of Euclidean m-space if for all points $\mathbf{x}^m,\mathbf{y}^m\in\Lambda$, $(\mathbf{x}^m+r\mathcal{B})\bigcap(\mathbf{y}^m+r\mathcal{B})=\emptyset$. The packing radius r_p is defined as $r_p:=\sup\{r:\Lambda+r\mathcal{B}\text{ is a packing}\}$.

Definition 3 (Covering and covering radius): Given an m-dimensional lattice Λ and a radius r, the set $\Lambda+r\mathcal{B}$ is a *covering* of Euclidean m-space if $\mathbb{R}^m\subset\Lambda+r\mathcal{B}$. The covering radius r_c is defined as $r_c:=\inf\{r:\Lambda+r\mathcal{B}\text{ is a covering}\}$.

Definition 4 (Packing-covering ratio): The *packing-covering ratio* (denoted by ξ) of a lattice Λ is the ratio of its covering radius to its packing radius, $\xi=\frac{r_c}{r_p}$.

III. Equivalence of two problems

In this section we show that the problem of Section II is equivalent to a problem with noise in evolution of state $\mathbf{X}^m$, but noiseless observation at the encoder, shown in Fig. 2(c).

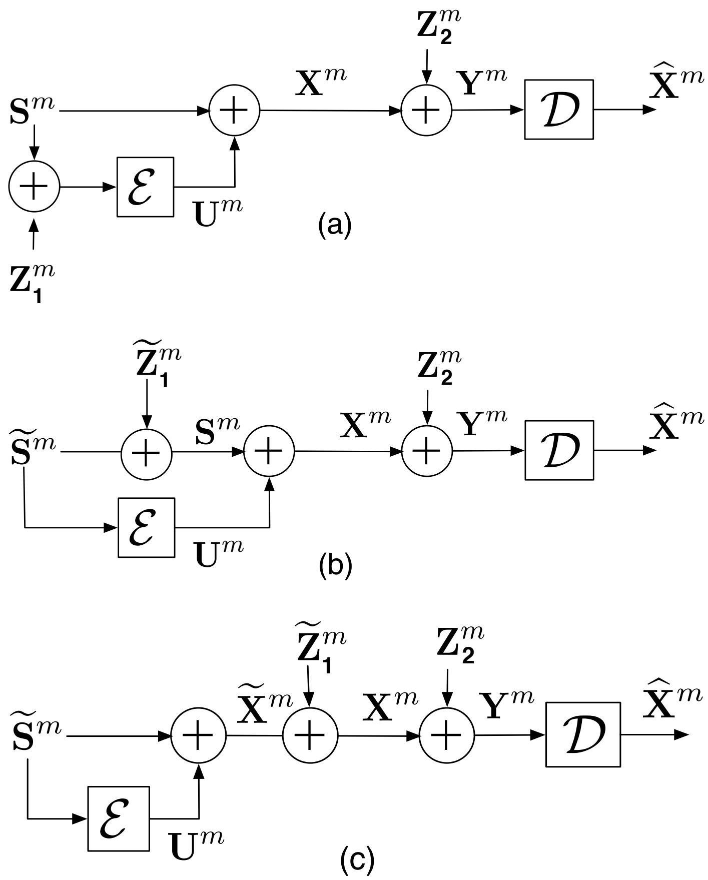

Fig. 2. The figures show how the signal cancelation problem shown in Fig. 1 is equivalent to a problem with noise in the evolution of state $\mathbf{X}^m$, instead of noise in the observation at the encoder. From (c), it is clear that the encoder can not help much in the reconstruction of $\widetilde{\mathbf{Z}}_1^m$ since its observations are independent of $\widetilde{\mathbf{Z}}_1^m$.

In the problem of Section II, the encoder takes an action based on its observation of $\mathbf{S}^m + \mathbf{Z}_1^m$. Define $\widetilde{\mathbf{S}}^m := \alpha(\mathbf{S}^m + \mathbf{Z}_1^m)$, the MMSE estimate of $\mathbf{S}^m$ given $\mathbf{S}^m + \mathbf{Z}_1^m$, where $\alpha = \frac{\sigma^2}{\sigma^2+N_1}$. Since $\widetilde{\mathbf{S}}^m$ can be obtained from $\mathbf{S}^m + \mathbf{Z}_1^m$ with an invertible mapping, we can equivalently assume that the encoder observes $\widetilde{\mathbf{S}}^m$. The initial state can be written as $\mathbf{S}^m = \widetilde{\mathbf{S}}^m + \widetilde{\mathbf{Z}}_1^m$, where $\widetilde{\mathbf{S}}^m \perp\!\!\!\perp \widetilde{\mathbf{Z}}_1^m$ (orthogonality principle), and $\widetilde{\mathbf{Z}}_1^m \sim \mathcal{N}\left(0, \frac{\sigma^2 N_1}{\sigma^2+N_1}\right)$. The resulting block diagram (which represents an equivalent problem) is shown in Fig. 2(b). By commutativity of addition, we get the equivalent problem with noise $\widetilde{\mathbf{Z}}_1^m$ in state evolution, as shown in Fig. 2(c). An intermediate state $\widetilde{\mathbf{X}}^m = \widetilde{\mathbf{S}}^m + \mathbf{U}^m$ is also introduced.

In summary, the equivalent noisy-state evolution problem is the following: the initial state $\widetilde{\mathbf{S}}^m \sim \mathcal{N}(0, \widetilde{\sigma}^2\mathbb{I})$ is observed noiselessly by the encoder $\mathcal{E}$, where $\widetilde{\sigma}^2 = \frac{\sigma^4}{\sigma^2+N_1}$. The encoder modifies the state using an input $\mathbf{U}^m$, resulting in the system state $\mathbf{X}^m$. State evolution noise $\widetilde{\mathbf{Z}}_1^m \sim \mathcal{N}(0, \widetilde{N}_1\mathbb{I})$ is added to the state $\widetilde{\mathbf{X}}^m$ resulting in state $\mathbf{X}^m$. Here, $\widetilde{N}_1 = \frac{\sigma^2 N_1}{\sigma^2+N_1}$. The objective, as before, is to minimize

$$\bar{J} = \frac{1}{m} k^2 \mathbb{E}\left[\|\mathbf{U}^m\|^2\right] + \frac{1}{m}\mathbb{E}\left[\|\mathbf{X}^m - \widehat{\mathbf{X}}^m\|^2\right], \tag{2}$$

where $\widehat{\mathbf{X}}^m$ is the estimate of $\mathbf{X}^m$ at the decoder based on noisy observations of $\mathbf{X}^m$.

IV. Approximately optimal solutions for distributed signal cancelation

A coarse lower bound on the average cost is given in the following.

Theorem 1:

$$\bar{J}_{opt} \geq \max\left\{\frac{\sigma^2 N_1}{\sigma^2 N_1+\sigma^2+N_1}, \inf_{P\geq 0} k^2 P + \left(\left(\sqrt{\widetilde{\kappa}(P)}-\sqrt{P}\right)^+\right)^2\right\},$$

where $\widetilde{\kappa}(P)=\frac{\widetilde{\sigma}^2}{(\widetilde{\sigma}+\sqrt{P})^2+1}$, and $\widetilde{\sigma}^2=\frac{\sigma^4}{\sigma^2+N_1}$.

Proof: Consider the equivalent problem of noise in state evolution of Section III. A lower bound can be derived as follows.

If the decoder is given side information $\widetilde{\mathbf{S}}^m$, it can simulate the encoder, reconstructing $\mathbf{U}^m$ perfectly. Thus the decoder only has to estimate $\widetilde{\mathbf{Z}}_1^m$, which is independent of $\widetilde{\mathbf{S}}^m$. The resulting MMSE is therefore given by $\frac{\widetilde{N}_1}{\widetilde{N}_1+1}=\frac{\sigma^2 N_1}{\sigma^2 N_1+\sigma^2+N_1}$, yielding the first term in the lower bound.

Alternatively, if side-information $\widetilde{\mathbf{Z}}_1^m$ is given to the decoder, the problem reduces to the vector Witsenhausen counterexample, where the encoder observes the source $\widetilde{\mathbf{S}}^m$ noiselessly and there is no noise $\widetilde{\mathbf{Z}}_1^m$ in state evolution. A lower bound can now be obtained from [4, Theorem 1] (using $\widetilde{\sigma}$ in place of σ), yielding the second term in the lower bound. ■

A. An upper bound on the total cost

Define $N_{sum}:=\frac{\sigma^2 N_1}{\sigma^2+N_1}+1$.

Theorem 2: For the noisy extension of Witsenhausen's counterexample of Section II, an upper bound on the optimal costs is

$$\bar{J}_{opt}\leq \min\left\{\bar{J}_{\widetilde{ZI}}, \bar{J}_{\widetilde{ZF}}, \bar{J}_{\widetilde{VQ}}\right\},$$

where $\bar{J}_{\widetilde{ZI}}=\frac{\sigma^2}{\sigma^2+1}$, $\bar{J}_{\widetilde{ZF}}=k^2\frac{\sigma^4}{\sigma^2+N_1}+\frac{\sigma^2 N_1}{\sigma^2+N_1+\sigma^2 N_1}$, and

$$\bar{J}_{\widetilde{VQ}} \leq \inf_{P\geq 0} k^2 P+\frac{\sigma^2 N_1}{\sigma^2+N_1}+ \left(\sqrt{N_{sum}}\sqrt{\psi\left(m+2,\sqrt{\frac{mP}{\xi^2 N_{sum}}}\right)} + \sqrt{\frac{P}{\xi^2}}\sqrt{\psi\left(m,\sqrt{\frac{mP}{\xi^2 N_{sum}}}\right)}\right)^2, \tag{3}$$

where $\psi(m,r):=\Pr(\mathbf{Z}^m\geq r)=\int_{\mathbf{z}^m\geq r}\frac{e^{-\frac{\mathbf{z}^m}{2}}}{\left(\sqrt{2\pi}\right)^m}d\mathbf{z}^m$ for $\mathbf{Z}^m\sim\mathcal{N}(0,\mathbb{I})$, and ζ is the packing-covering ratio of a lattice in $\mathbb{R}^m$.

Proof: We provide three strategies. Depending on k, σ, and N_1, we use the best of the three. The obtained upper bound is therefore the minimum of the costs attained by the three startegies. The strategies are defined on the equivalent problem of noise in the state evolution (of Section III).

The first strategy is the Zero-Input ($\widetilde{ZI}$) strategy, where the input $\mathbf{U}_1^m=0$. The decoder merely estimates $\widetilde{\mathbf{S}}^m+\widetilde{\mathbf{Z}}_1^m=\mathbf{S}^m$ from the noisy observation $\mathbf{S}^m+\mathbf{Z}_2^m$. Since $\mathbf{Z}_2^m\sim\mathcal{N}(0,\mathbb{I})$, the LLSE error is given by

$$MMSE=\frac{\sigma^2}{\sigma^2+1}, \tag{4}$$

which is also the attained cost since $P=0$.

Our second strategy is a Zero-Forcing ($\widetilde{ZF}$) strategy, applied to the equivalent noisy state-evolution problem. The first input forces the state $\widetilde{\mathbf{S}}^m$ to zero, requiring an average power of $P=\widetilde{\sigma}^2=\frac{\sigma^4}{\sigma^2+N_1}$. The decoder merely performs an LLSE estimation for $\widetilde{\mathbf{Z}}_1^m \sim \mathcal{N}(0,\widetilde{N})$. The $MMSE$ error is therefore given by

$$MMSE_{\widetilde{ZF}}=\frac{\widetilde{N}}{\widetilde{N}+1}=\frac{\sigma^2 N_1}{\sigma^2+N_1+\sigma^2 N_1}. \tag{5}$$

The cost for $\widetilde{ZF}$ is, therefore, $\bar{J}_{\widetilde{ZF}}=k^2\frac{\sigma^4}{\sigma^2+N_1}+\frac{\sigma^2 N_1}{\sigma^2+N_1+\sigma^2 N_1}$.

The first two strategies are linear, and therefore somewhat uninteresting. Our third strategy is a nonlinear strategy that is based on the idea of implicit communication (see, for example, [13]).

This strategy, which we call the Vector Quantization ($\widetilde{VQ}$) strategy, uses a lattice $\Lambda\subset\mathbb{R}^m$ of covering radius r_c, packing radius r_p, and packing-covering ratio of $\zeta=\frac{r_c}{r_p}$ as follows. The encoder uses the input $\mathbf{U}^m$ to force $\widetilde{\mathbf{S}}^m$ to a lattice point. The decoder declares $\widehat{\mathbf{X}}^m$ to be the quantization point that is within a distance r_p of its observation $\mathbf{Y}^m$, if any such quantization point exists. If there is no such quantization point, the decoder declares $\widehat{\mathbf{X}}^m=\mathbf{Y}^m$. It is shown in Appendix I that choosing $r_c=\sqrt{mP}$ so that $r_p=\sqrt{\frac{mP}{\xi^2}}$, the costs attained by $\widetilde{VQ}$ are bounded by the expression in (3). Note that this upper bound depends on ξ, the packing-covering ratio for the chosen lattice $\Lambda\subset\mathbb{R}^m$.

The upper bound can now be obtained by using the best of $\widetilde{ZI}$, $\widetilde{ZF}$, and $\widetilde{VQ}$ strategies depending on the values of k and σ. ■

It is also shown in Appendix I that by loosening the upper bound in (3), one can obtain the following bound

$$\begin{aligned}\bar{J}_{\widetilde{VQ}} \leq & \inf_{P>\zeta^2 N_{sum}} k^2 P+\frac{\sigma^2 N_1}{\sigma^2+N_1} \\ & +\left(\sqrt{N_{sum}}+\sqrt{\frac{P}{\zeta^2}}\right)^2 e^{-\frac{mP}{2\zeta^2 N_{sum}}+\frac{m+2}{2}\left(1+\ln\left(\frac{P}{\zeta^2 N_{sum}}\right)\right)}.\end{aligned}$$

It follows that, for any $P>\xi^2 N_{sum}$, for the asymptotic problem

$$\limsup_{m\to\infty}\bar{J}_{\widetilde{VQ}}\leq k^2 P+\frac{\sigma^2 N_1}{\sigma^2+N_1}.$$

B. Approximate asymptotic optimality

Theorem 3 (Approximate asymptotic optimality): In the limit of $m\to\infty$,

$$\begin{aligned} & \max\Big\{\tfrac{\sigma^2 N_1}{\sigma^2 N_1+\sigma^2+N_1}, \\ & \inf_{P\geq 0} k^2 P+\left(\left(\sqrt{\kappa(P)}-\sqrt{P}\right)^+\right)^2\Big\} \\ & \leq \bar{J}_{opt} \\ & \leq \gamma \max\Big\{\tfrac{\sigma^2 N_1}{\sigma^2 N_1+\sigma^2+N_1}, \\ & \inf_{P\geq 0} k^2 P+\left(\left(\sqrt{\kappa(P)}-\sqrt{P}\right)^+\right)^2\Big\},\end{aligned}$$

where $\gamma\leq 80$.

Proof: See Appendix II. ■

Numerical evaluations (shown in Fig. 3) show that the ratio is actually bounded by 10.

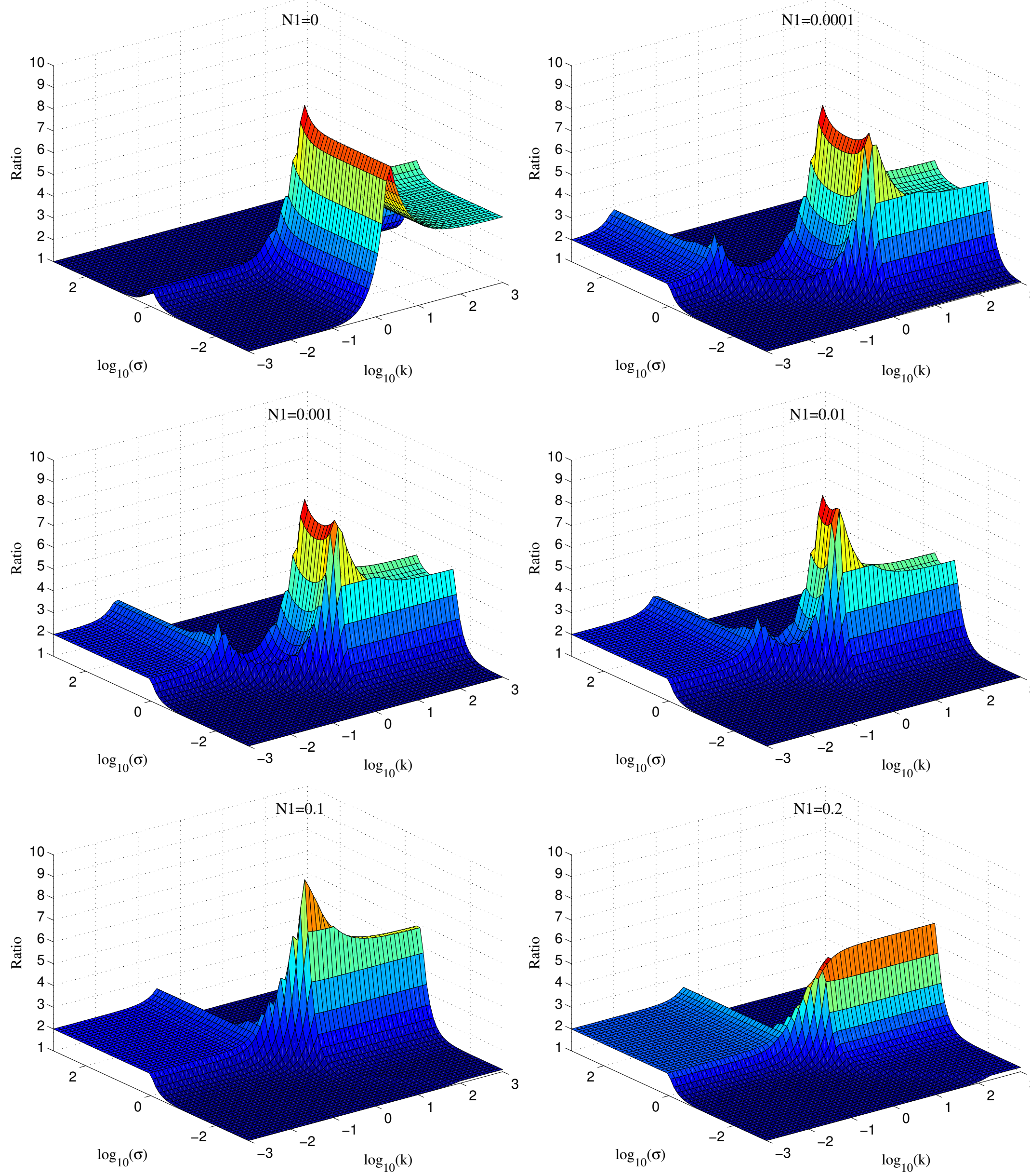

Fig. 3. Ratio of the upper and lower bound for varying values of N_1 in the asymptotic case. The maximum was observed to be smaller than 10 for all values of N_1 that were tested.

C. Approximate optimality for finite lengths

In this section, we derive a lower bound on the cost for finite lengths. While Theorem 1 gives one such bound that is valid for each vector length m, it is not sufficient to show approximate optimality of the quantization-based schemes. The lower bound in this section is derived using the bound in [12] which was derived for the original counterexample.

Theorem 4:

$$\begin{aligned}\bar{J}_{opt} \quad &\geq \quad \max\Bigg\{\frac{\sigma^2 N_1}{\sigma^2 N_1 + \sigma^2 + N_1},\\ &\qquad \inf_{P\geq 0}\sup_{\sigma_G^2\geq 1, L>0} k^2 P + \eta(P,\widetilde{\sigma}^2,\sigma_G^2,L)\Bigg\},\end{aligned}$$

where

$$\eta(P,\widetilde{\sigma}^2,\sigma_G^2,L) = \frac{\sigma_G^m}{c_m(L)}\exp\left(-\frac{mL^2(\sigma_G^2-1)}{2}\right) \left(\left(\sqrt{\kappa_2(P,\widetilde{\sigma}^2,\sigma_G^2,L)}-\sqrt{P}\right)^+\right)^2,$$

where $\kappa_2(P,\widetilde{\sigma}^2,\sigma_G^2,L) :=$

$$\frac{\widetilde{\sigma}^2\sigma_G^2}{c_m^{\frac{2}{m}}(L)e^{1-d_m(L)}\left((\widetilde{\sigma}+\sqrt{P})^2+d_m(L)\sigma_G^2\right)},$$

$c_m(L) := \frac{1}{\Pr(\|\mathbf{Z}_2^m\|^2\leq mL^2)} = (1-\psi(m,L\sqrt{m}))^{-1}$, $d_m(L) := \frac{\Pr(\|\mathbf{Z}_2^{m+2}\|^2\leq mL^2)}{\Pr(\|\mathbf{Z}_2^m\|^2\leq mL^2)} = \frac{1-\psi(m+2,L\sqrt{m})}{1-\psi(m,L\sqrt{m})}$, $0 < d_m(L) < 1$, and $\psi(m,r) = \Pr(\|\mathbf{Z}_2^m\| \geq r)$.

Proof: Follows along the same lines as that of Theorem 1. When side information about $\widetilde{\mathbf{Z}}_1^m$ is supplied to the decoder, the lower bound from [12, Theorem 3] is used instead of that from [4]. ■

Theorem 5 (Approximate optimality at finite lengths): For the signal cancelation problem of vector length m described in Section II,

$$\max\left\{\frac{\sigma^2N_1}{\sigma^2N_1+\sigma^2+N_1}, \inf_{P\geq 0}\sup_{\sigma_G^2\geq 1,L>0} k^2P+\eta(P,\widetilde{\sigma}^2,\sigma_G^2,L)\right\}$$
$$\leq \bar{J}_{opt}$$
$$\leq 400\zeta^2\max\left\{\frac{\sigma^2N_1}{\sigma^2N_1+\sigma^2+N_1}, \inf_{P\geq 0}k^2\sup_{\sigma_G^2\geq 1,L>0} P+\eta(P,\widetilde{\sigma}^2,\sigma_G^2,L)\right\},$$

Proof: See Appendix III. ■

For any $m \in \mathbb{Z}^+$, there exists a lattice with $\zeta \leq 4$, and in the limit $m \to \infty$, $\zeta \leq 2$ [17]. Thus the problem is solved to within a constant factor for all vector lengths.

V. Discussions

Even though these constants seem large, the actual ratios obtained by numerical evaluations are much smaller (e.g. see [12]). A straightforward improvement in the upper bound can be obtained using dirty-paper coding [4], and in the lower bound using techniques from [10].

It is easy to check that the equivalent problem in Section III can represent any problem of noise in state evolution by varying the parameters σ and N_1 in the original problem of noisy observation at the encoder. Thus the results of this paper also provide characterizations of optimal cost for noise in state evolution within a constant factor. Noise in the first input $\mathbf{U}^m$ and in evolution of state $\mathbf{S}^m$ can also be lumped with the noise in state evolution. Thus the solution here extends easily to a general problem with noises in all state evolutions, inputs, and observations.

An interesting feature that shows up is that the zero-input strategy is approximately optimal (to within a factor of 2) for $N_1 > N_2$. This suggests that the controllers operate opportunistically — if the encoder is noisier, it does not do any work. We expect that this feature will be retained in extensions of the problem where there are multiple controllers operating sequentially.

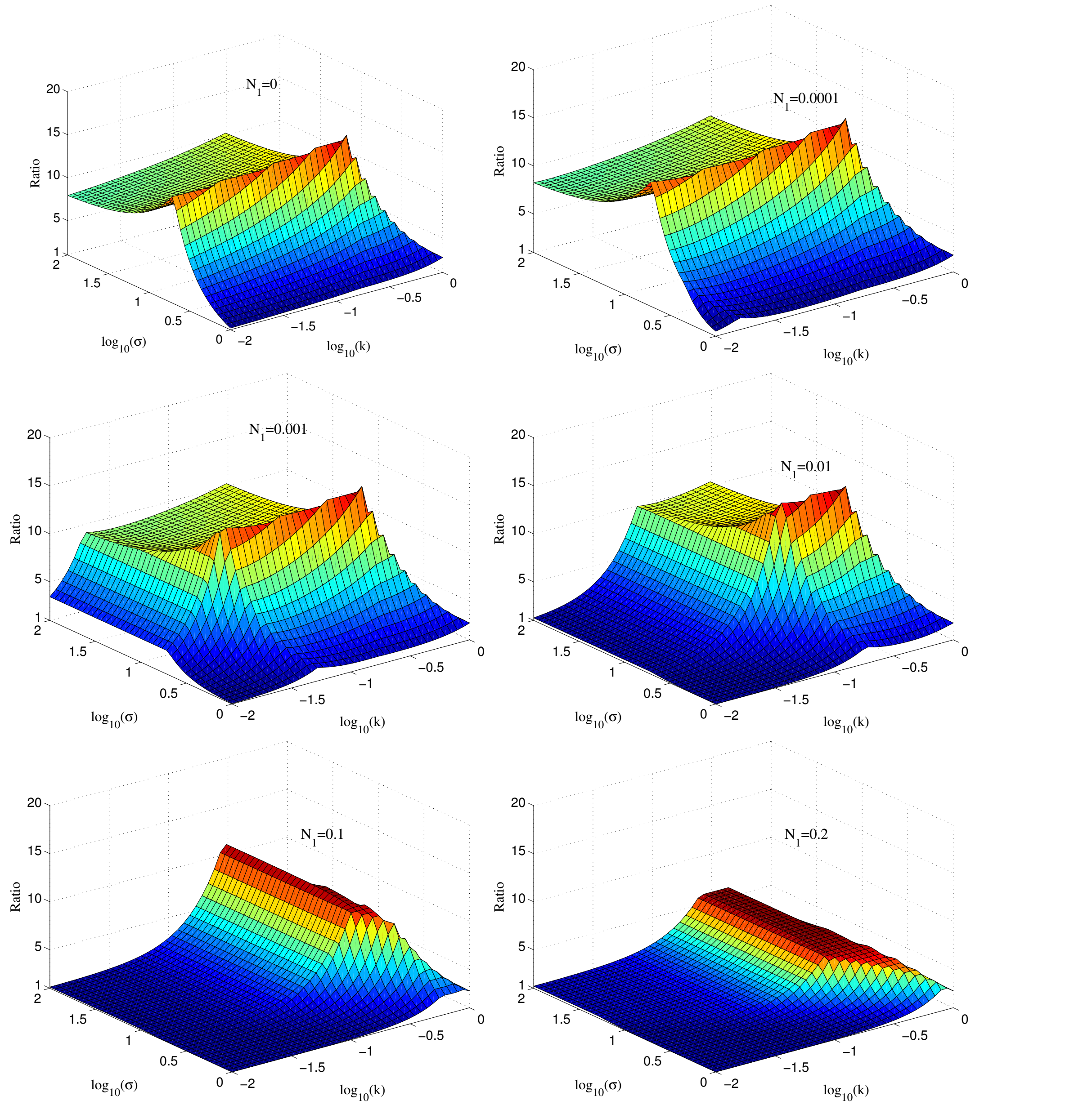

Fig. 4. Ratio of the upper and lower bound for varying values of N_1 for the scalar case. The maximum was observed to be smaller than 20 for all values of N_1 that were tested.

Acknowledgments

Helpful discussions with Hari Palaiyanur are gratefully acknowledged.

References

[1] Dimitri Bertsekas. *Dynamic Programming*. Athena Scientific, Belmont, MA, 1995.

[2] Y.-H. Kim, A. Sutivong, and T. M. Cover. State amplification. *IEEE Trans. Inform. Theory*, 54(5):1850–1859, May 2008.

[3] N Merhav and S Shamai. Information rates subject to state masking. *IEEE Trans. Inform. Theory*, 53(6):2254–2261, June 2007.

[4] P Grover and A Sahai. Vector Witsenhausen counterexample as assisted interference suppression. *Special issue on Information Processing and Decision Making in Distributed Control Systems of the International Journal on Systems, Control and Communications (IJSCC)*, 2010.

[5] Nuno C Martins. Witsenhausen's counter example holds in the presence of side information. *Proceedings of the 45th IEEE Conference on Decision and Control (CDC)*, pages 1111–1116, 2006.

[6] H. S. Witsenhausen. A counterexample in stochastic optimum control. *SIAM Journal on Control*, 6(1):131–147, January 1968.

[7] Krishnan Eswaran. *Remote Source Coding and AWGN CEO Problems*. PhD thesis, UC Berkeley, Berkeley, CA, 2006.

[8] GM Lipsa and NC Martins. Finite horizon optimal memoryless control of a delay in Gaussian noise: A simple counterexample. *Proceedings of the 47th IEEE Conference on Decision and Control (CDC)*, pages 1628–1635, 2008.

[9] Daniela Tuninetti and Christina Fragouli. Processing along the way: forwarding vs. coding. In *International Symposium on Information Theory and its Applications (ISITA)*, 2004.

[10] Pulkit Grover, Aaron B Wagner, and Anant Sahai. Information embedding meets distributed control. *Proceedings of the IEEE Information Theory Workshop (ITW)*, 2010.

[11] P Grover, A Sahai, and SY Park. The finite-dimensional Witsenhausen counterexample. In *7th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt), Workshop on Control over Communication Channels (ConCom)*, 2009.

[12] P Grover, SY Park, and A Sahai. The finite-dimensional Witsenhausen counterexample. *IEEE Transactions on Automatic Control, Submitted*, 2010.

[13] Pulkit Grover, Se Yong Park, and Anant Sahai. On the generalized Witsenhausen counterexample. In *Proceedings of the Allerton Conference on Communication, Control, and Computing*, Monticello, IL, October 2009.

[14] AS Avestimehr, S Diggavi, and DNC Tse. A deterministic approach to wireless relay networks. In *Proc. of the Allerton Conference on Communications, Control and Computing*, October 2007.

[15] R Etkin, DNC Tse, and H Wang. Gaussian interference channel capacity to within one bit. *IEEE Trans. Inform. Theory*, 54(12), December 2008.

[16] Pulkit Grover and Anant Sahai. Distributed signal cancelation inspired by Witsenhausen's counterexample. http://www.eecs.berkeley.edu/∼sahai/Papers/ISIT10Witsenhausen.pdf, 2010.

[17] Uri Erez, Simon Litsyn, and Ram Zamir. Lattices which are good for (almost) everything. *IEEE Trans. Inform. Theory*, 51(10):3401–3416, October 2005.

APPENDIX I
COSTS ATTAINED BY THE VECTOR QUANTIZATION $\widetilde{VQ}$ STRATEGY

Define the error event $\mathcal{E}_m := \{\|\widetilde{\mathbf{z}}_1^m + \mathbf{z}_2^m\| > r_p\}$.

$$\begin{aligned}&\mathbb{E}\left[\|\mathbf{X}^m - \widehat{\mathbf{X}}^m\|^2\right]\\ &= \mathbb{E}\left[\|\mathbf{X}^m - \widehat{\mathbf{X}}^m\|^2 \mathbb{1}_{\{\mathcal{E}_m\}}\right] + \mathbb{E}\left[\|\mathbf{X}^m - \widehat{\mathbf{X}}^m\|^2 \mathbb{1}_{\{\mathcal{E}_m^c\}}\right].\end{aligned}$$

The event $\mathcal{E}_m$ can happen in two ways. In the first case, the decoder estimate $\widehat{\mathbf{x}}^m \in \Lambda$, which is the case when $\mathbf{y}^m$ falls in a packing sphere. In this case,

$$\begin{aligned}\|\mathbf{x}^m - \widehat{\mathbf{x}}^m\| &= \|\mathbf{x}^m - \mathbf{y}^m + \mathbf{y}^m - \widehat{\mathbf{x}}^m\|\\ &\leq \|\mathbf{x}^m - \mathbf{y}^m\| + \|\mathbf{y}^m - \widehat{\mathbf{x}}^m\| \leq \|\mathbf{z}_2^m\| + r_p.\end{aligned}$$

If $\widehat{\mathbf{x}}^m \notin \Lambda$, $\widehat{\mathbf{x}}^m = \mathbf{y}^m$, and therefore,

$$\|\mathbf{x}^m - \widehat{\mathbf{x}}^m\| = \|\mathbf{z}_2^m\| \leq \|\mathbf{z}_2^m\| + r_p. \tag{6}$$

Therefore, in either case, under the error event $\mathcal{E}_m$, $\|\mathbf{x}^m - \widehat{\mathbf{x}}^m\| = \|\mathbf{z}_2^m\| \leq \|\mathbf{z}_2^m\| + r_p$. We now need the following lemma

Lemma 1: For $P > \zeta^2 N_{sum}$,

$$\begin{aligned}&\frac{1}{m}\mathbb{E}\left[\left(\|\mathbf{Z}_2^m\| + r_p\right)^2 \mathbb{1}_{\{\mathcal{E}_m\}}\right]\\ &\leq N_{sum}\left(1 + \sqrt{\frac{P}{\zeta^2 N_{sum}}}\right)^2 e^{-\frac{mP}{2\zeta^2 N_{sum}} + \frac{m+2}{2}\left(1 + \ln\left(\frac{P}{\zeta^2 N_{sum}}\right)\right)}.\end{aligned}$$

Proof: Using the Cauchy-Schwartz inequality,

$$\begin{aligned}&\mathbb{E}\left[\left(\|\mathbf{Z}_2^m\|+r_p\right)^2\mathbb{1}_{\{\mathcal{E}_m\}}\right]\\ \leq\ &\left(\sqrt{\mathbb{E}\left[\|\mathbf{Z}_2^m\|^2\mathbb{1}_{\{\mathcal{E}_m\}}\right]}+\sqrt{\mathbb{E}\left[r_p^2\mathbb{1}_{\{\mathcal{E}_m\}}\right]}\right)^2.\end{aligned}\tag{7}$$

Defining $\mathbf{Z}_3^m=-\mathbf{Z}_2^m$,

$$\begin{aligned}\mathbb{E}\left[\left(\widetilde{\mathbf{Z}}_1^m\right)^T\mathbf{Z}_2^m\mathbb{1}_{\{\mathcal{E}_m\}}\right] &= -\mathbb{E}\left[\left(\widetilde{\mathbf{Z}}_1^m\right)^T\mathbf{Z}_3^m\mathbb{1}_{\left\{\|\widetilde{\mathbf{Z}}_1^m-\mathbf{Z}_3^m\|>r_p\right\}}\right]\\ &\overset{(a)}{=} -\mathbb{E}\left[\left(\widetilde{\mathbf{Z}}_1^m\right)^T\mathbf{Z}_2^m\mathbb{1}_{\{\mathcal{E}_m\}}\right]\\ &= 0,\end{aligned}$$

where (a) follows from the fact that $\Pr\left((\widetilde{\mathbf{Z}}_1^m,\mathbf{Z}_2^m)\in\mathcal{A}\right)=\Pr\left((\widetilde{\mathbf{Z}}_1^m,\mathbf{Z}_2^m)\in\mathcal{A}\right)=\Pr\left((\widetilde{\mathbf{Z}}_1^m,\mathbf{Z}_3^m)\in\mathcal{A}\right)$ for any set $\mathcal{A}\subset\mathbf{R}^{2m}$. Thus,

$$\begin{aligned}\mathbb{E}\left[\|\widetilde{\mathbf{Z}}_1^m+\mathbf{Z}_2^m\|^2\mathbb{1}_{\{\mathcal{E}_m\}}\right] &= \mathbb{E}\left[\|\widetilde{\mathbf{Z}}_1^m\|^2\mathbb{1}_{\{\mathcal{E}_m\}}\right]+\mathbb{E}\left[\|\mathbf{Z}_2^m\|^2\mathbb{1}_{\{\mathcal{E}_m\}}\right]\\ &\quad +2\mathbb{E}\left[\left(\widetilde{\mathbf{Z}}_1^m\right)^T\mathbf{Z}_2^m\mathbb{1}_{\{\mathcal{E}_m\}}\right]\\ &= \mathbb{E}\left[\|\widetilde{\mathbf{Z}}_1^m\|^2\mathbb{1}_{\{\mathcal{E}_m\}}\right]+\mathbb{E}\left[\|\mathbf{Z}_2^m\|^2\mathbb{1}_{\{\mathcal{E}_m\}}\right]\\ &\geq \mathbb{E}\left[\|\mathbf{Z}_2^m\|^2\mathbb{1}_{\{\mathcal{E}_m\}}\right].\end{aligned}$$

Thus, using (7)

$$\begin{aligned}&\mathbb{E}\left[\left(\|\mathbf{Z}_2^m\|+r_p\right)^2\mathbb{1}_{\{\mathcal{E}_m\}}\right]\\ \leq\ &\left(\sqrt{\mathbb{E}\left[\|\mathbf{Z}_1^m+\mathbf{Z}_2^m\|^2\mathbb{1}_{\{\mathcal{E}_m\}}\right]}+\sqrt{\mathbb{E}\left[r_p^2\mathbb{1}_{\{\mathcal{E}_m\}}\right]}\right)^2.\end{aligned}$$

The proof now follows from Lemma 1 of [12] (by using a noise of variance N_{sum} instead of 1). ■

Now, under the event $\mathcal{E}_m^c$, $\widehat{\mathbf{x}}^m=\widetilde{\mathbf{x}}^m$. Thus,

$$\|\mathbf{x}^m-\widehat{\mathbf{x}}^m\|=\|\left(\widetilde{\mathbf{x}}^m+\widetilde{\mathbf{z}}_1^m\right)-\widetilde{\mathbf{x}}^m\|=\|\widetilde{\mathbf{z}}_1^m\|.$$

Therefore,

$$\begin{aligned}\mathbb{E}\left[\|\mathbf{X}^m-\widehat{\mathbf{X}}^m\|^2\mathbb{1}_{\{\mathcal{E}_m^c\}}\right] &= \mathbb{E}\left[\|\widetilde{\mathbf{Z}}_1^m\|^2\mathbb{1}_{\{\mathcal{E}_m^c\}}\right]\\ &\leq \mathbb{E}\left[\|\widetilde{\mathbf{Z}}_1^m\|^2\right]=\frac{\sigma^2N_1}{\sigma^2+N_1}.\end{aligned}\tag{8}$$

The result now follows from Lemma 1 and (8).

Appendix II
Constant factor optimality for asymptotically infinite vector length

The proof involves showing that the ratio of the upper bound of Theorem 2 and the lower bound of Theorem 1 is no larger than 80. This is done by dividing the (k,σ,N_1) space into different regions, which are dealt with separately.

An optimal value of P that attains the minimum in the second expression in the lower bound of Theorem 1 is denoted by P^*.

Case 1: $N_1 \geq 1$.

A lower bound is

$$\begin{aligned}\bar{J}_{opt} &\geq \frac{\sigma^2 N_1}{\sigma^2 N_1 + \sigma^2 + N_1} \\ &\overset{(N_1 \geq 1)}{\geq} \frac{\sigma^2}{\sigma^2 + \sigma^2 + 1} = \frac{\sigma^2}{2\sigma^2 + 1}.\end{aligned}$$

The zero-input upper bound $\bar{J}_{\widetilde{ZI}} = \frac{\sigma^2}{\sigma^2+1}$. The ratio of the upper and lower bounds is therefore smaller than

$$\frac{2\sigma^2 + 1}{\sigma^2 + 1} < 2. \tag{9}$$

Case 2: $\sigma^2 < N_1 < 1$.

If $N_1 > \sigma^2$, using the first term in the lower bound of Theorem 1,

$$\begin{aligned}\bar{J}_{opt} &\geq \frac{\sigma^2 N_1}{\sigma^2 N_1 + \sigma^2 + N_1} \\ &\overset{(N_1 > \sigma^2)}{>} \frac{\sigma^2 \sigma^2}{\sigma^2 \sigma^2 + \sigma^2 + \sigma^2} = \frac{\sigma^4}{\sigma^4 + 2\sigma^2} \\ &\overset{(\sigma^2 < 1)}{>} \frac{\sigma^4}{\sigma^2 + 2\sigma^2} = \frac{\sigma^2}{3}.\end{aligned}$$

The $\widetilde{ZI}$ upper bound $\bar{J}_{\widetilde{ZI}} = \frac{\sigma^2}{\sigma^2+1} < \sigma^2$. Thus the ratio of upper and lower bounds is smaller than 3.

Case 3: $N_1 < \sigma^2 < 1$.

Case 3a: $P^* \geq \frac{\sigma^2}{16}$.

Since the lower bound is the larger of the two terms in Theorem 1, it is larger than any convex combination of the two terms as well. That is,

$$\begin{aligned}\bar{J}_{opt} &\geq \frac{1}{2}\left(k^2 P^* + \left(\left(\sqrt{\widetilde{\kappa}} - \sqrt{P^*}\right)^+\right)^2\right) \\ &\quad + \frac{1}{2}\frac{\sigma^2 N_1}{\sigma^2 N_1 + \sigma^2 + N_1} \\ &\overset{\left(P^* \geq \frac{\sigma^2}{16}\right)}{\geq} \frac{k^2\sigma^2}{32} + \frac{\sigma^2 N_1}{2(\sigma^2 N_1 + \sigma^2 + N_1)}.\end{aligned}$$

Now for the upper bound, we use the zero-forcing strategy

$$\begin{aligned}\bar{J}_{\widetilde{ZF}} &= \frac{k^2\sigma^4}{\sigma^2 + N_1} + \frac{\sigma^2 N_1}{\sigma^2 N_1 + \sigma^2 + N_1} \\ &\leq \frac{k^2\sigma^4}{\sigma^2} + \frac{\sigma^2 N_1}{\sigma^2 N_1 + \sigma^2 + N_1} \\ &= k^2\sigma^2 + \frac{\sigma^2 N_1}{\sigma^2 N_1 + \sigma^2 + N_1}.\end{aligned}$$

The ratio of upper and lower bound is therefore smaller than $\max\{32, 2\} = 32$.

Case 3b: $P^* < \frac{\sigma^2}{16}$.

Since $N_1 < \sigma^2$,

$$\widetilde{\sigma}^2 = \frac{\sigma^4}{\sigma^2 + N_1} \overset{(N_1 < \sigma^2)}{\geq} \frac{\sigma^4}{\sigma^2 + \sigma^2} = \frac{\sigma^2}{2}.$$

Thus,

$$\begin{aligned}\widetilde{\kappa} &= \frac{\widetilde{\sigma}^2}{(\widetilde{\sigma}+\sqrt{P^*})^2+1} \geq \frac{\sigma^2/2}{\left(\frac{\sigma}{\sqrt{2}}+\frac{\sigma}{4}\right)^2+1} \\ &\overset{(\sigma^2\leq 1)}{\geq} \frac{\sigma^2}{2\left(\frac{1}{\sqrt{2}}+\frac{1}{4}\right)^2+1} \geq \frac{\sigma^2}{3}.\end{aligned}$$

Thus,

$$(\sqrt{\widetilde{\kappa}}-\sqrt{P^*})^2 \geq \sigma^2\left(\frac{1}{\sqrt{3}}-\frac{1}{4}\right)^2 > 0.1\sigma^2.$$

Using $\bar{J}_{\widetilde{ZI}} = \frac{\sigma^2}{\sigma^2+1} < \sigma^2$, the ratio of the upper and lower bounds is smaller than 10.

Case 4: $N_1 \leq 1 < \sigma^2$.

Case 4a: $P^* \leq \frac{1}{9}$.

In this case,

$$\widetilde{\sigma}^2 = \frac{\sigma^4}{\sigma^2+N_1} \overset{(N_1\leq 1\leq \sigma^2)}{\geq} \frac{\sigma^4}{\sigma^2+\sigma^2} = \frac{\sigma^2}{2}$$

Therefore,

$$\widetilde{\kappa} = \frac{\widetilde{\sigma}^2}{(\widetilde{\sigma}+\sqrt{P^*})^2+1} \geq \frac{\sigma^2/2}{\left(\frac{\sigma}{\sqrt{2}}+\frac{1}{3}\right)^2+1} \geq 0.24.$$

Thus, $\left(\left(\sqrt{\widetilde{\kappa}}-\sqrt{P^*}\right)^+\right)^2 \geq 0.024$. The zero-input upper bound is smaller than 1. Thus the ratio is smaller than $\frac{1}{0.024} < 41$.

Case 4b: $P^* > \frac{1}{9}$

A lower bound is

$$\begin{aligned}\bar{J}_{opt} &\geq \max\left\{\frac{k^2}{9}, \frac{\sigma^2 N_1}{\sigma^2 N_1+\sigma^2+N_1}\right\} \\ &\geq \frac{k^2}{9}\times\frac{9}{10} + \frac{\sigma^2 N_1}{\sigma^2 N_1+\sigma^2+N_1}\times\frac{1}{10} \\ &= \frac{k^2}{10} + \frac{\sigma^2 N_1}{10(\sigma^2 N_1+\sigma^2+N_1)} \end{aligned} \tag{10}$$

Now, we use the asymptotic vector quantization upper bound of

$$\lim_{m\to\infty} \bar{J}_{\widetilde{VQ}} \leq k^2\xi^2\left(\frac{\sigma^2 N_1}{\sigma^2+N_1}+1\right) + \frac{\sigma^2 N_1}{\sigma^2+N_1}. \tag{11}$$

As $m\to\infty$, there exist lattices whose packing covering ratio is asymptotically at most 2. Since $N_1 < 1$, this upper bound is smaller than $2\xi^2 k^2 + \frac{\sigma^2 N_1}{\sigma^2 N_1+\sigma^2+N_1} \leq 8k^2 + \frac{\sigma^2 N_1}{\sigma^2 N_1+\sigma^2+N_1}$. The ratio of the first terms in the upper bound and the lower bound of (10) is at most 80. The ratio of the second terms is

$$\begin{aligned}\frac{\sigma^2 N_1}{\sigma^2+N_1}\times\frac{10(\sigma^2 N_1+\sigma^2+N_1)}{\sigma^2 N_1} &= 10\frac{\sigma^2 N_1}{\sigma^2+N_1}+10 \\ &\leq 10+10 = 20.\end{aligned}$$

Thus the ratio of the upper and lower bounds is no larger than $\max\{8, 80\} = 80$.

Thus the ratio of upper and lower bounds is at most 80.

Appendix III
Constant factor optimality for finite vector lengths

Proof: Again, we divide the (k, σ, N_1)-space into different regions and prove constant factor optimality for each of them.

Case 1: $N_1 \geq 1$.

Case 1 in the proof for theorem 3 shows that the ratio of zero-input upper bound and the infinite-length lower bound is smaller than 2. This works even in the finite-length case because zero-input strategy has the same cost for any vector length.

Case 2: $N_1 < 1, \sigma^2 < 1$.

Again, *Case 2* and *Case 3* of proof for theorem 3 show that zero-input and zero-forcing strategies attain within a factor of 32 of the optimal for this case.

Case 3: $N_1 < 1, \sigma^2 \geq 1, P^* \geq \frac{\widetilde{\sigma}^2}{100}$.

In this case, the first term in the lower bound can be lower bounded as follows

$$\begin{aligned} k^2 P^* &\geq k^2 \frac{\widetilde{\sigma}^2}{100} \\ &= k^2 \frac{\sigma^4}{100(\sigma^2 + N_1)} \\ &\overset{N_1 < 1 \leq \sigma^2}{\geq} k^2 \frac{\sigma^4}{100(\sigma^2 + \sigma^2)} = k^2 \frac{\sigma^2}{200}. \end{aligned}$$

Thus, a lower bound to the costs is given by

$$\begin{aligned} \bar{J}_{opt} &\geq \max\left\{k^2 \frac{\sigma^2}{200}, \frac{\sigma^2 N_1}{\sigma^2 + N_1 + \sigma^2 N_1}\right\} \\ &\geq k^2 \frac{\sigma^2}{200} \times \frac{100}{101} + \frac{\sigma^2 N_1}{\sigma^2 + N_1 + \sigma^2 N_1} \times \frac{1}{101} \\ &= k^2 \frac{\sigma^2}{202} + \frac{\sigma^2 N_1}{101(\sigma^2 + N_1 + \sigma^2 N_1)}. \end{aligned}$$

An upper bound on the costs is given by $\bar{J}_{\widetilde{ZF}} = k^2 \sigma^2 + \frac{\sigma^2 N_1}{\sigma^2 + N_1 + \sigma^2 N_1}$.

The ratio of upper and lower bounds for this case is therefore smaller than 202.

Case 4: $N_1 < 1, \sigma^2 \geq 1, P^* < \frac{\widetilde{\sigma}^2}{100}, \widetilde{\sigma}^2 \leq 16$.

In this case, we use the lower bound of Theorem 1, which is a special case of the lower bound of Theorem 4

$$\begin{aligned} \widetilde{\kappa} &= \frac{\widetilde{\sigma}^2}{(\widetilde{\sigma} + \sqrt{P^*})^2 + 1} \overset{\left(P^* < \frac{\widetilde{\sigma}^2}{100}\right)}{\geq} \frac{\widetilde{\sigma}^2}{\widetilde{\sigma}^2 \left(1 + \frac{1}{\sqrt{100}}\right)^2 + 1} \\ &\overset{(\widetilde{\sigma}^2 < 16)}{\geq} \frac{\widetilde{\sigma}^2}{16\left(1 + \frac{1}{\sqrt{100}}\right)^2 + 1} = \frac{\widetilde{\sigma}^2}{20.36} \geq \frac{\widetilde{\sigma}^2}{21}. \end{aligned}$$

Thus, for $\widetilde{\sigma}^2 < 16$ and $P^* \leq \frac{\widetilde{\sigma}^2}{100}$,

$$\bar{J}_{opt} \geq \left((\sqrt{\widetilde{\kappa}} - \sqrt{P^*})^+\right)^2 \geq \widetilde{\sigma}^2 \left(\frac{1}{\sqrt{21}} - \frac{1}{\sqrt{100}}\right)^2 \approx 0.014\widetilde{\sigma}^2 \geq \frac{\widetilde{\sigma}^2}{72}.$$

Using the upper bound $\bar{J}_{\widetilde{ZI}} = \frac{\sigma^2}{\sigma^2+1}$, the ratio of upper and lower bounds is smaller than

$$\begin{aligned}\frac{\sigma^2}{\sigma^2+1} \times \frac{72}{\widetilde{\sigma}} &= \frac{\sigma^2}{\sigma^2+1} \times \frac{72(\sigma^2+1)}{\sigma^4} \\ &= \frac{72}{\sigma^2} < 72,\end{aligned}$$

since $\sigma^2 > 1$.

Case 5: $N_1 < 1, \sigma^2 \geq 1, P^* < \frac{\widetilde{\sigma}^2}{100}, \widetilde{\sigma}^2 > 16, P^* < \frac{1}{2}$.

Again using the lower bound of Theorem 1,

$$\begin{aligned}\widetilde{\kappa} &= \frac{\widetilde{\sigma}^2}{(\widetilde{\sigma} + \sqrt{P^*})^2 + 1} \overset{(P^* \leq \frac{1}{2})}{\geq} \frac{\widetilde{\sigma}^2}{(\widetilde{\sigma} + \sqrt{0.5})^2 + 1} \\ &\overset{(a)}{\geq} \frac{16}{(\sqrt{16} + \sqrt{0.5})^2 + 1} \approx 0.6909 \geq 0.69,\end{aligned}$$

where (a) uses $\widetilde{\sigma}^2 \geq 16$ and the observation that $\frac{x^2}{(x+b)^2+1} = \frac{1}{\left(1+\frac{b}{x}\right)^2 + \frac{1}{x^2}}$ is an increasing function of x for $x, b > 0$. Thus,

$$\left((\sqrt{\widetilde{\kappa}} - \sqrt{P^*})^+\right)^2 \geq ((\sqrt{0.69} - \sqrt{0.5})^+)^2 \approx 0.0153 \geq 0.015.$$

Using the upper bound of $\frac{\sigma^2}{\sigma^2+1} < 1$, the ratio of the upper and the lower bounds is smaller than $\frac{1}{0.015} < 67$.

Case 6: $N_1 < 1, \sigma^2 \geq 1, P^* < \frac{\widetilde{\sigma}^2}{100}, \widetilde{\sigma}^2 > 16, P^* \geq \frac{1}{2}$.

This is the most interesting case because nonlinear strategies are needed here to obtain constant factor results. But first we concentrate on the lower bound.

Using $L = 2$ in the first term of the lower bound,

$$\begin{aligned}c_m(L) &= \frac{1}{\Pr(\|\mathbf{Z}^m\|^2 \leq mL^2)} = \frac{1}{1 - \Pr(\|\mathbf{Z}^m\|^2 > mL^2)} \\ &\overset{(\text{Markov's ineq.})}{\leq} \frac{1}{1 - \frac{m}{mL^2}} \overset{(L=2)}{=} \frac{4}{3},\end{aligned}$$

Similarly,

$$\begin{aligned}d_m(2) &= \frac{\Pr(\|\mathbf{Z}^{m+2}\|^2 \leq mL^2)}{\Pr(\|\mathbf{Z}^m\|^2 \leq mL^2)} \\ &\geq \Pr(\|\mathbf{Z}^{m+2}\|^2 \leq mL^2) = 1 - \Pr(\|\mathbf{Z}^{m+2}\|^2 > mL^2) \\ &\overset{(\text{Markov's ineq.})}{\geq} 1 - \frac{m+2}{mL^2} = 1 - \frac{1+\frac{2}{m}}{4} \overset{(m\geq 1)}{\geq} 1 - \frac{3}{4} = \frac{1}{4}.\end{aligned}$$

In the bound, we are free to use any $\sigma_G^2 \geq 1$. Using $\sigma_G^2 = 6P^* > 1$,

$$\begin{aligned}\kappa_2 &= \frac{\sigma_G^2 \widetilde{\sigma}^2}{\left((\widetilde{\sigma} + \sqrt{P^*})^2 + d_m(2)\sigma_G^2\right) c_m^{\frac{2}{m}}(2) e^{1-d_m(2)}} \\ &\overset{(a)}{\geq} \frac{6P^* \widetilde{\sigma}^2}{\left((\widetilde{\sigma} + \frac{\widetilde{\sigma}}{10})^2 + \frac{6\widetilde{\sigma}^2}{100}\right) \left(\frac{4}{3}\right)^{\frac{2}{m}} e^{\frac{3}{4}}} \overset{(m\geq 1)}{\geq} 1.255 P^*.\end{aligned}$$

where (a) uses $\sigma_G^2 = 6P^*, P^* < \frac{\widetilde{\sigma}}{100}, c_m(2) \leq \frac{4}{3}$ and $1 > d_m(2) \geq \frac{1}{4}$. Thus,

$$\left((\sqrt{\kappa_2} - \sqrt{P^*})^+\right)^2 \geq P^*(\sqrt{1.255} - 1)^2 \geq \frac{P^*}{70}. \tag{12}$$

Now, using the first term in lower bound on the total cost from Theorem 4, and substituting $L = 2$,

$$\begin{aligned}
\bar{J}_{opt} &\geq k^2 P^* + \frac{\sigma_G^m}{c_m(2)} \exp\left(-\frac{mL^2(\sigma_G^2 - 1)}{2}\right)\left(\left(\sqrt{\kappa_2} - \sqrt{P^*}\right)^+\right)^2 \\
&\overset{(\sigma_G^2 = 6P^*)}{\geq} k^2 P^* + \frac{(6P^*)^m}{c_m(2)} \exp\left(-\frac{4m(6P^* - 1)}{2}\right) \frac{P^*}{70} \\
&\overset{(a)}{\geq} k^2 P^* + \frac{3^m}{\frac{4}{3}} e^{2m} e^{-12P^* m} \frac{1}{70 \times 2} \\
&\overset{(m \geq 1)}{\geq} k^2 P^* + \frac{3 \times 3 \times e^2}{4 \times 70 \times 2} e^{-12mP^*} \\
&> k^2 P^* + \frac{1}{9} e^{-12mP^*},
\end{aligned} \tag{13}$$

where (a) uses $c_m(2) \leq \frac{4}{3}$ and $P^* \geq \frac{1}{2}$. Using the second term in the lower bound, we obtain the following bound

$$\begin{aligned}
\bar{J}_{opt} &> \max\left\{k^2 P^* + \frac{1}{9} e^{-12mP^*}, \frac{\sigma^2 N_1}{\sigma^2 + N_1 + \sigma^2 N_1}\right\} \\
&\geq \frac{k^2 P^*}{2} + \frac{1}{18} e^{-12mP^*} + \frac{\sigma^2 N_1}{2\left(\sigma^2 + N_1 + \sigma^2 N_1\right)}.
\end{aligned} \tag{14}$$

For the upper bound, we loosen the lattice-based upper bound of Theorem 2 and massage it to obtain a form similar to (13). Here, P is a part of the optimization:

$$\begin{aligned}
&\bar{J}_{opt}(m, k^2, \sigma_0^2) \\
\leq\ & \inf_{P > \xi^2 N_{sum}} k^2 P + N_{sum}\left(1 + \sqrt{\frac{P}{\xi^2 N_{sum}}}\right)^2 e^{-\frac{mP}{2\xi^2 N_{sum}} + \frac{m+2}{2}\left(1 + \ln\left(\frac{P}{\xi^2 N_{sum}}\right)\right)} + \frac{\sigma^2 N_1}{\sigma^2 + N_1} \\
\leq\ & \inf_{P > \xi^2 N_{sum}} k^2 P + \frac{1}{9} e^{-\frac{0.5mP}{\xi^2 N_{sum}} + \frac{m+2}{2}\left(1 + \ln\left(\frac{P}{\xi^2 N_{sum}}\right)\right) + 2\ln\left(1 + \sqrt{\frac{P}{\xi^2 N_{sum}}}\right) + \ln(9)} + \frac{\sigma^2 N_1}{\sigma^2 + N_1} \\
\leq\ & \inf_{P > \xi^2 N_{sum}} k^2 P + \frac{1}{9} e^{-m\left(\frac{0.5P}{\xi^2 N_{sum}} - \frac{m+2}{2m}\left(1 + \ln\left(\frac{P}{\xi^2 N_{sum}}\right)\right) - \frac{2}{m}\ln\left(1 + \sqrt{\frac{P}{\xi^2 N_{sum}}}\right) - \frac{\ln(9)}{m}\right)} + \frac{\sigma^2 N_1}{\sigma^2 + N_1} \\
=\ & \inf_{P > \xi^2 N_{sum}} k^2 P + \frac{1}{9} e^{-\frac{0.12mP}{\xi^2 N_{sum}}} \times e^{-m\left(\frac{0.38P}{\xi^2 N_{sum}} - \frac{1 + \frac{2}{m}}{2}\left(1 + \ln\left(\frac{P}{\xi^2 N_{sum}}\right)\right) - \frac{2}{m}\ln\left(1 + \sqrt{\frac{P}{\xi^2 N_{sum}}}\right) - \frac{\ln(9)}{m}\right)} + \frac{\sigma^2 N_1}{\sigma^2 + N_1} \\
\overset{(m \geq 1)}{\leq}\ & \inf_{P > \xi^2 N_{sum}} k^2 P + \frac{1}{9} e^{-\frac{0.12mP}{\xi^2 N_{sum}}} e^{-m\left(\frac{0.38P}{\xi^2 N_{sum}} - \frac{3}{2}\left(1 + \ln\left(\frac{P}{\xi^2 N_{sum}}\right)\right) - 2\ln\left(1 + \sqrt{\frac{P}{\xi^2 N_{sum}}}\right) - \ln(9)\right)} + \frac{\sigma^2 N_1}{\sigma^2 + N_1} \\
\leq\ & \inf_{P \geq 34\xi^2 N_{sum}} k^2 P + \frac{1}{9} e^{-\frac{0.12mP}{\xi^2 N_{sum}}} + \frac{\sigma^2 N_1}{\sigma^2 + N_1},
\end{aligned} \tag{15}$$

where the last inequality follows from the fact that $\frac{0.38P}{\xi^2 N_{sum}} > \frac{3}{2}\left(1 + \ln\left(\frac{P}{\xi^2 N_{sum}}\right)\right) + 2\ln\left(1 + \sqrt{\frac{P}{\xi^2 N_{sum}}}\right) + \ln(9)$ for $\frac{P}{\xi^2 N_{sum}} > 34$. This can be checked easily by plotting it.[3]

Now notice that for $N_1 < 1$, $N_{sum} = \widetilde{N}_1 + 1 < 2$, because $\widetilde{N}_1 = \frac{\sigma^2 N_1}{\sigma^2 + N_1} \leq 1$ for $N_1 < 1$. Thus,

$$\bar{J}_{opt} \leq \inf_{P \geq 68\xi^2} k^2 P + \frac{1}{9} e^{-\frac{0.12mP}{2\xi^2}} + \frac{\sigma^2 N_1}{\sigma^2 + N_1}.$$

Using $P = 200\xi^2 P^* \geq 100\xi^2$ (since $P^* \geq \frac{1}{2}$), which is larger than $68\xi^2$. Thus, we obtain from (15)

$$\bar{J}_{opt} \leq k^2 200\xi^2 P^* + \frac{1}{9} e^{-\frac{12mP}{\xi^2}} + \frac{\sigma^2 N_1}{\sigma^2 + N_1}. \tag{16}$$

[3]It can also be verified symbolically by examining the expression $g(b) = 0.38b^2 - \frac{3}{2}(1 + \ln b^2) - 2\ln(1 + b) - \ln(9)$, taking its derivative $g'(b) = 0.76b - \frac{3}{b} - \frac{2}{1+b}$, and second derivative $g''(b) = 0.76 + \frac{3}{b^2} + \frac{2}{(1+b)^2} > 0$. Thus $g(\cdot)$ is convex-$\cup$. Further, $g'(\sqrt{34}) \approx 3.62 > 0$, and $g(\sqrt{34}) \approx 0.09$ and so $g(b) > 0$ whenever $b \geq \sqrt{34}$.

Now notice that $\frac{\sigma^2 N_1}{\sigma^2+N_1} \times \frac{\sigma^2 N_1+\sigma^2+N_1}{\sigma^2 N_1} = \frac{\sigma^2 N_1}{\sigma^2+N_1} + 1 < 2$ for $N_1 < 1$. Thus, comparing (14) and (16), the ratio of the upper and lower bounds is smaller than $400\xi^2$. ■