Recursive Consistent Estimation with Bounded Noise

Sundeep Rangan, *Member, IEEE,* and Vivek K Goyal, *Member, IEEE*

Abstract—Estimation problems with bounded, uniformly distributed noise arise naturally in reconstruction problems from over complete linear expansions with subtractive dithered quantization. We present a simple recursive algorithm for such bounded-noise estimation problems. The mean-square error (MSE) of the algorithm is "almost" $O(1/n^2)$, where n is the number of samples. This rate is faster than the $O(1/n)$ MSE obtained by standard recursive least squares estimation and is optimal to within a constant factor.

Index Terms—Consistent reconstruction, dithered quantization, frames, overcomplete representations, overdetermined linear equations.

I. INTRODUCTION

It is common to analyze systems including quantizers by modeling each quantizer as a source of signal-independent additive white noise. This model is precisely correct only when one uses subtractive dithered quantization, but for simplicity it is often assumed to hold for coarse, undithered quantization [1]–[3]. What can easily be lost in using this model is that the distribution of the quantization noise can be important, especially its boundedness.

This correspondence focuses on solving an overdetermined linear system of equations from quantized data. Assuming subtractive dither, this can be abstracted as the estimation of an unknown vector $x \in \mathbb{R}^r$ from measurements $y_k \in \mathbb{R}$

$$y_k = a_k' x + e_k, \qquad 0 \le k \le n - 1 \qquad (1)$$

where each $a_k \in \mathbb{R}^r$ is a known vector and the e_k's are independent and identically distributed (i.i.d.) random variables distributed uniformly on $[-\delta, \delta]$. [1] The maximum noise magnitude $\delta > 0$ is half of the quantization step size and is known *a priori*. Estimation problems of this form may arise elsewhere as well. At issue are the quality of reconstruction that is possible and the efficient computation of good estimates.

The classical method for estimating the unknown vector x is least squares estimation, which attempts to find $\hat{x}$ such that the ℓ_2-norm of the residual sequence $y_k - a_k' \hat{x}$ is minimized [4], [5]. Least squares estimators have been extensively studied and admit efficient implementations. Under mild assumptions, least squares estimates are guaranteed to converge to the true value as the number of samples grows to infinity.

However, least squares estimation may produce an estimate which not only differs from the maximum-likelihood (ML) and minimum mean-squared error estimates, but also is inconsistent with the bounds on the quantization noise. With the bound on e_k, each sample y_k in (1) places certain hard constraints on the location of the unknown vector

x. Least squares estimates are not in general consistent with these constraints. Since the constraints are convex, least squares estimates can be improved by projecting onto a set of estimates that are consistent.

Recently, it has been suggested that this improvement can result in faster order of convergence [6]–[9]. Numerical tests showed that, after applying consistency constraints, estimates can attain an $O(1/n^2)$ mean-squared error (MSE). Classical least squares estimation, which does not, in general, satisfy the hard constraints, attains only an $O(1/n)$ MSE.

The behavior and implementation of consistent estimation methods are not fully understood. While the $O(1/n^2)$ MSE for consistent estimation has been observed in a number of simulations, the decay rate has only been proven for certain sets $\{a_k\}$. The most general conditions under which $O(1/n^2)$ MSE is provably attainable are not currently known.

In addition, consistent estimation is difficult to implement recursively. Given n data points, finding a consistent estimate requires the solution of a linear program with r variables and $2n$ constraints. No recursive implementation of this computation is presently known. The linear program must be recomputed with each new observation and the size of the problem grows to infinity.

This correspondence introduces a simple, recursively implementable estimator with a provable $O(1/n^2)$ MSE. The proposed estimator is similar to the consistent estimation method of [7], [9], except that the estimates are only guaranteed to be consistent with the most recent data point. The estimator can be realized with an extremely simple update rule which avoids any linear programming. Our main results show that, under suitable assumptions on the vectors a_k, the simple estimator "almost" achieves the conjectured $O(1/n^2)$ MSE.

We will also show that under mild conditions on the *a priori* probability density of x, the MSE decay rate of any reconstruction algorithm is bounded below by $O(1/n^2)$. Thus the proposed estimator is optimal to within a constant factor. An $O(1/n^2)$ lower bound has also been shown in [10] under weaker assumptions that do not require uniformly distributed white noise. However, with the uniformly distributed white-noise model considered here, we will be able to derive a simple expression for the constant in this lower bound.

A. Summary of Contribution

As noted above, $O(1/n^2)$ MSE results have already appeared in the literature. This work has two distinguishing features: First, $O(1/n^2)$ MSE is obtained with an extremely simple algorithm that works recursively, i.e., uses each observation only once, with no increase in memory usage with time. Second, the requirement on the set of measurement "directions" $\{a_k\}$ is very mild (see Theorem 2). Until recently, the only published $O(1/n^2)$ MSE upper bounds for finite-dimensional signal spaces were derived from the analogous result for oversampled analog-to-digital (A/D) conversion of periodic band-limited signals [6], [7]. Thus, they were applicable to a particular family of $\{a_k\}$ sets known as Fourier frames [9]. A new approach reported in [11]—not based on consistency—attains $O(1/n^2)$ MSE more generally when the a_k's are uniform samples from a closed curve in $\mathbb{R}^r$; still, Theorem 2 given here is more general.

The previous paragraph requires a note of moderation because the estimation problem in this correspondence differs somewhat from those in [6]–[11]. These previous works used measurements from an (undithered) uniform quantizer $y_k = q(a_k' x)$. The bounds are for the squared error in estimating a fixed vector x while increasing the number of measurements n; constant factors in the bounds depend on x. Furthermore, when each a_k has equal norm—as assumed in these works—signal vectors x within a small ball centered at the origin

[1]All vectors are real column vectors. For a vector v, v' denotes its transpose and $\|v\|$ denotes its Euclidean norm. Expectation and probability are denoted with E and Pr, respectively.

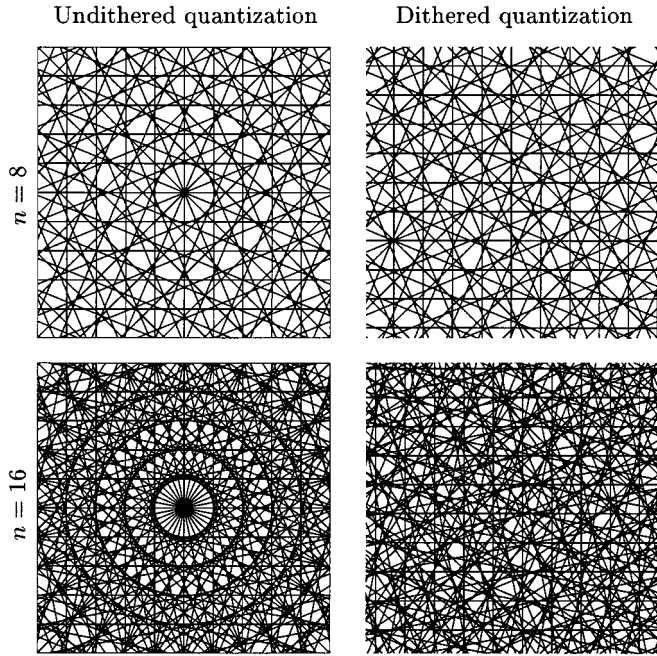Undithered quantization   Dithered quantization

Fig. 1. Geometric partitions induced by representing $x \in \mathbb{R}^2$ by quantized versions of $\{a_k' x\}_{k=0}^{n-1}$. Discrete Fourier frames are used [9]. Comparing the first and second row shows that doubling n roughly quadruples the number of cells; this is the fundamental reason for $O(1/n^2)$ MSE behavior. When undithered quantizers are used (left column), the properties of the partition depend on the distance from the origin. Analysis is easier with dithered quantization (right column) because the partition is shift-invariant.

must be excluded from consideration. As shown in Fig. 1, if x is close to the origin its radial component is not refined as n increases.

Use of dithered quantization gives us the simple model (1). The partition this induces is shift-invariant, as shown in Fig. 1. We are able to analyze performance by taking expectations over $\{e_k\}$, obtaining results that do not depend on x.

B. Related Work

To further contextualize this work, we should mention several more lines of related research. Zamir and Feder [12], [13] have studied the rate-distortion performance of a system which uses entropy-coded dithered quantization of an *oversampled* continuous-time signal and linear least squares reconstruction. Our main result (Theorem 2) suggests that these results could be revisited for alternative reconstruction strategies, and that the rate-distortion performance would be improved. Since our result is asymptotic in the oversampling ratio and does not yield simple expressions for distortion, this analysis may be difficult. This result may also be of interest to harmonic analysts studying the robustness of various overcomplete representations, such as nonorthonormal discrete wavelet expansions [14].

Also related to bounded noise estimation are various deterministic analyses (see [15], [16] and the references therein). Deterministic analysis concerns worst case estimation performance subject to bounds on the noise e_k. This formulation does not incorporate any statistical information on the noise and will not be considered here.

Our analysis of the recursive algorithm is based loosely on a standard stochastic approximation argument. A comprehensive survey of stochastic approximation methods can be found in the books by Kushner and Lin [17] and Ljung [4]. The lower bound is derived from a recently developed version of the Ziv–Zakai bound [18] presented in [19].

II. PROPOSED ALGORITHM AND CONVERGENCE PROPERTIES

Suppose $x \in \mathbb{R}^r$ is an unknown vector, and we obtain a set of observations y_k given by (1). We wish to find an estimate $\hat{x}_k$ of the unknown vector x from the data y_i and a_i for $i = 0, 1, \ldots, k-1$. The noise e_k is unknown, but bounded: $|e_k| \leq \delta$ for all k.

We propose the following simple recursive scheme:

$$\hat{x}_{k+1} = \hat{x}_k + \frac{a_k}{a_k' a_k} \phi(y_k - a_k' \hat{x}_k) \qquad (2)$$

where

$$\phi(e) = \begin{cases} 0, & \text{if } |e| \leq \delta \\ e - \delta, & \text{if } e > \delta \\ e + \delta, & \text{if } e < -\delta. \end{cases} \qquad (3)$$

(ϕ is a soft-thresholding function.) Any initial estimate $\hat{x}_0$ may be used.

The motivation behind this estimator is simple. If an observation y_k is consistent with the estimate $\hat{x}_k$ (i.e., $|y_k - a_k' \hat{x}_k| \leq \delta$), then the estimate is unchanged; that is, $\hat{x}_{k+1} = \hat{x}_k$. If the observation is not consistent, then $\hat{x}_{k+1}$ is taken to be the closest point to $\hat{x}_k$ consistent with the observation.

We will prove two results concerning this algorithm. The first result states that the estimation error decreases monotonically for any noise sequence e_k with $|e_k| \leq \delta$. No statistical assumptions are made.

Theorem 1: Fix a vector $x \in \mathbb{R}^r$, and consider the algorithm (2) acting on a sequence of observations y_k given by (1). If $|e_k| \leq \delta$, then

$$\|x - \hat{x}_{k+1}\| \leq \|x - \hat{x}_k\|.$$

Proof: See Appendix A.   □

For the second result, we impose the following assumption.

Assumption 1: For the measurements (1) and algorithm (2)

a) e_k and a_k are independently distributed random processes, independent from each another;
b) e_k is uniformly distributed on $[-\delta, \delta]$; and
c) there exist constants $M > 0$ and $\sigma > 0$ such that for all k, $\|a_k\|^2 \leq M$, and

$$\boldsymbol{E}|a_k' z| \geq \sigma \|z\| \qquad \forall z \in \mathbb{R}^r. \qquad (4)$$

The assumptions provide the simplest scenario in which to examine the algorithm, and are similar to those used in the classical analysis of the least mean squares (LMS) algorithm (see, for example, [4], [5]). The assumption (4), in particular, is a standard and mild persistent excitation condition.

The independence assumption on the vectors a_k is, however, somewhat restrictive, especially for analysis with deterministic a_k. While it is possible that this assumption can be replaced with suitable averaging conditions as in [4], [17], the analysis is considerably more difficult and will not be considered here. It should be noted that Assumption 1 does not require the vectors a_k to be identically distributed or have zero mean.

Theorem 2: Fix a vector $x \in \mathbb{R}^r$, and consider the algorithm (2) acting on a sequence of observations y_k given by (1). If Assumption 1 is satisfied then, for every $p < 1$

$$\|x - \hat{x}_k\| k^p \to 0 \text{ almost surely.}$$

Proof: See Appendix B.   □

Theorem 2 is our main result on the performance of the algorithm (2). The result states that, under suitable assumptions, the estimation error

$\|x - \hat{x}_k\|^2$ converges to zero, and for all $p < 1$, the rate of convergence is $o(k^{-2p})$. In this sense, the MSE is "almost" $O(1/k^2)$. As stated in Section I, this rate is superior to the $O(1/k)$ attained by classical least squares estimation.

III. BAYESIAN LOWER BOUNDS AND OPTIMALITY

We now derive a lower bound to the MSE. A consequence of this bound will be that the $O(1/n^2)$ rate of convergence of the recursive estimator (2) is asymptotically optimal.

Our derivation is based on a recently developed version of the Ziv–Zakai bound given in [18], [19]. Unlike the basic form of the better known Cramér–Rao bound, the Ziv–Zakai bound does not require smooth distributions or unbiased estimators. We derive the bound under the following assumption.

Assumption 2: For the observations (1)

a) e_k is an i.i.d. process with e_k uniformly distributed on $[-\delta, \delta]$.
b) x is an unknown vector with a continuous distribution function $p_X(x)$.

In these assumptions, the unknown vector x is modeled as a random variable with given *a priori* distribution, and the vectors a_k in (1) are assumed to be known and deterministic. This formulation provides a simple framework for deriving a lower bound, although the assumptions differ somewhat from Section II.

The following theorem is the main result of this section. In the statement of this theorem, an *estimator* will simply mean any sequence of functions on the data: $\hat{x}_k = \hat{x}_k(y_0, \ldots, y_{k-1}) \in \mathbb{R}^r$.

Theorem 3 (Ziv–Zakai Bound): Fix a set of vectors a_k and consider the observations y_k in (1). Under Assumption 2, the following holds for any estimator $\hat{x}_k$:

$$\liminf_{k \to \infty} k^2 \boldsymbol{E} \|x - \hat{x}_k\|^2 \geq \sum_{i=1}^{r} 2\delta^2 \Lambda(q_i)^{-2}$$

where q_i is the ith standard unit vector and

$$\Lambda(v) = \min_{w: w'v=1} \limsup_{k \to \infty} \frac{1}{k} \sum_{j=0}^{k-1} |a_j' w|. \qquad (5)$$

for $v \in \mathbb{R}^r$.

Proof: See Appendix C.   □

Theorem 3 provides a lower bound on the MSE in terms of the vectors a_k and noise magnitude δ. For large k, $\Lambda(v)$ can be approximated by

$$\Lambda(v) \approx \min_{w: w'v=1} \frac{1}{k} \sum_{j=0}^{k-1} |a_j' w|$$

and the minimization can be solved by linear programming. This provides a method for computing a lower bound on the achievable MSE for large k. Alternatively, one can take $w = v/v'v$, and use the bound

$$\Lambda(v) \leq \limsup_{k \to \infty} \frac{1}{k} \sum_{j=0}^{k-1} \frac{|a_j'v|}{v'v}.$$

This last estimate shows that if $\|a_j\|$ is bounded over all j, then $\Lambda(v) < \infty$ for any v. Now Theorem 3 implies that if $\Lambda(q_i) < \infty$ for any i, then

$$\liminf_{k \to \infty} k^2 \boldsymbol{E} \|x - \hat{x}_k\|^2 > 0.$$

Consequently, if the norms $\|a_j\|$ are bounded, the MSE of any estimator $\hat{x}_k$ is bounded below by $O(1/k^2)$. Since we have shown that the recursive estimator (2) achieves an MSE $o(1/k^{2p})$ for all $p < 1$, we can conclude that the proposed recursive estimator has an "almost" optimal rate of convergence.

IV. A NUMERICAL EXAMPLE

As a numerical test, we compared the performance of the proposed recursive algorithm against two other reconstruction methods: a linear programming (LP) algorithm and a classical recursive least squares (RLS) algorithm. For each of the algorithms, we measured the average MSE as a function of the number of samples.

The LP algorithm selects the vector x which minimizes the ℓ_∞ norm of the residual sequence $y_k - a_k'x$. This estimate corresponds to the ML estimate when both x and δ are treated as unknown. The computation of the LP estimate involves the solution of a linear program with $r + 1$ variables and $2n$ constraints, where n is the number of samples. This computation cannot be implemented recursively, and the linear program must be recomputed with each new sample. The LP estimate is the most computationally demanding of the three algorithms tested, but is the only one that produces estimates consistent with the noise bounds on all the samples available.

The RLS algorithm selects the vector x which minimizes the ℓ_2 norm of the residual sequence $y_k - a_k'x$. The RLS estimate is not, in general, consistent with the noise bounds on the data, but can be computed with a simple recursive update [5].

For the test, data in (1) was generated with $r = 4$ and $\{a_k\}$ being an i.i.d. process, uniformly distributed on the unit sphere in $\mathbb{R}^4$. We used a noise bound of $\delta = 1$. The algorithms were started with an initial error of $x - \hat{x}_0 = [1, 1, 1, 1]'$. Fig. 2(a) shows the results of a single simulation. As expected, the proposed recursive method yields nonincreasing distortion.

Fig. 2(b) shows the averaged results of 1000 simulations. Also plotted is the Ziv–Zakai MSE lower bound from Theorem 4. The asymptotic slopes of the curves confirm the $O(1/n)$ MSE for least squares estimation and the $O(1/n^2)$ MSE for the consistent LP estimation and the proposed algorithm. While very simple and recursive, the proposed algorithm performs only a constant factor worse than the nonrecursive consistent reconstruction and the theoretical lower bound.

V. IMPLICATIONS FOR SOURCE CODING AND DECODING

Thus far our discussion has been limited to the problem of estimating x given the y_k and a_k sequences. In this section, we consider the implications for using an entropy-coded version of y_k as a source encoding for x. Specifically, let $x \in \mathbb{R}^r$ be an arbitrary source vector. A representation of x can be formed through $y = Q_K(Ax)$, where A is an $n \times r$ matrix and $Q_K(\cdot)$ is an optimal K-dimensional entropy-coded dithered lattice quantizer [20], [21]. (The $K = 1$ case is the uniform scalar quantizer used in previous sections.)

If x comes from sampling a band-limited, periodic signal at the Nyquist rate and A is a Fourier matrix, this corresponds to the encoding for band-limited continuous-time signals considered by Zamir

and Feder [12]. [2] They showed that, *using least squares reconstruction*, the MSE for the scheme is fixed as long as the ratio of the oversampling factor to the second moment of the quantizer is kept constant. They also showed that as the dimension of the lattice quantizer is increased, the performance of this scheme for a Gaussian source and MSE distortion measure approaches the rate-distortion bound [13].

Instead of least squares reconstruction, we now consider using algorithm (2) for estimating the vector x from the quantized data $y = Q_K(Ax)$. Although our analysis does not directly apply to the case when A is a deterministic matrix, Assumption 1c) is satisfied under any arbitrarily small random perturbation of the rows of A. Thus, our analysis for random matrices A should apply to *generic* deterministic matrices as well.

Also, although we have described the algorithm (2) only for the scalar quantization case $K = 1$, it is straightforward to extend the algorithm to the lattice quantization case when $K > 1$. For the $K = 1$ case, we have taken each sample to specify a pair of hyperplane constraints on x. This can be viewed as a rotated and shifted version of the Cartesian product of a one-dimensional (1-D) lattice quantizer cell (an interval) and $\mathbb{R}^{r-1}$. For general K, each set of K samples specifies a constraint on x that is a rotated and shifted version of the Cartesian product of a K-dimensional cell with $\mathbb{R}^{r-K}$. An iterative reconstruction algorithm could update its estimate every K samples with the nearest point of this set.

Our results show that the reconstruction algorithm (2), or the extension of the algorithm for lattice vector quantization, will attain an $O(1/R^2)$ MSE, where $R = n/r$ is the oversampling ratio. This rate is superior to the $O(1/R)$ rate attained by least squares reconstruction. We conclude that a reconstruction algorithm which utilizes the hard bounds on the quantization error will have rate-distortion performance better than that described in [12], [13]. In the limit of high oversampling, the MSE would remain fixed when the ratio of the *square* of the oversampling ratio to the second moment of the quantizer is kept constant. Furthermore, with the extension to lattice quantization, the performance approaches the rate-distortion bound more quickly as the lattice dimension is increased.

VI. Conclusion

We have presented a simple, recursively implementable algorithm for estimation with uniformly distributed noise. This algorithm exhibits an $O(1/n^2)$ MSE where n is the number of samples. It is shown that this rate is asymptotically optimal to within a constant factor. Moreover, the rate is faster than the $O(1/n)$ MSE attained by classical least squares methods.

However, while the proposed estimator has an optimal order of convergence, there is still considerable potential for improvement. Our numerical tests indicated a large gap in performance between the recursive estimator, and both the nonrecursive linear programming method and the theoretical lower limit. We are currently investigating various methods to reduce this gap, while maintaining the recursive algorithm's computational simplicity. One promising avenue is to employ the ellipsoidal bounding methods presented in [23] which have been used in deterministic estimation problems [15], [16].

Of course, if we are willing to sacrifice recursiveness we may cycle through the data more than once, or we may reuse the data according to a sliding window. In either case, the performance would be improved with an increase in complexity proportional to the number of times each data point is used. Since our focus was on obtaining the best *order of convergence*, this was not explored. We may also define a consistent set to project to not based on a single sample, but rather based on some

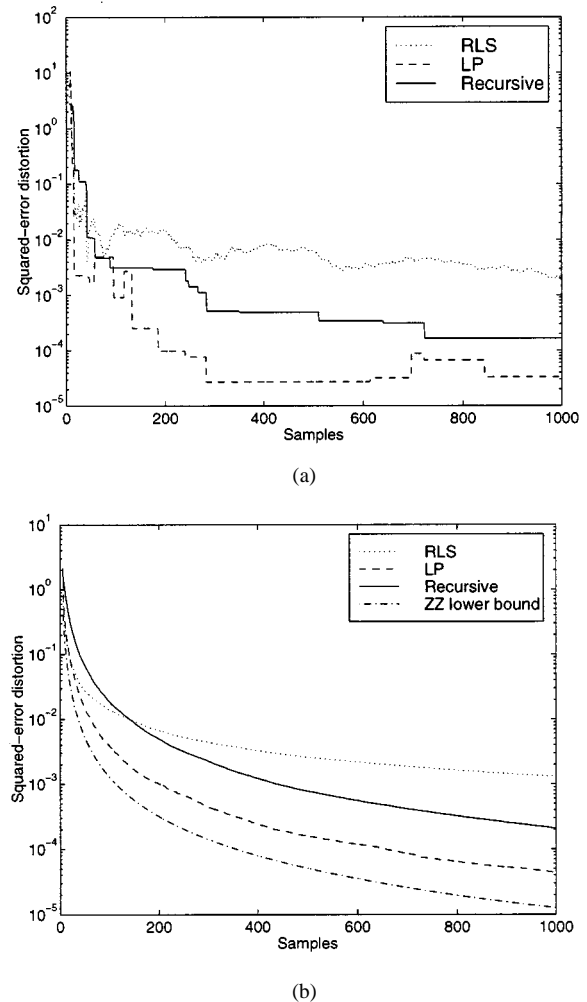[2]A similar situation without the constraint of periodicity is studied in [22].

(a)

(b)

Fig. 2.   Comparison of three reconstruction algorithms. "RLS" refers to the recursive minimum ℓ_2-error reconstruction; "LP" refers to a linear program reconstruction which computes an estimate consistent with the smallest possible noise δ; and "Recursive" refers to the proposed algorithm. "ZZ lower bound" is the Ziv–Zakai theoretical lower bound. (a) Single simulation. (b) Average of 1000 simulations.

number m of samples. This would again improve performance but increase complexity; the increase in complexity jumps when m exceeds r because the consistent set is not a Cartesian product of intervals.

These results serve as a reminder that simple linear filtering is not optimal for removing non-Gaussian additive noise, even if it is white. Accordingly, the improvement from consistent reconstruction in [6], [7], [9] is not because of the determinism of quantization noise, but because of its boundedness. It is an open and interesting question whether the results presented in this correspondence can be extended to undithered quantization, thus at least partially settling a conjecture of [9] in the affirmative.

Motivated by a source-coding application, and for concreteness in the proof of Theorem 2, uniformly distributed noise was assumed. However, the algorithm itself uses only the bound on the noise δ. This raises the broader issue of the value of hard information. It seems that hard information may be fundamentally more informative than "soft," or probabilistic, information. In many systems, all the signals—including the noise—can be bounded using certain physical considerations. This sort of "hard" information should be exploited fully.

APPENDIX A
PROOF OF THEOREM 1

Let $w_k = x - \hat{x}_k$. We must show that $\|w_{k+1}\| \leq \|w_k\|$. Rewrite (2) as

$$w_{k+1} = w_k - \frac{a_k}{a_k' a_k} \phi(a_k' w_k + e_k).$$

Taking the norm of both sides and manipulating the result gives

$$\|w_{k+1}\|^2 = \|w_k\|^2 + \frac{\phi(u_k + e_k)}{a_k' a_k}[-2u_k + \phi(u_k + e_k)] \quad (6)$$

where $u_k = a_k' w_k$. Now, it can be verified from the definition of ϕ in (3) that for any $e \in [-\delta, \delta]$ and $u \in \mathbb{R}$

$$\phi(u + e) \geq 0 \iff u \geq 0$$
$$\phi(u + e) - 2u \geq 0 \iff u \leq 0.$$

Thus, since $e_k \in [-\delta, \delta]$

$$\phi(u_k + e_k)(\phi(u_k + e_k) - 2u_k) \leq 0.$$

Hence, (6) shows that $\|w_{k+1}\| \leq \|w_k\|$ for all k.

APPENDIX B
PROOF OF THEOREM 2

It suffices to prove the theorem for $p \in (1/2, 1)$. Thus, fix any $p \in (1/2, 1)$ and let $z_k = k^p(x - \hat{x}_k)$. We must show that $z_k \to 0$ almost surely.

The proof will follow that of [17, Theorem 5.4.2]. For $\lambda > 0$, define the set

$$Q_\lambda = \{z \in \mathbb{R}^r \mid \|z\|^2 \leq \lambda\}.$$

We will show that there exists a constant $C_1 > 0$ such that for all $\lambda > 0$, the following two events occur almost surely:

a) the set Q_λ is recurrent, i.e., z_k returns to Q_λ infinitely often; and
b) there are at most a finite number of transitions of z_k from Q_λ to $Q_{(1+C_1)\lambda}^c$.

Together, these two assertions imply that, for all $\lambda > 0$ and all k sufficiently large, $z_k \in Q_{(1+C_1)\lambda}$. That is,

$$\limsup_k \|z_k\|^2 \leq (1 + C_1)\lambda.$$

Since this is true for all $\lambda > 0$, $z_k \to 0$ almost surely, which is precisely the statement of the theorem.

We begin by proving assertion a).

Lemma 1: For all $\lambda > 0$, the set Q_λ is recurrent.

Proof: Using $z_k = (x - \hat{x}_k)k^p$, rewrite (2) as

$$z_{k+1} = \left(\frac{k+1}{k}\right)^p z_k - \frac{(k+1)^p a_k}{a_k' a_k} \phi(a_k' z_k k^{-p} + e_k). \quad (7)$$

Assumption 1a) implies that z_k is a Markov process.

Let $V(z) = \|z\|^2$ and $V_k = V(z_k)$. Denote by $\boldsymbol{E}_k(\cdot)$ the conditional expectation given z_k. The lemma will be proven with the following standard Martingale result (see, for example, [17, Theorem 4.4.4]): if there exist $T > 0$ and $K > 0$ such that

$$\boldsymbol{E}_k V_{k+1} \leq V_k - K, \qquad \text{when } k > T \text{ and } V_k \geq \lambda \quad (8)$$

then the set $Q_\lambda = \{z \mid V(z) \leq \lambda\}$ is recurrent. Thus, the lemma reduces to showing (8) for some $T > 0$ and $K > 0$.

Unfortunately, proving (8) requires us to estimate $\boldsymbol{E}_k V_{k+1} - V_k$, which demands a somewhat long and tedious calculation. For space considerations, we will omit many of the details.

Squaring both sides of (7), one can obtain

$$\|z_{k+1}\|^2 = \left(\frac{k+1}{k}\right)^{2p} \|z_k\|^2 + h_k \quad (9)$$

where

$$h_k = \frac{(k+1)^{2p}}{a_k' a_k} \left[\phi^2(u_k + e_k) - 2\phi(u_k + e_k)u_k\right]$$
$$u_k = k^{-p} a_k' z_k. \quad (10)$$

Since $p \in (1/2, 1)$, the map $w \mapsto w^{2p}$ is convex in w so

$$\left(\frac{k+1}{k}\right)^{2p} - 1 \leq \frac{2p}{k}.$$

Using this inequality along with (9) and the fact that $V_k = \|z_k\|^2$ we get

$$V_{k+1} \leq V_k + \frac{2p}{k}\|z_k\|^2 + h_k. \quad (11)$$

We next compute $\boldsymbol{E}_k h_k$. To this end, suppose u is a constant and e is uniformly distributed on $[-\delta, \delta]$. Then it can be verified that

$$\boldsymbol{E}\left[\phi^2(u + e) - 2u\phi(u + e)\right] = -\varphi(|u|) \quad (12)$$

where

$$\varphi(u) = \begin{cases} \frac{1}{3\delta}|u|^3, & |u| \leq 2\delta \\ |u|^2 - \frac{4}{3}\delta^2, & |u| > 2\delta. \end{cases} \quad (13)$$

Also, by explicitly computing the derivative of $\varphi(u)$, it can be shown that the derivative is monotonically increasing and hence φ is convex.

Now, since e_k is independent of a_k and z_k, we can apply (12) to (10) to obtain

$$\boldsymbol{E}(h_k | z_k, a_k) = -\frac{(k+1)^{2p}}{a_k' a_k} \varphi(|a_k' z_k k^{-p}|)$$
$$\leq -\frac{(k+1)^{2p}}{M} \varphi(|a_k' z_k k^{-p}|)$$

where in the last step we used the bound in Assumption 1c). Taking expectations over a_k, using the fact that a_k is independent of z_k and using the estimate in Assumption 1c)

$$\boldsymbol{E}_k h_k = \boldsymbol{E}(h_k | z_k) \leq -\frac{(k+1)^{2p}}{M} \boldsymbol{E}_k \varphi(|a_k' z_k k^{-p}|)$$
$$\leq -\frac{(k+1)^{2p}}{M} \varphi(\boldsymbol{E}_k |a_k' z_k k^{-p}|)$$
$$\leq -\frac{(k+1)^{2p}}{M} \varphi(\sigma\|z_k\|k^{-p}). \quad (14)$$

Note that we have used the convexity of φ along with Jensen's inequality. Finally, using (13), it can be shown that for any $\lambda > 0$, there exists a $K > 0$ and $T > 0$ such that

$$\varphi(\sigma|z|k^{-p}) > \frac{2p}{k}\|z\|^2 + K$$

when $k > T$ and $\|z\|^2 \geq \lambda$. Using this along with (11) and (14) shows (8), and the proof of the lemma is complete. □

We now turn to proving assertion b) stated at the beginning of the proof. This will be done in the next three lemmas. We will continue to use the notation in Lemma 1.

Lemma 2: There exist constants $C_1 > 0$ and $C_2 > 0$ such that for all $k > 0$ and z_k

$$\boldsymbol{Pr}(h_k < -2C_1\|z_k\|^2 | z_k) > C_2 k^{-p}$$

where h_k is defined in (10).

Proof: For $k \geq 0$, define the events

$$A_k = \{e_k \in [\delta - u_k/2, \delta] \cup [-\delta, -\delta - u_k/2]\}$$
$$B_k = \{|a_k' z_k| \geq \sigma \|z_k\|/2\}$$

where, as before, $u_k = k^{-p} a_k' z_k$. Using the definition of ϕ in (3), it can be verified that if the event A_k occurs, then

$$\phi(u_k + e_k)^2 - 2u_k \phi(u_k + e_k) \leq -u_k^2/2 = -k^{-2p}|a_k'z_k|^2/2.$$

Hence, (10) and Assumption 1c) show that if both A_k and B_k occur

$$h_k \leq -\frac{(k+1)^{2p}}{2k^{2p}a_k'a_k}|a_k'z_k|^2 \leq -\frac{\sigma^2}{2M}\|z_k\|^2.$$

We set $C_1 = \sigma^2/4M$, so that the lemma will be proven, if we can show

$$\boldsymbol{Pr}(A_k \cap B_k | z_k) > C_2 k^{-p} \qquad (15)$$

for some $C_2 > 0$ and all $k > 0$ and z_k.

To find a $C_2 > 0$ such that (15) holds, first observe that since e_k is uniformly distributed on $[-\delta, \delta]$ and is independent of a_k and z_k

$$\boldsymbol{Pr}(A_k | z_k, a_k) = \frac{|u_k|}{4\delta} = \frac{k^{-p}|a_k'z_k|}{4\delta}.$$

Hence

$$\boldsymbol{Pr}(A_k | B_k, z_k) \geq \frac{\sigma}{8k^p\delta}\|z_k\|. \qquad (16)$$

Also, using Assumption 1c)

$$\sigma\|z_k\| \leq \boldsymbol{E}(|a_k'z_k| \,|\, z_k)$$
$$\leq \frac{1}{2}\sigma\|z_k\| + \sqrt{M}\|z_k\|\boldsymbol{Pr}(|a_k'z_k| \geq \frac{1}{2}\sigma\|z_k\| \,|\, z_k)$$

and, therefore,

$$\boldsymbol{Pr}(B_k | z_k) = \boldsymbol{Pr}(|a_k'z_k| \geq \frac{1}{2}\sigma\|z_k\| \,|\, z_k) \geq \frac{\sigma}{2\sqrt{M}}. \qquad (17)$$

Thus, if we define

$$C_2 = \frac{\sigma^2}{16\delta\sqrt{M}}$$

(16) and (17) show that (15) holds for all $k > 0$, and the lemma is proven. $\qquad \square$

Now, fix a $\lambda > 0$. The sequence z_k will be said to have a *run* from λ to μ at n if there exists an $m > 0$ such that $\|z_n\|^2 \leq \lambda, \|z_{n+m}\|^2 \geq \mu$, and $\|z_k\|^2 \in (\lambda, \mu)$ for all $k \in (n, n+m)$. The number m will be called the *length* of the run. The next lemma will show that as n increases, the probability of a run at n becomes small.

Lemma 3: Let $\lambda > 0$ and let C_1 be as in Lemma 2. Then there exist constants $N > 0$ and $\rho \in (0, 1)$ such that for $n \geq N$, the probability of a run at n from λ to $(1 + C_1)\lambda$ is less than $\rho^{n^{1-p}}$.

Proof: We first claim that there exists an $\alpha > 0$ such that the length of any run from λ to $(1 + C_1)\lambda$ at n is at least αn. To see this, suppose z_k has a run at n of length m. Then

$$\lambda \geq \|z_n\|^2 = n^{2p}\|w_n\|^2$$
$$(1 + C_1)\lambda \leq \|z_{n+m}\|^2 = (n + m)^{2p}\|w_{n+m}\|^2$$

where $w_k = x - \hat{x}_k$ as before. By Theorem 1, $\|w_{n+m}\| \leq \|w_n\|$, and, therefore,

$$\frac{1}{1 + C_1} \geq \frac{n^{2p}\|w_n\|^2}{(n + m)^{2p}\|w_{n+m}\|^2} \geq \left(\frac{n}{n + m}\right)^{2p}.$$

This implies that with $\alpha = (1 + C_1)^{1/2p} - 1 > 0$, we have $m \geq \alpha n$. Thus, the length of any run at n is at least αn.

Now choose N such that $k \geq N$ implies that

$$\left(1 + \frac{1}{k}\right)(1 + C_1) \leq 1 + 2C_1 \qquad (18)$$

and

$$\alpha k - 2 \geq \alpha k/2. \qquad (19)$$

Then, if $k \geq N$, $\|z_k\|^2 \in (\lambda, (1 + C_1)\lambda)$ and $\|z_{k+1}\|^2 \geq \lambda$, (11) and (18) show that

$$h_k \geq \|z_{k+1}\|^2 - (1 + 1/k)\|z_k\|^2$$
$$\geq \lambda - (1 + 2C_1)\lambda = -2C_1\lambda \geq -2C_1\|z_k\|^2.$$

Therefore, Lemma 2 implies that if $k \geq N$ and $\|z_k\|^2 \in (\lambda, (1 + C_1)\lambda)$

$$\boldsymbol{Pr}(\|z_{k+1}\|^2 \geq \lambda | z_k) \leq \boldsymbol{Pr}(h_k \geq -2C_1\|z_k\|^2 | z_k)$$
$$\leq 1 - k^{-p}C_2\|z_k\|^2 \leq 1 - k^{-p}C_2\lambda.$$

Therefore,

$$\boldsymbol{Pr}(\|z_{k+1}\|^2 \geq \lambda| \|z_k\|^2 \in (\lambda, (1 + C_1)\lambda)) \leq 1 - k^{-p}C_2\lambda. \quad (20)$$

Now, if z_k has a run from λ to $(1 + C_1)\lambda$ at n, we have shown above that the length of this run must be at least αn. Hence,

$$\|z_k\|^2 \in (\lambda, (1 + C_1)\lambda), \qquad \text{for all } k \in (n, n + \alpha n).$$

Using (19), (20), the Markov property of z_k, and the inequality $\log(1 - x) \leq -x$, the probability of a run at n must be less than

$$\prod_{k=n+1}^{n+\alpha n-1}(1 - k^{-p}C_2\lambda) \leq (1 - n^{-p}C_2\lambda)^{\alpha n-2}$$
$$= \exp[(\alpha n - 2)\log(1 - n^{-p}C_2\lambda)]$$
$$\leq \exp(-(\alpha n/2)n^{-p}C_2\lambda) = \rho^{n^{1-p}}$$

where

$$\rho = e^{-\alpha C_2\lambda/2} < 1.$$

This completes the proof of the lemma. $\qquad \square$

Our final lemma proves assertion b) made at the beginning of the proof.

Lemma 4: Let $\lambda > 0$ and let C_1 be as in Lemma 2. Then there can be at most a finite number of runs from λ to $\lambda(1 + C_1)$.

Proof: Let P_n denote the probability of a run at n from λ to $\lambda(1 + C_1)$. Lemma 3 states that for $n \geq N$

$$P_n \leq \rho^{n^{1-p}}.$$

Hence

$$\sum_{n=N}^{\infty} P_n \leq \sum_{n=N}^{\infty} \rho^{n^{1-p}} \leq \int_{N-1}^{\infty} \rho^{x^{1-p}} \, dx$$
$$= \frac{1}{1-p} \int_{(N-1)^{1-p}}^{\infty} u^{p/(1-p)}\rho^u \, du < \infty.$$

The result now follows from the Borel–Cantelli lemma. $\qquad \square$

Lemmas 1 and 4 prove assertions a) and b) made at the beginning of the proof. As argued there, these two facts prove the theorem.

APPENDIX C
PROOF OF THEOREM 3

We begin with describing the general Ziv–Zakai bound. For any vector $x_0 \in \mathbb{R}^r$, let $\boldsymbol{Pr}(\cdot | x_0)$ denote the conditional probability given the unknown vector $x = x_0$. Given vectors $x_0, x_1 \in \mathbb{R}^r$, let

$$P_{\min, k}(x_1, x_0) = \min \frac{1}{2}[\boldsymbol{Pr}(\hat{x}_k = x_1 | x_0) + \boldsymbol{Pr}(\hat{x}_k = x_0 | x_1)]$$

$$(21)$$

where the minimum is taken over all estimators $\hat{x}_k$. The quantity $P_{\min, k}(x_1, x_0)$ represents the minimum probability of error in

estimating x given that x is *a priori* known to be either x_1 or x_0. The Ziv–Zakai bound can now be stated as follows.

Theorem 4: Suppose that $P_{\min, k}(x_1, x_0)$ is only a function of $x_1 - x_0$, i.e., $P_{\min, k}(x_1, x_0) = P_{\min, k}(x_1 - x_0)$. Then for any vector $v \in \mathbb{R}^r$ and estimator $\hat{x}_k$

$$\boldsymbol{E}|v'(x - \hat{x}_k)|^2 \geq \int_0^\infty \max_{w:\, w'v=h} [hA(w)P_{\min, k}(w)]\, dh \qquad (22)$$

where

$$A(w) = \int \min(p_X(x),\, p_X(x + w))\, dx \qquad (23)$$

and $p_X(x)$ is the *a priori* distribution of x.

Proof: See [19, Property 4]. $\qquad\square$

To apply this bound, we must derive expressions for $A(w)$ and $P_{\min, k}(w)$. We begin the derivation with a simple computation. For any sequence y_i, let

$$\Omega_k(y) = \{x \in \mathbb{R}^r : |y_j - a_j'x| \leq \delta,\ \forall j = 0, \ldots, k\}.$$

That is, $\Omega(y)$ is the set of vectors x consistent with the observations y_i up to time k. Our first lemma provides a simple expression for the probability that a given vector lies within this consistent set.

Lemma 5: For any constant vectors $x_0, x_1 \in \mathbb{R}^r$

$$\boldsymbol{Pr}(x_1 \in \Omega_k(y)|x_0) = P_k(x_1 - x_0)$$

where

$$P_k(v) = \prod_{j=0}^k \left[1 - \frac{|a_j'v|}{2\delta}\right]_+ \qquad (24)$$

and $[u]_+ = \max(u, 0)$.

Proof: For any j, e_j is uniformly distributed on $[-\delta, \delta]$, and, therefore,

$$\boldsymbol{Pr}(|y_j - a_j'x_1| \leq \delta|x_0) = \boldsymbol{Pr}(|a_j'(x_0 - x_1) + e_j| \leq \delta)$$
$$= \left[1 - \frac{|a_j'(x_1 - x_0)|}{2\delta}\right]_+.$$

Since the e_j's are independent

$$\boldsymbol{Pr}(x_1 \in \Omega_k(y)|x_0) = \boldsymbol{Pr}(|y_j - a_j'x_1| \leq \delta,\ j = 0, \ldots, k)$$
$$= \prod_{j=0}^k \left[1 - \frac{|a_j'v|}{2\delta}\right]_+ = P_k(x_1 - x_0). \quad\square$$

We next compute $P_{\min, k}(x_1, x_0)$ defined in (21).

Lemma 6: For x_0 and $x_1 \in \mathbb{R}^r$

$$P_{\min, k}(x_1, x_0) = \frac{1}{2}P_k(x_1 - x_0) \qquad (25)$$

where P_k is given in (24).

Proof: A standard hypothesis testing result states that the estimator achieving the minimum in (21) is given by the ratio test

$$\hat{x}_k^*(y) = \begin{cases} x_1, & \text{if } p_{Y|X}(y|x_1) \geq p_{Y|X}(y|x_0) \\ x_0, & \text{else} \end{cases} \qquad (26)$$

where $p_{Y|X}(y|x)$ is the conditional probability distribution of $y_0, \ldots, y_k$ given x. This estimator is optimal in that

$$P_{\min, k}(x_1, x_0) = \frac{1}{2}[\boldsymbol{Pr}(\hat{x}_k^* = x_1|x_0) + \boldsymbol{Pr}(\hat{x}_k^* = x_0|x_1)]. \quad (27)$$

Now, since $y_j = a_j'x + e_j$ and e_j is i.i.d. with e_j uniformly distributed on $[-\delta, \delta]$

$$p_{Y|X}(y|x_i) = \mathbf{1}_{\{x_i \in \Omega_k(y)\}}(2\delta)^{-k}.$$

Thus, the estimator (26) simplifies to

$$\hat{x}_k^*(y) = \begin{cases} x_1, & \text{if } x_1 \in \Omega_k(y) \\ x_0, & \text{else.} \end{cases}$$

Using Lemma 5

$$\boldsymbol{Pr}(\hat{x}_k^* = x_1|x_0) = \boldsymbol{Pr}(x_1 \in \Omega_k(y)|x_0) = P_k(x_1 - x_0) \quad (28)$$
$$\boldsymbol{Pr}(\hat{x}_k^* = x_0|x_1) = 1 - \boldsymbol{Pr}(x_1 \in \Omega_k(y)|x_1) \qquad (29)$$
$$= 1 - P_k(x_1 - x_1) = 0. \qquad (30)$$

Substituting (28) and (30) into (27) proves the result. $\qquad\square$

Now, using the above expression for $P_{\min, k}(x_1, x_0)$ in the Ziv–Zakai bound, we obtain the following.

Lemma 7: If $v \in \mathbb{R}^r$ and $\hat{x}_k$ is any estimator

$$\liminf_{k \to \infty} k^2 \boldsymbol{E}|v'(x - \hat{x}_k)|^2 \geq 2\delta^2 \Lambda(v)^{-2}.$$

where $\Lambda(v)$ is defined in (5).

Proof: Substituting (25) into the Ziv–Zakai bound (22) and performing a change of variables gives

$$k^2 \boldsymbol{E}|v'(x - \hat{x}_k)|^2 \geq \frac{k^2}{2} \int_0^\infty \max_{w:\, w'v=h} [hA(w)P_k(w)]\, dh$$
$$= \frac{1}{2} \int_0^\infty \max_{w:\, w'v=1} \left[hA\left(\frac{hw}{k}\right) P_k\left(\frac{hw}{k}\right)\right] dh \qquad (31)$$

where $A(w)$ is defined in (23). The lemma can now be proven by simply taking the limits as $k \to \infty$.

First consider the function $A(w)$. Using the Dominated Convergence Theorem and the continuity of $p_X(x)$

$$\lim_{\|w\| \to 0} A(w) = \int \lim_{\|w\| \to 0} \min(p_X(x + w),\, p_X(w))\, dx$$
$$= \int p_X(x)\, dx = 1. \qquad (32)$$

For the function $P_k(w)$, we note that for small w

$$P_k(w) = \exp\left[\sum_{j=0}^k \log\left(1 - \frac{|a_j'w|}{2\delta}\right)\right]$$
$$\approx \exp\left[-\frac{1}{2\delta} \sum_{j=0}^k |a_j'w|\right]$$

where we have used the fact that $\log(1 - x) \approx -x$ for small x. Thus

$$\liminf_{k \to \infty} P_k\left(\frac{w}{k}\right) = \liminf_{k \to \infty} \exp\left[-\frac{1}{2\delta k} \sum_{j=0}^k |a_j'w|\right]$$
$$= \exp\left[-\frac{1}{2\delta}\Lambda_0(w)\right] \qquad (33)$$

where

$$\Lambda_0(w) = \limsup_{k \to \infty} \frac{1}{k} \sum_{j=0}^k |a_j'w|.$$

Note that

$$\Lambda(v) = \min_{w:\, w'v=1} \Lambda_0(w). \qquad (34)$$

Using (31)–(34) and Fatou's lemma

$$\liminf_{k \to \infty} k^2 \boldsymbol{E}|v'(x - \hat{x}_k)|^2$$
$$\geq \frac{1}{2} \int_0^\infty \liminf_{k \to \infty} \left[\max_{w:\, w'v=1} hA\left(\frac{hw}{k}\right) P_k\left(\frac{hw}{k}\right)\right] dh$$
$$\geq \frac{1}{2} \int_0^\infty \max_{w:\, w'v=1} \left[\liminf_{k \to \infty} hA\left(\frac{hw}{k}\right) P_k\left(\frac{hw}{k}\right)\right] dh$$

$$= \frac{1}{2} \int_0^\infty \max_{w:\, w'v=1} h \exp\left(\frac{-h\Lambda_0(w)}{2\delta}\right) dh$$

$$= \frac{1}{2} \int_0^\infty h \exp\left(\frac{-h\Lambda(v)}{2\delta}\right) dh$$

$$= 2\delta^2 \Lambda(v)^{-2}$$

and the lemma is proven. $\qquad\square$

The theorem now follows as a straightforward consequence of Lemma 7. If q_i is the ith standard unit vector, then

$$\|x - \hat{x}_k\|^2 = \sum_{i=1}^r |q_i'(x - \hat{x}_k)|^2.$$

Thus, Lemma 7 implies that

$$\liminf_{k \to \infty} k^2 \boldsymbol{E} \|x - \hat{x}_k\|^2 \geq 2\delta^2 \sum_{i=1}^r \Lambda(q_i)^{-2}.$$

ACKNOWLEDGMENT

REFERENCES

[1] R. M. Gray, "Quantization noise spectra," *IEEE Trans. Inform. Theory*, vol. 36, pp. 1220–1244, Nov. 1990.

[2] S. P. Lipshitz, R. A. Wannamaker, and J. Vanderkooy, "Quantization and dither: A theoretical survey," *J. Audio Eng. Soc.*, vol. 40, no. 5, pp. 355–375, May 1992.

[3] R. M. Gray and T. G. Stockham Jr., "Dithered quantizers," *IEEE Trans. Inform. Theory*, vol. 39, pp. 805–812, May 1993.

[4] L. Ljung, *Theory and Practice of Recursive Identification*. Cambridge, MA: MIT Press, 1983.

[5] S. S. Haykin, *Adaptive Filter Theory*, 3rd ed. Upper Saddle River, NJ: Prentice-Hall, 1996.

[6] N. T. Thao and M. Vetterli, "Reduction of the MSE in R-times oversampled A/D conversion from $O(1/R)$ to $O(1/R^2)$," *IEEE Trans. Signal Processing*, vol. 42, pp. 200–203, Jan. 1994.

[7] ——, "Deterministic analysis of oversampled A/D conversion and decoding improvement based on consistent estimates," *IEEE Trans. Signal Processing*, vol. 42, pp. 519–531, Mar. 1994.

[8] Z. Cvetković, "Overcomplete expansions for digital signal processing," Ph.D. dissertation, Univ. California, Berkeley, 1995.

[9] V. K. Goyal, M. Vetterli, and N. T. Thao, "Quantized overcomplete expansions in $\mathbb{R}^N$: Analysis, synthesis, and algorithms," *IEEE Trans. Inform. Theory*, vol. 44, pp. 16–31, Jan. 1998.

[10] N. T. Thao and M. Vetterli, "Lower bound on the mean-squared error in oversampled quantization of periodic signals using vector quantization analysis," *IEEE Trans. Inform. Theory*, vol. 42, pp. 469–479, Mar. 1996.

[11] Z. Cvetković, "Source coding with quantized redundant expansions: Accuracy and reconstruction," in *Proc. IEEE Data Compression Conf.*, J. A. Storer and M. Cohn, Eds. Snowbird, UT, Mar. 1999, pp. 344–353.

[12] R. Zamir and M. Feder, "Rate-distortion performance in coding bandlimited sources by sampling and dithered quantization," *IEEE Trans. Inform. Theory*, vol. 41, pp. 141–154, Jan. 1995.

[13] ——, "Information rates of pre/post-filtered dithered quantizers," *IEEE Trans. Inform. Theory*, vol. 42, pp. 1340–1353, Sept. 1996.

[14] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA: Soc. Industr. Appl. Math., 1992.

[15] M. Cwikel and P. O. Gutman, "Convergence of an algorithm to find maximal state constraint sets for discrete-time linear dynamical systems with bounded controls and states," *IEEE Trans. Automat. Contr.*, vol. AC-31, pp. 457–459, May 1986.

[16] M. Milanese and V. Vicino, "Optimal estimation theory for dynamic systems with set membership uncertainty: An overview," *Automatica*, vol. 27, no. 6, pp. 997–1009, Nov. 1991.

[17] H. J. Kushner and C. G. Lin, *Stochastic Approximation Algorithms and Applications*. New York: Springer-Verlag, 1997.

[18] J. Ziv and M. Zakai, "Some lower bounds on signal parameter estimation," *IEEE Trans. Inform. Theory*, vol. IT-15, pp. 386–391, May 1969.

[19] K. L. Bell, Y. Steinberg, Y. Ephraim, and H. L. Van Trees, "Extended Ziv–Zakai lower bound for vector parameter estimation," *IEEE Trans. Inform. Theory*, vol. 43, pp. 624–637, Mar. 1997.

[20] J. Ziv, "On universal quantization," *IEEE Trans. Inform. Theory*, vol. IT-31, pp. 344–347, May 1985.

[21] R. Zamir and M. Feder, "On universal quantization by randomized uniform/lattice quantization," *IEEE Trans. Inform. Theory*, vol. 38, pp. 428–436, Mar. 1992.

[22] Z. Cvetković and M. Vetterli, "Error-rate characteristics of oversampled analog-to-digital conversion," *IEEE Trans. Inform. Theory*, vol. 44, pp. 1961–1964, Sept. 1998.

[23] M. J. Todd, "On minimum volume ellipsoids containing part of a given ellipsoid," *Math. Oper. Res.*, vol. 7, no. 2, pp. 253–261, May 1982.

Bounds for Sparse Planar and Volume Arrays

Yann Meurisse and Jean-Pierre Delmas, *Member, IEEE*

Abstract—This correspondence improves and extends bounds on the numbers of sensors, redundancies, and holes for sparse linear arrays to sparse planar and volume arrays. As an application, the efficiency of regular planar and volume arrays with redundancies but no holes is deduced. Also, examples of new redundancy and hole square arrays, found by exhaustive computer search, are given.

Index Terms—Difference base, minimum hole array, minimum redundancy array, sparse planar array, sparse volume array.

I. INTRODUCTION

When the number of antenna sensors available for an array is limited, the problem of optimum array geometry naturally arises. From the beam width and the sidelobe level of the associated beam pattern [1] or from the direction of arrival (DOA) estimation accuracy [2] point of view, array configurations known as linear minimum-redundancy (MR) arrays or linear minimum-hole (MH) arrays (also called optimum nonredundant arrays) are often proposed. Linear MR arrays have been extensively studied; see [3] and [4], and the references therein. In particular, much attention has been given to bounds on the ratio M^2/A [4], [5] where M and A denote, respectively, the number of sensors and the aperture of the linear array. Linear MH arrays were considered in [3] and [6]. Whereas specific structures were designed to optimize some performance criteria (e.g., [7] for DOA algorithms with DOA prior information and [1] for beam patterns with various sidelobe level/beamwidth tradeoffs); redundancy and hole concepts do not embrace any such optimality criterion directly. Thus, the MR and MH