Interference suppression in the presence of quantization errors

Omar Bakr Mark Johnson Raghuraman Mudumbai Upamanyu Madhow

Electrical Engineering and Computer Sciences, UC Berkeley, Berkeley, CA 94720
Electrical and Computer Engineering, UC Santa Barbara, Santa Barbara, CA 93106
Email:{ombakr, mjohnson}@eecs.berkeley.edu, {raghu, madhow}@ece.ucsb.edu

Abstract

Multi-element antennas offer the possibility of increasing the spatial reuse of wireless spectrum by "nulling" out interfering signals. However, the interference suppression performance is highly sensitive to small errors in the gains applied to the antenna elements. In this paper, we examine in detail the effect of one specific source of error that arises from quantizing array weights. We show that a simple approach based on scalar quantization that ignores the correlation of the quantization errors fails to fully utilize the interference suppression capability of the array: the residual interference level does not decrease with the number of antennas. Unfortunately, the optimum approach to computing the weights involves vector quantization over a space that grows exponentially with the number of antennas and number of quantization bits, and is therefore computationally intractable. We propose instead a simple suboptimal method that greedily optimizes SIR coefficient by coefficient. Simulations show that this greedy approach provides substantial SIR gains over the naive approach, with SIR growing polynomially in the number of antennas. We derive analytical bounds that indicate that even larger SIR gains (exponential in the number of antennas) are potentially achievable, so that finding tractable algorithms that improve upon our suboptimal approach is an important open problem.

I. Introduction

The use of multiple antennas offers the potential of significantly improving the performance and capacity of interference-limited wireless communication systems, with applications including spatial multiplexing over a point-to-point link, multipacket reception, and uncoordinated spectrum sharing among different networks. Standard interference suppression techniques include beamforming using the zero-forcing (ZF) or linear minimum mean-squared error (MMSE) criterion. In modern "mostly digital" transceiver architectures, beamforming is typically implemented digitally, using quantized weight vectors. In this paper, we examine the effect of such coefficient quantization on interference suppression performance, specifically focusing on the performance of quantized versions of a ZF receive beamformer which steers nulls in the direction of the interfering signals. Since the MMSE beamformer tends to the ZF beamformer at high signal-to-noise ratio (SNR), we expect our results to also apply, at least qualitatively, to the MMSE beamformer in regimes where interference, rather than noise, is the dominant impairment.

We find that quantization of the beamforming weights can dramatically degrade interference suppression performance if performed in naive fashion, which could severely limit our ability to exploit the large spatial multiplexing gains that are theoretically available with multiple antennas. Specifically, while the contribution of the desired signal to the beamformer output is highly robust to quantization error in the beamforming weights, far more precise control of the weights is required to effectively steer nulls. Indeed, it is clear that, with probability one, perfect nulling of even a single interferer is no longer possible. What we would still hope for, however, is that the Signal to Interference Ratio (SIR) scales up as quickly as possible with N, the number of antennas, and degrades as slowly as possible with K, the number of interferers.

Our main results are summarized as follows:

1) A naive strategy of coefficient-by-coefficient quantization of the ZF beamforming weights has poor performance: the SINR scales as N/K, as it would for a spatial matched filter. That is, we obtain the combining gains for the desired signal, but increasing the number of degrees of freedom N does not lead to any improved scaling in the residual interference power at the output of the beamformer.

2) We provide analytical estimates that indicate that we should be able to perform much better than the naive strategy. An upper bound implies that the SINR could potentially grow exponentially fast with N, while a pessimistic estimate indicates that the SINR should grow at least as fast as N^2. This motivates the search for improvements to the naive "scalar quantization" strategy.
3) We propose a suboptimal "vector quantization" strategy in which each coefficient is optimized sequentially to maximize SINR, starting from the naive solution. This greedy sequential strategy yields substantial gains in SINR over the naive strategy, demonstrating the polynomial growth of the kind predicted by the pessimistic analytical estimates.
4) We provide extensive simulation results for various strategies. While optimal vector quantization of the beamforming weights is computationally infeasible for large N, exhaustive search for small N is found to perform significantly better than the greedy sequential strategy, possibly even leading to exponential dependence on N. Finding computationally efficient vector quantization algorithms that can approach the performance of the optimal solution therefore remains an important open problem.

Related Work. To the best of our knowledge, this paper is the first systematic examination of the effect of quantization of antenna weights on interference suppression. However, quantization in MIMO wireless systems has been extensively studied in many contexts in previous work. The most notable example of this is the work on the design of codebooks in channels with limited channel-state feedback [1], [2], where the goal is to find transmit beamforming vectors that maximize the *desired* signal contribution at the receiver. This problem has been shown to be related to the problem of sphere-packing in a Grassmanian manifold [3], and the geometry of vectors on the multi-dimensional hypersphere proves to be important to understanding the codebook design problem [4]. While the geometry of the hypersphere also provides insights into interference suppression and gives us an analytical upper-bound on the interference suppression levels, it turns out that the problem of finding vectors *orthogonal* to a given vector in the hypersphere is more involved than the sphere-packing problem. We briefly remark on some open issues related to this in Section VI.

Outline: In Section II, we describe a simple interference suppression problem and show that naive scalar quantization incurs a severe performance penalty. Section III contains analytical estimates of the performance that can be achieved with vector quantization, with optimistic estimates predicting exponential gains of SIR with N and pessimistic estimates predicting polynomial gains. In Section IV, we describe an algorithm that can significantly improve the performance using a greedy sequential vector quantization strategy. Extensive simulation results are provided in Section V, while Section VI contains our conclusions.

II. THE INTERFERENCE SUPPRESSION PROBLEM

We begin by considering a multi-antenna receive beamforming system, where the goal is to maximize the array gain in the direction of a desired transmitter[1]. In this section, we analyze the effect of errors in the choice of antenna weights. For clarity we focus on a simple setup where a receiver with a N element antenna array receives the desired signal that has the (complex baseband) channel gain $\mathbf{h}_0 \doteq \left[h_0[1],\ h_0[2],\ \ldots\ h_0[N]\right]$ and an interfering signal with channel gain $\mathbf{h}_1 \doteq \left[h_1[1],\ h_1[2],\ \ldots\ h_1[N]\right]$[2], where the components of the channels $\mathbf{h}_0,\ \mathbf{h}_1$ are drawn from iid complex Gaussian distributions i.e. Rayleigh fading. This setup can be readily generalized to multiple interferers and to Rician or Line-of-Sight channels and we later present some simulation results to show that the same ideas extend to more general setups. Indeed, we later present arguments that show that a Rayleigh fading channel represents a worst-case scenario in dealing with quantized weights.

A. Problem Statement

We begin by presenting a quick review of the theory behind computations of the optimum beamforming weights. The incoming signal at the input of the array $\mathbf{y}[n]$ is the sum of the desired signal and interference and noise:

$$\mathbf{y}[n] = \mathbf{h}_0 d[n] + \mathbf{h}_1 d_1[n] + \mathbf{v}[n]$$

[1] A similar analysis also applies to a transmit array that seeks to minimize its interference at given locations. We focus on the receive array in this paper for clarity.

[2] We will denote scalars in lower case, vectors in bold lower case, and matrices in bold upper case.

where $d[n]$ is the desired signal, $d_1[n]$ is interfering signal, and $\mathbf{v}[n]$ is the white noise vector at the receiver with variance σ_v^2). For simplicity, we shall assume that the desired and interfering signals have the same power i.e. are drawn from Rayleigh distributions with the same variance which we set to unity without loss of generality.

Using array weights $\mathbf{w},\ ||\mathbf{w}|| = 1$, the signal at the output of the array will be $\mathbf{w}^H\mathbf{y}[n]$ and the resulting output SINR is given by:

$$\text{SINR}_{out} = \frac{|\mathbf{w}^H\mathbf{h}_0|^2}{|\mathbf{w}^H\mathbf{h}_1|^2 + \sigma_v^2}$$

where $(\cdot)^H$ denotes the complex conjugate transpose. In the absence of interference, the output signal to noise ratio (SNR) is maximized by choosing

$$\mathbf{w}_{opt} = \frac{\mathbf{h}_0}{||\mathbf{h}_0||},\ \text{SNR}_{opt} = \frac{||\mathbf{h}_0||^2}{\sigma_v^2} \tag{1}$$

Note that when averaged over the fading channel gain $\mathbf{h}_0$, the SNR varies as

$$\text{E}\big[\text{SNR}_{opt}\big] = \frac{\text{E}[||\mathbf{h}_0||^2]}{\sigma_v^2} \equiv \frac{N}{\sigma_v^2} \tag{2}$$

In other words, the SNR increases linearly with the number of antennas N. Physically a larger array permits higher antenna directivities which accounts for this increase. Note that the "effective noise power" $|\mathbf{w}^H\mathbf{v}|^2 \equiv \sigma_v^2$ does not increase with N.

When an interferer is present, complete interference rejection can be achieved by choosing a beamforming weight vector $\mathbf{w}$ that is the projection of the desired vector $\mathbf{h}_0$ onto the subspace orthogonal to the interference vector $\mathbf{h}_1$:

$$\mathbf{w}_{opt} = \mathbf{w}_{projection} = \mathbf{h}_0 - \frac{1}{\mathbf{h}_1^H\mathbf{h}_1}\mathbf{h}_1\mathbf{h}_1^H\mathbf{h}_0 \tag{3}$$

If we let $\theta_1 \doteq \cos^{-1}\left(\frac{|\mathbf{h}_1^H\mathbf{h}_0|}{||\mathbf{h}_0||||\mathbf{h}_1||}\right)$ be the "angle" between the desired and interference channel vectors[3], we can write the SNR from the projection beamformer as

$$\text{SNR}_{opt} = \frac{||\mathbf{h}_0||^2\sin^2\theta_1}{\sigma_v^2} \tag{4}$$

When averaged over the fading channel gains $\mathbf{h}_0,\ \mathbf{h}_1$, we still get a linear increase in SNR with N:

$$\text{E}\big[\text{SNR}_{opt}\big] = \frac{\text{E}[||\mathbf{h}_0||^2\sin^2\theta_1]}{\sigma_v^2} \geq \frac{N}{2\sigma_v^2} \tag{5}$$

The final inequality in (5) is based on the observation that $\text{E}\big[\sin^2\theta_1\big] \geq \frac{1}{2}$, which is derived in Equation (27) in Appendix I.

The projection-based beamformer (3) does not take noise into account. In general, maximizing the output SINR does not necessarily require complete interference rejection; reducing the interference to the noise level may be sufficient. Optimizing the output SINR leads to the Minimum Variance Distortionless Response (MVDR) beamformer [5]. If we define the noise+interference correlation matrix $\mathbf{R}_{N+I} \doteq \mathbf{h}_1\mathbf{h}_1^H + \sigma_\nu^2 I_N$ where I_N is the $N \times N$ identity matrix, then the output SINR can be maximized by choosing $\mathbf{w}_{opt}$:

$$\mathbf{w}_{opt} = \mathbf{w}_{\text{MVDR}} = \frac{\mathbf{R}_{N+I}^{-1}\mathbf{h}_0}{\mathbf{h}_0^H\mathbf{R}_{N+I}^{-1}\mathbf{h}_0} \tag{6}$$

The denominator in (6) is a normalizing factor. When the interference power is much larger than the noise power, both projection and MVDR yield virtually identical results. This is typically the case in bandwidth-limited wireless links and we focus on this case exclusively in this paper; therefore the resulting SINR is still approximately given by (4) and its variation with N by (5).

[3]Note that θ_1 is an angle in N-dimensional space, and does not correspond to a physical direction.

B. Effect of errors in weights

In practice the weights that are actually applied to the array elements may differ from the optimum desired weight from (3) for many reasons such as errors in channel estimation, calibration errors and quantization effects. Let $\hat{\mathbf{w}}$ be the weight applied to the array. We will assume that $\hat{\mathbf{w}} \equiv \frac{(\mathbf{w}_{opt}+\mathbf{\Delta w})}{||(\mathbf{w}_{opt}+\mathbf{\Delta w})||}$ where the real and imaginary parts (in-phase and quadrature) of each component of the weight error vector $\mathbf{\Delta w}$ are $i.i.d.$ zero mean random variables with variance $\frac{1}{2N}\sigma_w^2$.

Remark. Since any scaled version of the vector $\mathbf{w}$ yields the same SINR, we constrain $\mathbf{w}$ to be a unit vector i.e. $||\mathbf{w}|| = 1$ in order to eliminate this ambiguity. In practice, the weight applied to each antenna element usually varies between fixed limits e.g. $(-1, 1)$, and this weight is quantized to a fixed number of bits. Thus the variance of each component of the quantization error vector $\mathbf{\Delta w}$ is also fixed. However, because of the scale invariance of the SINR, we choose to impose the unit vector constraint for analytical convenience (instead of the fixed dynamic range constraint). As a result, the components of $\mathbf{w}$ scale as $\frac{1}{\sqrt{N}}$, and the variance of the quantization noise also scales inversely with N since quantization error varies proportionally with the dynamic range of the elements of $\mathbf{w}$ itself. In other words, $\mathrm{E}\big[||\mathbf{\Delta w}||^2\big] = \sigma_w^2$ independent of N.

Intuitively, the errors $\mathbf{\Delta w}$ result in $\hat{\mathbf{w}}$ deviating from $\mathbf{w}_{opt}$ by an "angle" $\theta_w \doteq \cos^{-1}\big(|\mathbf{w}_{opt}^H\hat{\mathbf{w}}|\big)$. This deviation will result in a reduction in the signal strength in the desired direction as well as an increase in the interference power, since $\hat{\mathbf{w}}$ will no longer be orthogonal the interference subspace. The desired power is proportional to $\cos\theta_w$, and the increase in interference (leakage) is proportional to $\sin\theta_w$ (see Figure 1). For small angles θ_w, we can use the standard approximations $\sin\theta_w \approx \theta_w$ and $\cos\theta_w \approx 1$. This explains why nulls are more sensitive than peaks to phase and amplitude errors, since $\sin\theta$ changes more rapidly than $\cos\theta$ when θ is small.

Thus taking the errors $\mathbf{\Delta w}$ into account, we can update (1) for the SNR in the absence of interference as

$$\begin{aligned} \mathrm{SNR} &= \frac{|\hat{\mathbf{w}}^H\mathbf{h}_0|^2}{||\mathbf{w}||^2\sigma_v^2} \\ &= \frac{|\mathbf{w}_{opt}^H\mathbf{h}_0 + \mathbf{\Delta w}^H\mathbf{h}_0|^2}{\sigma_v^2} \\ &\geq \mathrm{SNR}_{opt}(1-||\mathbf{\Delta w}||)^2 \end{aligned} \tag{7}$$

This implies that the SNR averaged over the fading still scales as N:

$$\mathrm{E}\big[\mathrm{SNR}\big] \geq \alpha \times \mathrm{E}\big[\mathrm{SNR}_{opt}\big] \tag{8}$$

with at most only a constant factor loss $\alpha = E\big[(1-||\mathbf{\Delta w}||)^2\big] \geq (1-\sigma_w^2)$ due to the errors.

Now, let us consider the effect of the weight errors $\mathbf{\Delta w}$ on the zero-forcing beamformer in the presence of an interferer. The SINR is given by

$$\begin{aligned} \mathrm{SINR} &= \frac{|\hat{\mathbf{w}}^H\mathbf{h}_0|^2}{|\hat{\mathbf{w}}^H\mathbf{h}_1|^2 + ||\hat{\mathbf{w}}||^2\sigma_v^2} \\ &= \frac{\big|||\mathbf{h}_0||\sin\theta_1 + \mathbf{\Delta w}^H\mathbf{h}_0\big|^2}{|\mathbf{\Delta w}^H\mathbf{h}_1|^2 + (1+||\mathbf{\Delta w}||^2)\sigma_v^2} \end{aligned} \tag{9}$$

where θ_1 is defined as in (4). The numerator of (9) when averaged over the weight errors $\mathbf{\Delta w}$ and the fading coefficients $\mathbf{h}_0$, $\mathbf{h}_1$ is lower-bounded by $\frac{N}{2}+\sigma_w^2$ which scales linearly with N, and the averaged denominator equals $\sigma_w^2 + (1+\sigma_w^2)\sigma_v^2$, which is independent of N.

Thus the ratio of average signal to interference and noise powers scales linearly with N. This seems good considering that the SNR in the no-interferer case given by (8) also scales linearly with N. Note however that the average interference power given by the denominator of (9) does not decrease with N even though the degrees of freedom for interference cancellation increases; this is starkly unlike the situation without weight errors, where a larger number of antennas permits the complete nulling of a larger number of interferers. This indicates that there might be some suboptimality in the scaling behavior in (9). We show next that indeed, a much better scaling behavior can be obtained by adopting a more optimal approach.

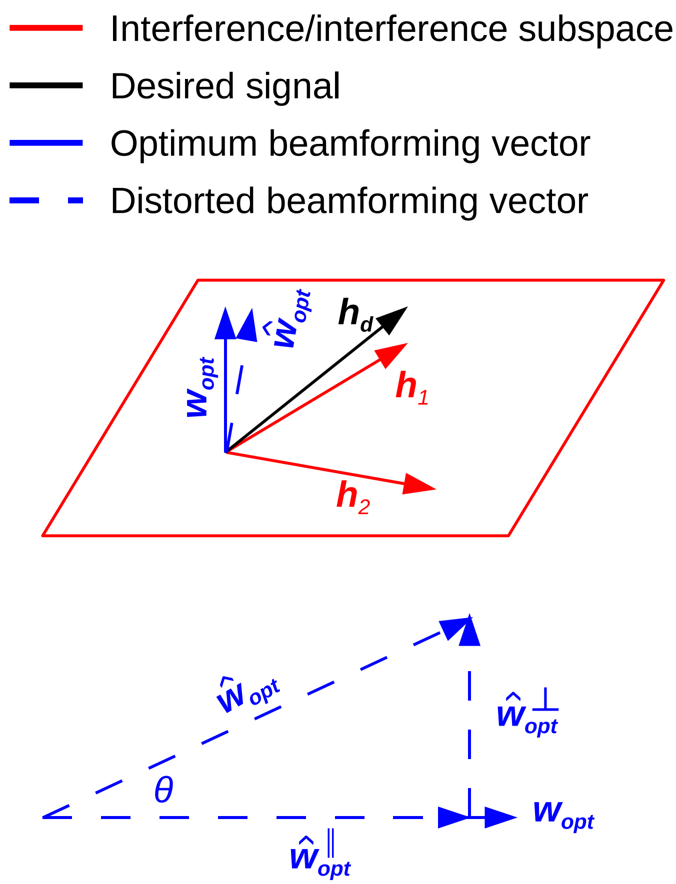

Fig. 1. The optimum beamforming vector $\mathbf{w}_{opt}$ can be viewed as a projection of the desired signal onto the subspace orthogonal to the interference subspace. The distorted beamforming vector $\hat{\mathbf{w}}_{opt}$ can decomposed into two orthogonal components: $\hat{\mathbf{w}}_{opt} = \hat{\mathbf{w}}_{opt}^{\perp} + \hat{\mathbf{w}}_{opt}^{\parallel}$. $\hat{\mathbf{w}}_{opt}^{\parallel}$, which is parallel to $\mathbf{w}_{opt}$, represents the potential loss in beamforming gain, and is proportional to $\cos\theta$. $\hat{\mathbf{w}}_{opt}^{\perp}$, which is orthogonal to $\mathbf{w}_{opt}$, represents the potential leakage into the interference subspace, and is proportional to $\sin\theta$.

III. Improving Interference Suppression

We now consider the problem of computing the optimal weights in the presence of quantization errors. Let us represent the i'th component of vector $\mathbf{x}$ by $x[i]$, e.g. the weight applied to antenna element i is $w[i]$ and so on. Let B be the number of quantization bits i.e. the I and Q components of each weight element is quantized to one of 2^B levels. There are then $N_w = 2^{2BN}$ total weights available from which to choose[4]. Let $\mathcal{W}$ be the set of these weight vectors. We can then formulate the problem of computing the optimum weight as a SINR maximization problem:

$$\mathbf{w}_{opt} \doteq \arg\max_{\mathbf{w}\in\mathcal{W}} \frac{|\hat{\mathbf{w}}^H \mathbf{h}_0|^2}{|\hat{\mathbf{w}}^H \mathbf{h}_1|^2 + ||\hat{\mathbf{w}}||^2 \sigma_v^2} \tag{10}$$

Remark. It is not a trivial task to compare two algorithms based on their SINR performance. If one algorithm outperforms another for every possible realization of the desired and interference channels, then it is clearly better. However, such a definition is too stringent to be useful, and in practice, we may have to settle for comparing some averages. At first glance, it seems as if the natural choice would be to compare the average SINR i.e. $\mathrm{E}[\mathrm{SINR}]$; however this measure suffers from some fundamental disadvantages. First of all, it is rather unwieldy analytically. Secondly, the average can be dominated by a small number of channel realizations where the SINR becomes very

[4]Note that some of these weight vectors may be scaled versions of others, and therefore N_w is actually an upper-bound on the number of distinct weight vectors available to choose from.

large. We, therefore choose a different measure: the ratio of average signal power to average interference power (we neglect noise):

$$\bar{\text{SINR}} = \frac{\text{E}[|\hat{\mathbf{w}}^H \mathbf{h}_0|^2]}{\text{E}[|\hat{\mathbf{w}}^H \mathbf{h}_1|^2]} \tag{11}$$

A. An approximate lower bound on achievable SIR

The problem of choosing the optimum weight $\mathbf{w}$ can also be considered as a problem of choosing the optimal $\mathbf{\Delta w}$. We have seen from (9) that the average signal power scales linearly with N in the presence of the errors $\mathbf{\Delta w}$. However, as we have seen, the average interference power is independent of N if the components of $\mathbf{\Delta w}$ are chosen independently. We show next that if the components of $\mathbf{\Delta w}$ are chosen *dependently* of each other, the interference power can be made to decrease significantly with N.

Given an interference vector $\mathbf{h}_1$, our goal is to understand the term $|\mathbf{\Delta w}^H \mathbf{h}_1|^2$. Let us first consider a simpler problem. For an interference vector $\mathbf{h}_1$, suppose we simply wish to minimize $|\mathbf{u}^H \mathbf{h}_1|^2$, where $\mathbf{u}$ takes all possible values in $\mathcal{U} = \frac{1}{\sqrt{N}}\{-1,+1\}^N$ (i.e. the elements of $\mathcal{U}$ are binary vectors normalized to unit norm). Let $Z(\mathbf{u}) \doteq \mathbf{u}^H \mathbf{h}_1$. Note that $Z(\mathbf{u}) \sim CN(0,1)$ for any $\mathbf{u} \in \mathcal{U}$ since the fading coefficients of $\mathbf{h}_1$ are assumed to be $CN(0,1)$. Note also that $Z(\mathbf{u}_1)$, $Z(\mathbf{u}_2)$ are jointly proper complex Gaussian with covariance (and normalized correlation) equal to $\mathbf{u}_1^H \mathbf{u}_2$. Let us now choose a subset $\mathcal{U}_o$ of $\mathcal{U}$ which forms an orthonormal basis in N dimensions: the Walsh-Hadamard codes. Then $\{Z(\mathbf{u}) : \mathbf{u} \in \mathcal{U}_o\}$ are N uncorrelated, and hence independent, $CN(0,1)$ random variables. The corresponding powers $\{|Z(\mathbf{u})|^2, \mathbf{u} \in \mathcal{U}_o\}$ are therefore exponential random variables with mean one.

The minimum interference power can now be upper bounded as

$$P_{min} = \min_{\mathbf{u} \in \mathcal{U}} |Z(\mathbf{u})|^2 \leq \min_{\mathbf{u} \in \mathcal{U}_o} |Z(\mathbf{u})|^2 \tag{12}$$

It is easy to show that the minimum of N i.i.d. exponential random variables of mean one is an exponential random variable with mean $\frac{1}{N}$. Thus, for our simplified model, the interference power scales down at least as fast as $\frac{1}{N}$ with N.

Let us now return to our original problem. Consider $Z(\mathbf{\Delta w}) \equiv \mathbf{\Delta w}^H \mathbf{h}_1$. For each coefficient of the ZF weight vector, suppose that we restrict ourselves to two options: round both I and Q coefficients up or round them both down. Thus, if the quantization interval is Δ, the corresponding coefficient of $\mathbf{\Delta w}$ is being set to $-(\Delta - X)$ or X, respectively, where X is the distance of the unquantized coefficient from the bottom-left edge of the quantization bin in which it falls. Thus, if we let +1 indicate rounding down a coefficient, and -1 indicate rounding up, there are 2^N possible choices of $\mathbf{\Delta w}$ with this strategy, which map to the vectors in $\mathcal{U} = \frac{1}{\sqrt{N}}\{-1,+1\}^N$ considered in our prior simplified example. While the Walsh-Hadamard vectors $\mathbf{u}_i$ no longer yield uncorrelated Gaussian random variables $Z(\mathbf{b}_i)$ when the distribution of the X's are taken into account, since $\mathcal{U}$ is an N-dimensional vector space, it is always possible to find a set of basis vectors $\mathbf{v}_i$ which yield uncorrelated random variables $Z(\mathbf{v}_i)$ and for which (12) holds. This leads us to expect that the minimum interference power here will also fall off at least as fast as $\frac{1}{N}$. We plan to continue to work on refining these arguments, and hope to present a sharper result at the conference.

B. A geometric upper-bound on achievable SIR

Consider the $2N$-dimensional hypersphere generated by the I and Q coefficients of the unit vector $\mathbf{x}$ corresponding to the weight errors i.e. $\mathbf{x} \doteq \frac{1}{||\mathbf{\Delta w}||}\left[\Re(\mathbf{\Delta w}[1]),\ \Im(\mathbf{\Delta w}[1]),\ \Re(\mathbf{\Delta w}[2]),\ \Im(\mathbf{\Delta w}[2]),\ \cdots\right]$. We want to choose $\mathbf{x}$ so as to minimize the interference power given by $|\mathbf{\Delta w}^H \mathbf{h}_1|^2 \equiv ||\mathbf{\Delta w}||^2 |\mathbf{x}^H \mathbf{h}_1|^2$, from the set $\mathcal{X}$ of available vectors $\mathbf{x}$:

$$\mathbf{x}_{opt} = \arg\min_{\mathbf{x} \in \mathcal{X}} |\mathbf{x}^H \mathbf{h}_1|^2 \tag{13}$$

We can also obtain an upper-bound on the achievable SIR by assuming that the vector $\mathbf{x}$ that minimizes the interference power also simultaneously achieves the maximum possible signal power.

Intuitively, when the number of antennas increases, we expect the minimum interference power as given in (13) to decrease for two reasons. First, as the dimensionality of the vector space of the weight vectors increases, a randomly chosen vector is "more orthogonal" to the interference vector (this is explained more precisely in

Appendix I). Second, the number of quantized weight vectors to choose from increases exponentially with N. We now quantify these factors and obtain an upper bound on the interference suppression capability of an array.

Without loss of generality, we assume that the interference vector $\mathbf{h}_1$ is aligned with one of the coordinate axes. Let θ_x be the angle between $\mathbf{h}_1$ and $\mathbf{x}$ and let $\phi_x \doteq \frac{\pi}{2} - \theta_x$. Our goal is to find $\mathbf{x}$ such that ϕ_x is as small as possible. If $\mathbf{x}$ is chosen randomly and uniformly on the unit hypersphere, the probability density function of ϕ_x is given by (see Appendix I):

$$f_\phi(\phi_x) = \frac{\cos^{2N-2}\phi_x}{2\ {}_2F_1\left(\frac{1}{2}, \frac{3-2N}{2}; \frac{3}{2}; 1\right)} \tag{14}$$

where ${}_2F_1(.,.;.;.)$ is the Gauss hypergeometric function[6]. Consider the cumulative distribution function corresponding to (14) given by

$$F_\phi(\phi) = \Pr(|\phi_x| \leq \phi) \equiv \int_{-\phi}^{\phi} f_\phi(\phi_x) d\phi_x \tag{15}$$

From (23) we have that $F_\phi(\phi)$ is an increasing function of N for any $\phi \in [0, \frac{\pi}{2}]$. Thus we have

$$F_\phi(\phi) \geq F_0(\phi) \equiv \frac{2\phi}{\pi} \tag{16}$$

Finally if all $N_w \equiv 2^{2BN}$ available weight vectors are assumed to be distributed uniformly over the surface of the unit hyperphere, there exists at least one vector $\mathbf{x}$ such that $F_\phi(\phi_x) \leq \frac{1}{N_w}$. Therefore using (16) we conclude that at least one weight vector $\mathbf{x}$ exists such that

$$\phi_x \leq \frac{\pi}{2N_w} \equiv \frac{\pi}{2^{1+2BN}} \tag{17}$$

and the average interference power corresponding to this vector $\mathbf{x}$ is upper-bounded by $\mathrm{E}\left[||\mathbf{h}_1||^2||\mathbf{\Delta w}||^2 \sin^2\phi_x\right] \leq \sigma_w^2 2^{-2-4BN}$ which decreases exponentially with N.

The key assumption in the above derivation is that the available weight vectors $\mathbf{x}$ are distributed uniformly over the hypersphere with respect to any arbitrary interference vector $\mathbf{h}_1$. This assumption is too optimistic; a more realistic assumption is that the weight vectors are uniformly distributed over a *hypercube*. Intuitively there are very few "sparse" weight vectors in $\mathcal{X}$, whereas most available vectors have significant (non-zero) coefficients over a large proportion of antenna elements. Thus for instance an interference vector that is highly sparse is difficult to suppress. To take an extreme example, an interference vector that is zero everywhere except a single antenna element will be difficult to suppress. Therefore it is not clear how tight the bound in (17) is. Also the above reasoning leads us to expect that the average interference suppression performance will depend strongly on the fading distribution e.g. the Rayleigh distribution is more likely to give "unbalanced" or sparse vectors than a line-of-sight distribution. We leave a more detailed exploration of these ideas to future work.

IV. Constructive algorithms for interference suppression

We have seen that the naive approach to quantizing the antenna weight vector, based on applying a scalar quantizer to each coefficient independently, does not achieve the optimum interference suppression. On the other hand, because there are a total of 2^{2BN} possible quantized vectors, implementing an optimal vector quantizer by exhaustively searching through the set of reconstruction levels has a computational cost that is exponential in the number of antennas and the number of quantization bits. Thus, vector quantization by exhaustive search is infeasible in practical scenarios.

We propose instead a simple, sub-optimal vector quantization scheme which is based on coordinate descent optimization. This scheme substantially improves over the naive method, yet has computational cost that is linear in the number of antennas.

The sub-optimal scheme greedily quantizes the weight vector by searching through the set of reconstruction levels for each element individually, instead of jointly as in the exhaustive search algorithm. We begin by computing an initial weight vector. This can be done in several ways, such as applying a scalar quantizer to the optimal MVDR weights given by Equation 6, applying a scalar quantizer to the desired channel response $\mathbf{h}_0$ (matched filter), or even randomly drawing a vector from the set of valid quantized weights. Our simulations show that initializing with a quantized version of the optimal weights or the matched filter yield far superior results to using a random initialization. In most of our simulations, we used the matched filter.

Next, we search through all 2^{2B} valid reconstruction levels of the first coefficient, while keeping the other $N-1$ coefficients fixed. The value which maximizes the output SINR is chosen as the quantized value of the first coefficient. We then proceed to quantize the second coefficient, keeping the first coefficient fixed at its quantized value and coefficients 3 through N fixed at their initial values, by a similar search. This method is applied to each of the N coefficients of $\mathbf{w}$, yielding a sub-optimal, vector-quantized weight vector.

The coordinate descent quantizer has a computational cost of $N \cdot 2^{2B}$. We note that the output of this quantizer depends on both the method of selecting the initial vector and the order in which the coefficients are quantized.

V. Simulation Results

In this section, we provide numerical results that verify the analysis of this paper, and compare the performance of the various quantization schemes.

A. General errors in antenna weights

Figure 2(a) demonstrates the relationship between interference power after beam-nulling and the mean square error in the antenna weights. The plot compares the case where independent Gaussian noise is added to the real and imaginary components of the optimal weight vector to the case where the noise is added to the magnitude and phase, as well as comparing the Rayleigh and LOS channels. In this simulation there were $N = 200$ antennas and $K = 20$ interferers. The results show that the interference leakage is directly proportional to the total error power, and does not depend on whether the error is modeled as additive in the Cartesian or polar coordinates. The channel has very little effect on the performance.

B. Scalar quantization of antenna weights

In this section, we present numerical results of the naive, scalar quantization scheme. With scalar quantization, the variance of the quantization noise grows as $\sigma_w^2 \sim 2^{-2B}$ and therefore the SIR is proportional to $\frac{N}{K}2^{2B}$. Figure 2(b) shows the SIR (normalized by the number of antennas, N) as a function of N, with the number of interferers fixed at $K = 20$ and the quantizer size fixed at $B = 4$ bits. Figure 2(c) shows the SIR (multiplied by K) as a function of K, with $N = 1000$ and $B = 4$. These plots demonstrate that the SIR is linearly proportional to N/K for the scalar quantization method. The slight downward slope in 2(c) is due to the fact that for a fixed number of antennas, as the dimension of the interference subspace increases more of the desired signal lies in that subspace. Hence, the signal power that is orthogonal to the interference begins to decrease. Figure 2(d) shows the SIR as a function of B, with $N = 1000$ and $K = 20$. The slope of the curve closely matches the predicted gain of 6dB for each quantization bit.

C. Vector quantization of antenna weights

We also simulated the coordinate descent vector quantizer, to quantify the performance and demonstrate the gains over scalar quantization. Figure 3(a) shows the SIR as a function of the number of antennas, for various channels and quantizer bit rates,with $K = 15$ interferers. The quantizer was initialized by applying a scalar quantizer to the matched filter. In contrast to the simulated performance of the scalar quantizer in Figure 2(b) where SIR$/N$ was constant, we see that for the coordinate descent vector quantizer SIR$/N$ grows approximately linearly. Thus, the SIR grows quadratically with the number of antennas. Figure 3(b) shows that the average interference power (scaled by the number of antennas) is constant in the same simulation. Thus, the average interference power with vector quantization decreases linearly with N. This is in contrast to the scalar quantizer, where the interference power was constant. Since the SIR is increasing quadratically in N and the interference is decreasing linearly, we know that the signal strength is growing linearly. Thus, the coordinate descent method does not cause a significant reduction in SNR.

Figure 3(c) shows the quantity $K \cdot$ SIR as a function of the number of interferers K, with $N = 100$ antennas and a $B = 2$ bit quantizer. The downward slope of this curve demonstrates that the SIR is decreasing at a rate faster than $1/K$. Figure 3(d) shows a plot of $K^2 \cdot$ SIR for the same setting, which shows that the SIR is decreasing approximately by a factor of $1/K^2$. Taken together, Figures 3(a) and (d) imply that for the coordinate descent

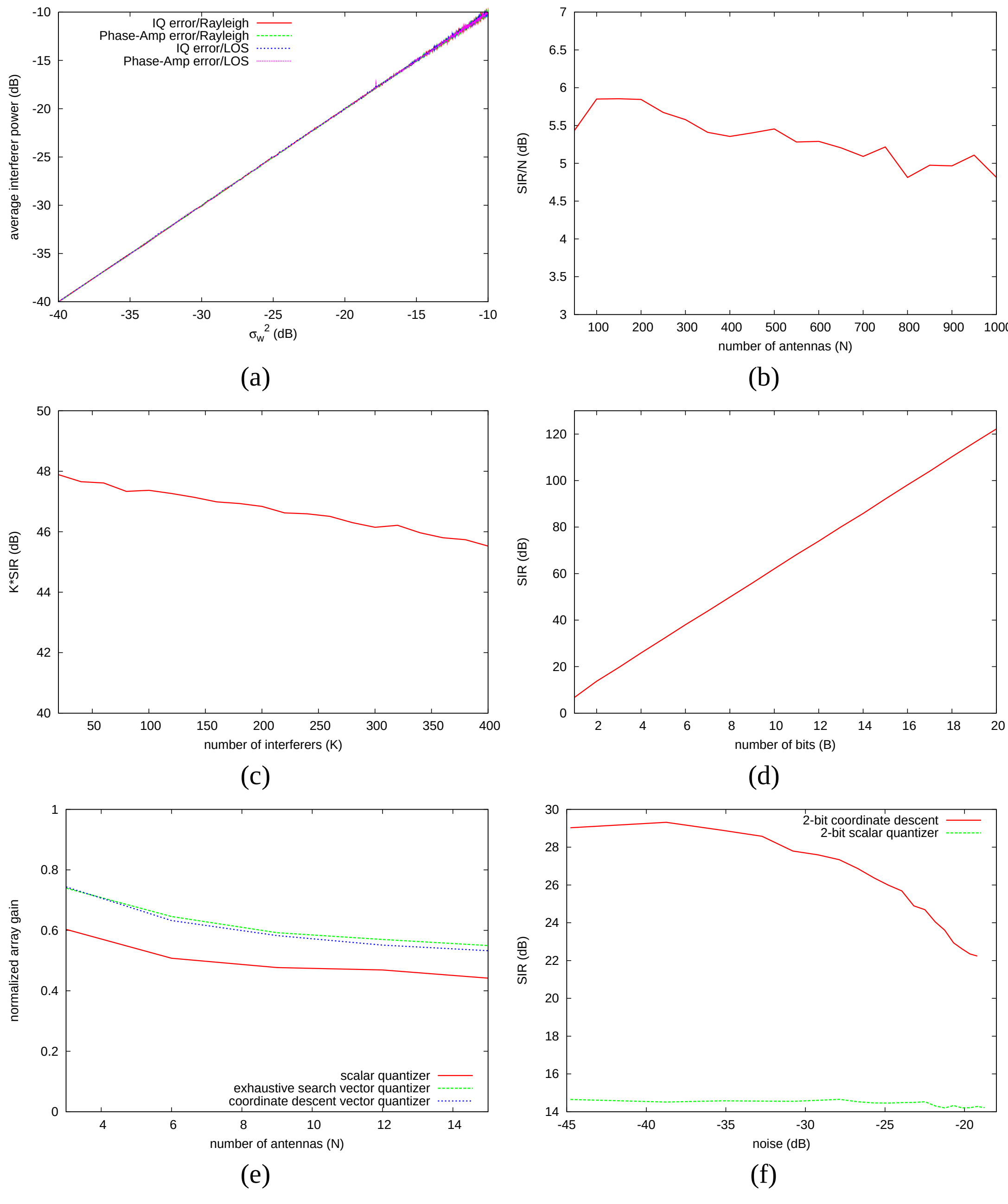

Fig. 2. (a) Simulated interference power as a function of errors in the weight vector. (b)-(d) Simulated effect of scalar quantization on the performance of the MVDR beamnulling scheme as a function of the number of antennas N, the number of interferers K, and the number of quantization bits B. (d) Comparison of the beamforming gain with scalar, exhaustive search vector, and coordinate descent vector quantizers. (f) Effect of post-quantization noise on scalar and coordinate descent vector quantizers.

vector quantizer, the SIR is proportional to N^2/K^2. Recall that with scalar quantization the SIR was proportional to N/K. The functional dependence of SIR on N and K through the ratio N/K is corroborated by Figure 3(e), which shows that the SIR is constant when N is varied and K is set to $N/4$.

The dependence of the coordinate descent method on the number of quantization bits is presented in Figure 3(f). Just as in scalar quantization, the SIR grows exponentially in the bit rate, with a gain of approximately 6 dB per bit.

Finally, we also simulated the performance of the exhaustive search vector quantizer. Figure 3(g) shows the SIR in dB as a function of N, with $K = 2$ and one bit quantization. Figure 3(h) shows the SIR in dB as a function of K, with $N = 10$ and $B = 1$ bit quantization. While the computational complexity of the exhaustive search method presents an obstacle to providing more extensive simulations, it appears that the SIR (in linear units) grows exponentially in N and decreases exponentially in K. Thus, there may still be a significant gap between this optimal vector quantizer and the coordinate descent algorithm, where SIR grew polynomially in N/K. This potential gain motivates future work on computationally efficient vector quantizers that outperform coordinate descent.

D. Beamforming gain

While we have focused in this work on the effect of various quantization schemes on SIR, vector quantization can also improve the beamforming gain. Figure 2(e) shows the normalized array gain of the scalar quantizer, as well as the exhaustive search and coordinate descent vector quantizers. In this simulation, all phases were quantized to two levels (0 or π). Without quantization, the normalized gain would equal 1. The gap between vector and scalar quantization is smaller because beamforming is more robust to errors than interference rejection, and thus the naive method does not incur a large penalty.

E. Effect of noise

Up to this point, we have neglected all sources of error other than quantization. However, in practice there may be small additional noise due to various hardware imperfections or thermal variations. Figure 2(f) shows the SIR as a function of the variance of a uniform random variable that is added to each component of the antenna weight vector after quantization. The plot shows that while the performance of the coordinate descent method degrades in the presence of noise, it is still superior to the naive scalar quantization scheme.

VI. Conclusion

We have shown that coefficient quantization in digital receiver implementations can have a profound effect on the performance of interference-limited multi-antenna systems. In particular, quantization of beamforming weights critically affects interference suppression performance, and must be performed carefully in order to exploit the degrees of freedom gains from using an increased number of antennas. While our suboptimal greedy sequential strategy provides large gains over the naive scalar quantization strategy, our analytical estimates indicate that it might be possible to do much better. Important open problems include refinement of the analysis to provide tight upper and lower bounds on SIR scaling, and devising efficient adaptive and non-adaptive algorithms for finding the quantized weights that approach the performance of the optimal quantized weights.

References

[1] D. J. Love, R. W. Heath, W. Santipach, and M. L. Honig, "What is the value of limited feedback for MIMO channels?" *IEEE Communications Magazine*, vol. 42, no. 10, pp. 54–59, Oct 2004.

[2] K. Mukkavilli, A. Sabharwal, B. Aazhang, and E. Erkip, "Performance limits on beamforming with finite rate feedback for multiple antenna systems," in *Proc. of the Thirty-Sixth Asilomar Conference on Signals, Systems and Computers*, vol. 1, Nov 2002, pp. 536–540.

[3] L. Zheng and D. N. C. Tse, "Sphere packing in the Grassmann manifold: a geometric approach to the noncoherent multi-antenna channel," in *Proc. IEEE International Symposium on Information Theory*, 2000, pp. 364–.

[4] D. J. Love and R. W. Heath, "Multi-mode precoding using linear receivers for limited feedback MIMO systems," in *Communications, 2004 IEEE International Conference on*, vol. 1, June 2004, pp. 448–452.

[5] D. Manolakis, V. Ingle, and S. Kogon, *Statistical and Adaptive Signal Processing: Spectral Estimation, Signal Modeling, Adaptive Filtering and Array Processing.* McGraw-Hill, 2000.

[6] E. W. Weisstein, "Hypergeometric function," *Wolfram MathWorld*. [Online]. Available: http://mathworld.wolfram.com/HypergeometricFunction.html

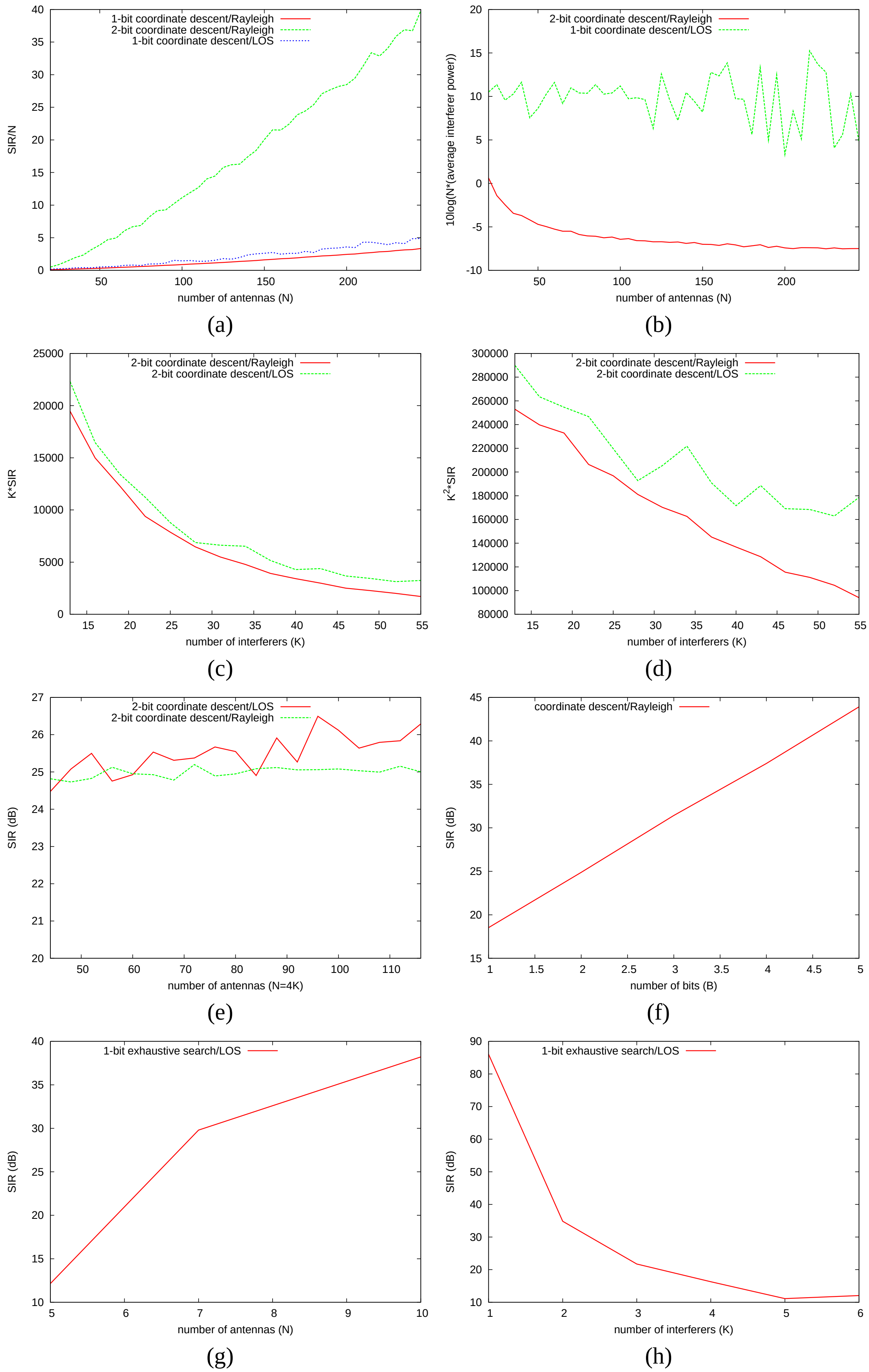

Fig. 3. (a)-(f) Simulated performance of the coordinate descent vector quantizer. (g)-(h) Simulated performance of the exhaustive search vector quantizer.

[7] J. S. Hicks and R. F. Wheeling, "An efficient method for generating uniformly distributed points on the surface of an n-dimensional sphere," *Commun. ACM*, vol. 2, no. 4, pp. 17–19, 1959.
[8] "Hypergeometric function," *Mathematics Wiki*. [Online]. Available: http://www.mathematics.thetangentbundle.net/wiki/spherical_coordinates

APPENDIX I
DISTRIBUTION OF THE ANGLE WITH RESPECT TO AN INTERFERENCE VECTOR

.

We start with the well-known formula for the area of a surface element of a d-dimensional unit hypersphere [7], [8]:

$$dS = \sin^{d-2}\theta_{d-2}d\theta_{d-2}\sin^{d-3}\theta_{d-3}d\theta_{d-3}\ldots\sin\theta_1 d\theta_1 d\theta_0 \tag{18}$$

where $\theta_{d-2},\ \theta_{d-3},\ \ldots,\ \theta_1,\ \theta_0$ are the $d-1$ angular coordinates of the hyper-spherical coordinate system. If we assume (without loss of generality) that the interference vector is aligned with one of the coordinate axes, then $\theta_x \equiv \theta_{d-2}$ is the "angle" made by the unit vector $\mathbf{x}$ on the surface element above with the interference vector. If the unit vector is assumed to be distributed uniformly on the surface of the hypersphere, then the probability density of the angle is proportional to the area element i.e. $f_\theta(\theta_x) \propto \sin^{d-2}\theta_{d-2}$. It is more convenient for us to work with the transformed angle $\phi_x \doteq \frac{\pi}{2} - \theta_x$, and we then have for the probability density of ϕ_x:

$$\begin{aligned} f_\phi(\phi_x) &= \frac{\sin^{d-2}\left(\frac{\pi}{2}-\phi_x\right)}{\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\sin^{d-2}\left(\frac{\pi}{2}-\phi_x\right)\,d\phi_x} \\ &\equiv \frac{\cos^{d-2}\phi_x}{\int_0^\pi \sin^{d-2}\phi_x\ d\phi_x} \end{aligned} \tag{19}$$

Define $I_n(t) \doteq \int_0^t \sin^n x\ dx$. We have the identity

$$I_n(t) \equiv -\cos t\ {}_2F_1\Big(\frac{1}{2}, \frac{1-n}{2}; \frac{3}{2}; \cos^2 t\Big) \tag{20}$$

where ${}_2F_1(.,.;.;.)$ is the Gauss hypergeometric function. However, it is more convenient to use the recursive formula:

$$I_n(t) \equiv -\frac{\sin^{n-1}t\cos t}{n} + \frac{n-1}{n}I_{n-2}(t) \tag{21}$$

We then have

$$\begin{aligned} \Pr(|\phi_x| \le \phi) &= 1 - 2\Pr\left(\phi < \phi_x \le \frac{\pi}{2}\right) \\ &= 1 - \frac{I_n(\phi)}{I_n(\frac{\pi}{2})} \end{aligned} \tag{22}$$

From (21), we have $I_n(\frac{\pi}{2}) \equiv \frac{n-1}{n}I_{n-2}(\frac{\pi}{2})$, and further using (21) in (22) we have

$$\Pr(|\phi_x| \le \phi) = \frac{\sin^{n-1}\phi\cos\phi}{(n-1)I_{n-2}(\frac{\pi}{2})} + \left(1 - \frac{I_{n-2}(\phi)}{I_{n-2}(\frac{\pi}{2})}\right) \tag{23}$$

We see from (23) that the cumulative probability distribution (cdf) of $|\phi_x|$, $F^{(N)}(\phi) \doteq \Pr(|\phi_x| \le \phi)$ is an increasing function of N. From this, we have

$$\Pr(|\phi_x| \le \phi) > \frac{2\phi_x}{\pi} \tag{24}$$

Finally we have the following lemma:

Lemma. Let $F_1(x),\ F_2(x)$ be two probability distributions (cdfs) in an interval $[a,b]$ such that $F_1(x) \le F_2(x),\ \forall x \in [a,b]$ and let $g(x)$ be a non-increasing differentiable function in $[a,b]$. Then $\mathrm{E}_{F_1}[g(x)] \le \mathrm{E}_{F_2}[g(x)]$.

Proof. Since $g(x)$ is a non-increasing differentiable function, we have $g'(x) \leq 0,\ \forall x \in [a, b]$. Consider

$$\mathrm{E}_{F_1}[g(x)] = \int_a^b g(x) dF_1(x) = \Big[g(x)F_1(x)\Big]_a^b - \int_a^b g'(x)F_1(x)\ dx \tag{25}$$

$$\equiv g(b) - \int_a^b g'(x)F_1(x)\ dx$$

$$\leq g(b) - \int_a^b g'(x)F_2(x)\ dx \equiv \mathrm{E}_{F_2}[g(x)] \tag{26}$$

where we used integration by parts in (25).

From the above lemma with $g(x) = \cos^2 x$ in $[0, \frac{\pi}{2}]$ and (23) we get

$$\mathrm{E}_{F^{(N_1)}}[\sin^2 \theta_x] \equiv \mathrm{E}_{F^{(N_1)}}[\cos^2 \phi_x] < \mathrm{E}_{F^{(N_2)}}[\sin^2 \theta_x],\ \forall N_1 < N_2 \tag{27}$$

If we set $N_1 = 0$, this gives $\mathrm{E}_{F^{(N)}}[\sin^2 \theta_x] > \frac{1}{2}$.