

Improved bounds on Gaussian MAC and sparse regression via Gaussian inequalities

Ilias Zadik, Yury Polyanskiy, and Christos Thrampoulidis

Abstract—We consider the Gaussian multiple-access channel with two critical departures from the classical asymptotics: a) number of users proportional to blocklength and b) each user sends a fixed number of data bits. We provide improved bounds on the tradeoff between the user density and the energy-per-bit. Interestingly, in this information-theoretic problem we rely on Gordon’s lemma from Gaussian process theory. From the engineering standpoint, we discover a surprising new effect: good coded-access schemes can achieve perfect multi-user interference cancellation at low user density.

In addition, by a similar method we analyze the limits of false-discovery in binary sparse regression problem in the asymptotic regime of number of measurements going to infinity at fixed ratios with problem dimension, sparsity and noise level. Our rigorous bound matches the formal replica-method prediction for some range of parameters with imperceptible numerical precision.

I. INTRODUCTION

In this paper we consider the following two problems. First is the K -user AWGN multiple-access channel (MAC). The classical information theory analysis assumes K fixed and takes the users’ data payload and blocklength both growing to infinity. Following [1] we, instead, consider fixed size of the payload and take both K and blocklength growing to infinity (the case of all three growing to infinity has been considered in [2] under the name of many-user MAC). For this problem we derive a new achievability bound, which outperforms the one in [1] and (essentially) matches the converse bound for a range of parameters. Our bounds lead us to confirm the effect suggested in [1]: there exists a critical density of users below which the optimal decoder is able to completely “tune-out” all other users and achieve performance as if each user operated in isolation. This perfect multi-user interference cancellation disappears for high user densities.

The second problem we address is sparse regression, where (Gaussian) random linear combinations of an unknown binary sparse vector are observed in (Gaussian) noise and the goal is to reconstruct the vector subject to bounds on probability of mis-detection and false-detection rate. (As explained in [1] this problem may be seen as the *random-access* version of the first problem.) For this well-studied problem we prove new

reconstruction bounds which almost match (i.e., indistinguishable numerically) the non-rigorous replica-method prediction in a range of parameters. Note that replica predictions for MMSE and mutual information have recently been made rigorous [3]. Here we study more singular metrics, probability of mis-detection and false-discovery rates, where, to the best of our knowledge, such rigorous results are not known.

We mention that all of our bounds are based on exponential-time-complexity decoders, and as such are not practically relevant except as benchmarks. For benchmarking, the bounds have to be possible to evaluate non-asymptotically, a property all our bounds share, but replica-predictions in principle are not. Other practical aspects of these problems, practical motivation, and detailed discussion of prior work can be found in [4]–[8].

Our proof proceeds by reducing either problem to that of bounding the extrema of a certain Gaussian process. We then proceed to bound the latter via the use of Gordon’s lemma [9]. A lot of work on sparse regression have been similarly reducing the analysis to question of Gaussian width of tangent cones via application of Gordon’s lemma. However, we consider our approach novel for the following reasons.

Gordon’s inequality has been previously used to study the performance of non-smooth convex optimization methods (e.g. ℓ_1 -minimization) in structured signal recovery. The relevance of Gordon’s result to such problems was first noted in [10] and was popularized mainly due to the works of Stojnic [11], [12]. Since then it has been successfully used in a series of papers, e.g. [13]–[18] under many different settings. Most of these, involve analysis of the mean-squared error, but more general error criteria (such as bit-error rate) have been recently studied [19]. The aforementioned literature is concerned with convex decoders; in contrast, in this work (similar to [20]) we are interested in analyzing non-convex decoders. In fact, we show in Section II that a “straightforward” application of Gordon’s inequality as suggested in the recent literature may yield suboptimal bounds on the probability of error of our (non-convex) decoders. We manage to get improvement by, instead, appropriately combining Gordon’s inequality with Chernoff bounds.

Notation: Let $\phi(\cdot)$ denote the Gaussian pdf and $Q^{-1}(\cdot)$ denote the inverse of the standard tail function $Q(\cdot)$ of the Gaussian distribution, i.e., $Q(t) = \int_t^\infty \phi(s)ds$. Also, for any $\chi \in \mathbb{R}$, we denote $(\chi)_+ = \max\{\chi, 0\}$ and, for $\epsilon \in (0, 1)$, we write $h(\epsilon) = -\epsilon \ln(\epsilon) - (1 - \epsilon) \ln(1 - \epsilon)$ for the entropy function.

Y.P. is with the Department of EECS, MIT, Cambridge, MA, email: yp@mit.edu. C.T. is with the Department of ECE, UC Santa Barbara, CA, email:cthrampo@ucsb.edu. I.Z. is with the Operations Research Center, MIT, Cambridge, MA, email: izadik@mit.edu. This research has been supported in part by the Center for Science of Information (CSoI), an NSF Science and Technology Center, under grant agreement CCF-09-39370, by the NSF CAREER award CCF-12-53205, and by the NSF award CCF-17-17842.

II. PROOF TECHNIQUES: BOUNDING EXTREMA OF GAUSSIAN PROCESSES

To illustrate the technical tools needed, we consider a simpler guiding problem. Let $G \in \mathbb{R}^{\epsilon n \times n}$ with entries i.i.d. $\mathcal{N}(0, 1)$ and $u \in \mathbb{R}^n$ fixed. Define

$$V_\alpha \triangleq \min_{x \in S_\alpha} \frac{1}{n} \|Gx + u\|_2, \quad (1)$$

where $S_\alpha \triangleq \{x \in \{0, 1\}^n : \|x\|_0 = \alpha n\}$. Our goal is to get tight lower bounds on the asymptotic $\liminf_{n \rightarrow \infty} \mathbb{E}[V_\alpha]$. The map $G \mapsto V_\alpha$ is $\sqrt{\frac{\alpha}{n}}$ -Lipschitz; thus by concentration of measure instead of bounding $\mathbb{E}[V_\alpha]$ we look for $r = r(\alpha, \epsilon, u)$ such that $\Pr(V_\alpha \leq r) \rightarrow 0$ as $n \rightarrow \infty$. Lemma 1 and 2 below show two ways to get such lower bound: a more straightforward way using Chernoff inequality and union bounds and a more sophisticated way using Gordon's inequality. It turns out neither method dominates the other and for this reason both are used to obtain our main results in Secs. III and IV.

Lemma 1 (Union bound & Chernoff). *Fix $\alpha, \epsilon \in (0, 1)$. Let $r \in (0, \alpha\epsilon]$ such that*

$$r^2 E\left(\frac{\|u\|_2^2}{r^2 n}, \sqrt{\alpha\epsilon}\right) > h(\alpha) \quad (2)$$

where for $v, a > 0$

$$E(v, a) = \min_{\lambda \geq 0} \left\{ \lambda - \frac{1}{2} \ln(1 + 2\lambda a^2) - \frac{\lambda v}{1 + 2\lambda a^2} \right\}. \quad (3)$$

Then for some $c > 0$, $\Pr[V_\alpha \leq r] \leq e^{-cn}$ as $n \rightarrow \infty$.

Remark 1. The minimizing λ for $E(v, a)$ defined in (3) equals $\lambda^* = \left| \frac{a^2 - 2 + \sqrt{a^4 + 4v}}{4a^2} \right|_+$.

Proof. First, for $H \sim \mathcal{N}(0, I_{\epsilon n})$ it follows from union bound that

$$\Pr[V_\alpha \leq r] \leq \binom{n}{\alpha n} \Pr[\|\sqrt{\alpha}H + u\|_2 \leq r\sqrt{n}] \quad (4)$$

To bound the last probability recall the identity

$$\mathbb{E} \left[e^{-\lambda \|aZ + u\|_2^2} \right] = e^{-\frac{\lambda \|u\|_2^2}{1 + 2a^2\lambda}} / (1 + 2a^2\lambda)^{\frac{m}{2}} \quad (5)$$

$\forall \gamma, a > 0, u \in \mathbb{R}^m, Z \sim \mathcal{N}(0, I_m)$. Applying Chernoff bound, thus, we get in the same notation:

$$\Pr[\|aZ + u\|_2^2 \leq m] \leq e^{mE(\|u\|^2/m, a)}, \quad (6)$$

where $E(v, a)$ is as in (3). Plugging (6) into (4) and using $\binom{n}{\alpha n} = e^{(1+o(1))h(\alpha)n}$ completes the proof. $\square$

Define

$$\gamma(s) \triangleq \frac{1}{\sqrt{2\pi}} e^{-(Q^{-1}(s))^2/2}, \quad (7)$$

with a useful (and asymptotically tight as $s \rightarrow 0$) upper bound $\gamma(s) \leq s\sqrt{-2\ln s}$.

Lemma 2 (Gaussian process (GP) bound). *For any*

$$r < \sqrt{\alpha\epsilon + \frac{\|u\|_2^2}{n}} - \gamma(\alpha), \quad (8)$$

where $\gamma(\cdot)$ was defined in (7), there exists $C > 0$ such that $\Pr[V_\alpha \leq r] \leq e^{-Cn}$ for all n .

Proof. The proof is based on Gordon's inequality [9]. By a now standard argument (we omit the details for

brevity but see e.g. [13, Cor. 3.3]) it holds that for all values r such that

$$r < \sqrt{\alpha \frac{\mathbb{E}[\|h\|_2^2]}{n} + \frac{\|u\|_2^2}{n}} - \frac{1}{n} \mathbb{E}[\max_{x \in S_\alpha} g^T x],$$

there exists constant $C > 0$ such that $\Pr[V_\alpha \leq r] \leq e^{-Cn}$. Here, $h \sim \mathcal{N}(0, I_{\epsilon n})$ and $g \sim \mathcal{N}(0, I_n)$. Note that $\max_{x \in S_\alpha} g^T x = \sum_{i=1}^{\alpha n} h_i^\downarrow$, where we denote the entries of h sorted in decreasing order as $h_1^\downarrow \geq h_2^\downarrow \geq \dots \geq h_n^\downarrow$. It can be shown that $\mathbb{E}[\sum_{i=1}^{\alpha n} h_i^\downarrow] = \gamma(\alpha) + \mathcal{O}_\alpha(n^{-1/2})$ [20]. Moreover, $\mathbb{E}[\|h\|_2^2] = (1 + o(1))\epsilon n$. These combined with (8) complete the proof. $\square$

III. MULTIPLE-ACCESS CHANNEL

A. Problem formulation

Consider n channel uses of a standard K -user Gaussian multiple-access channel (GMAC, cf. [21, Section 4.7]) with $K = \mu n$ users, and where each user is sending $\log_2 M$ bits. In coding-theoretic terms each user j has a codebook of M vectors $c_1^{(j)}, \dots, c_M^{(j)} \in \mathbb{R}^n$, and the receiver observes Y given by, $Y = \sum_{j=1}^K c_{W_j}^{(j)} + Z$, $Z \sim \mathcal{N}(0, I_n)$, where $W_j \stackrel{iid}{\sim} \text{Unif}[M]$ correspond to user messages. Receiver's task is to find estimates $\hat{W}_j$ of true messages W_j with a per-user probability of error (PUPE):

$$\frac{1}{K} \sum_{j=1}^K \Pr[W_j \neq \hat{W}_j] \leq \epsilon. \quad (9)$$

Notice that by linearity of expectation, (9) is equivalent with at most ϵK expected number of estimate errors. Designing the codebooks is crucially constrained by the energy requirement:

$$\|c_m^{(j)}\|_2^2 \leq 2\mathcal{E} \cdot \log_2 M \quad \forall j \in [K], m \in [M], \quad (10)$$

where $\mathcal{E}$ is the energy-per-bit, or $\frac{E_b}{N_0}$. If $\{c_m^{(j)}\}_{j \in [K], m \in [M]}$ satisfy (10) and there exists decoder satisfying (9), we say that $\{c_m^{(j)}\}_{j \in [K], m \in [M]}$ is an $(n, M, \epsilon, \mathcal{E}, K)$ -code. Our goal is to characterize the following quantity

$$\mathcal{E}^*(M, \mu, \epsilon) = \lim_{n \rightarrow \infty} \inf \{ \mathcal{E} : \exists (n, M, \epsilon, \mathcal{E}, \mu n)\text{-code} \}, \quad (11)$$

where as usual the limit is understood as $\liminf$ or $\limsup$ depending on whether an upper or a lower bound is given. The practical meaning of the quantity $\mathcal{E}^*$ is the fundamental energy requirement for a massive number of (frame-synchronized and pre-identified) users sharing the same spectrum at density of μ users per real degree of freedom (rdof) and communicating with fixed-size packets of $\log_2 M$ data bits per packet.

Remark 2 (Finite payload). The idea of scaling K with n has appeared before as *many-user channel* formulation in [2]. However, the analysis there was done for the fixed power per user (i.e. $\|c_m^{(j)}\|_2^2 = nP$ for a fixed $P > 0$ and $n \rightarrow \infty$). This always results in $\mathcal{E} \rightarrow \infty$ and $\log_2 M \rightarrow \infty$, and thus did not capture neither the effects of finite payload size $\log_2 M$ nor the stringent energy-per-bit requirements of IoT.

Remark 3 (Reasonable value of μ). Consider a metropolitan area with 10^6 - 10^7 devices. The sub-GHz

ISM band is about 20 MHz wide. Therefore, if each of the devices is active a few times per hour, we get the ratio of 10^3 - 10^4 rdoF per active user at any given moment, i.e. $\mu = 10^{-4}$ to 10^{-3} . Note that this number is unlikely to vary significantly in the near future.

Remark 4 (Block-sparse regression formulation). We can describe equivalently the encoding strategy by a $n \times KM$ matrix X where columns $M(j-1)+1$ to Mj correspond to codewords of user $j \in [K]$, $c_1^{(j)}, \dots, c_M^{(j)}$. For a certain vector $\beta \in \{0, 1\}^{KM}$ the receiver observes $Y = X\beta + Z$ where $Z \sim \mathcal{N}(0, I_n)$. The vector β is uniformly distributed on the subset T_M of block-sparse vectors defined as

$$\left\{ \beta \in \{0, 1\}^{KM} : \sum_{i=M(j-1)+1}^{Mj} \beta_i = 1, \quad \forall j \in [K] \right\}$$

The receiver's task is to produce a block-sparse vector $\hat{\beta} \in T_M$ with $\mathbb{E}[d_H(\beta, \hat{\beta})] \leq 2\epsilon K$.

B. New bound based on Gaussian process inequalities

In Theorem 3 below, we present our main result, which is an upper bound on $\mathcal{E}^*$ (cf. (11)). Let us first define the following function,

$$\psi(b, \theta, \mu) \triangleq (\sqrt{1 + b^2\theta\mu} - b\mu\gamma(\theta))_+, \quad (12)$$

where $\gamma(\cdot)$ is defined in (7).

Theorem 3 (Bound-MAC). Fix M, μ and ϵ . Then

$$\mathcal{E}^*(M, \mu, \epsilon) \leq \inf \frac{b^2}{2\log_2 M},$$

where infimum is over all $b > 0$ such that for all $\theta \in [\epsilon, 1]$ it holds that

$$\theta\mu \ln M + \mu h(\theta) < \max_{\lambda \geq 0} \left\{ \frac{1}{2} \ln(1 + 2b^2\theta\mu\lambda) + \lambda \frac{\psi(b, \theta, \mu)^2}{1 + 2b^2\theta\mu\lambda} - \lambda \right\}, \quad (13)$$

where ψ is defined in (12).

Remark 5. It can be checked that the optimal value of λ in the RHS of (13) is $\lambda^* = \frac{r + \sqrt{r^2 + 4\psi^2} - 2}{4r}$ with $r = b^2\theta\mu$ and $\psi = \psi(b, \theta, \mu)$.

Proof sketch of Thm. 3 The bound is achieved via an appropriate random coding scheme. Specifically, we generate KM codewords $c_m^{(j)} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, PI_n)$, $P = \frac{b^2}{n}$. For analysis, we adopt the equivalent description of Remark 4 of the block-sparse regression setting. We have $X = \frac{b}{\sqrt{n}}G$, where G is an $n \times (\mu M)n$ matrix with iid $\mathcal{N}(0, 1)$ entries. For a decoder, we use the maximum likelihood estimator $\hat{\beta}$:

$$\hat{\beta} = \underset{\beta' \in T_M}{\operatorname{argmin}} \|Y - X\beta'\|_2^2, \quad (14)$$

over all the possible messages of the K users. (Note that the decoder is in principle suboptimal from the point of view of error criterion (9)). We prove that for any b satisfying (13) there is a $c > 0$ such that $\mathbb{P}[d_H(\hat{\beta}, \beta) = 2t] \leq e^{-cn}$ for all $t \geq \epsilon K$. To that end, condition on $\beta = \beta_0$ for any fixed $\beta_0 \in T_M$ and decompose $\beta' = \beta'_0 + \beta'_1$ with $\beta'_0 = \beta' \cap \beta_0$ (where we view the 0/1-vectors as subsets of $[KM]$). Denote $\mathcal{W} = \{\beta_0 - \beta' : \beta' \in T_M\}$ and $\mathcal{W}_t = \{\beta_0 -$

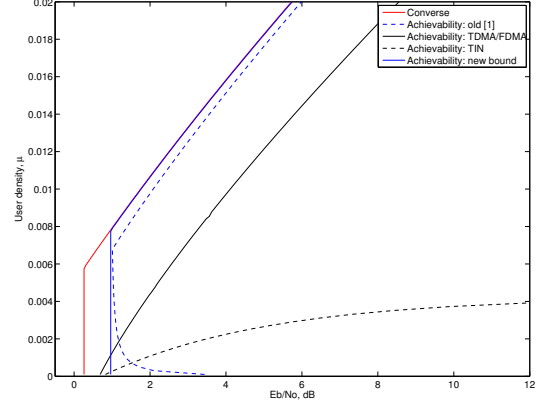


Fig. 1. Bounds on the range of feasible μ and E_b/N_0 for $k = 100$ bits per packet and $\text{PUPE} = 10^{-3}$.

$\beta' : d_H(\beta', \beta_0) = 2t, \beta' \in T_M\}$. Then we have self-evident bounds

$$\begin{aligned} & \mathbb{P}[d_H(\hat{\beta}, \beta) = 2t] \\ & \leq \mathbb{P}[\min_{w \in \mathcal{W}_t} \|Xw + Z\|_2 \leq \min_{w \in \mathcal{W} \setminus \mathcal{W}_t} \|Xw + Z\|_2] \\ & \leq \mathbb{P}[\min_{w \in \mathcal{W}_t} \|Xw + Z\|_2 \leq \|Z\|_2]. \end{aligned} \quad (15)$$

We next show that when $t = \theta K = \theta\mu n$ this probability is upper bounded by e^{-nE} where $E = \text{LHS} - \text{RHS}$ of (13). We first easily prove that by rotational invariance of X and concentration of $\|Z\|_2$ it suffices to show

$$\mathbb{P}[\min_{w \in \mathcal{W}_t} \|bGw + \sqrt{n}\mathbf{1}\|_2 \leq n]$$

is at most e^{-nE} where $\mathbf{1}$ denotes the all-1 vector. In this way, we reduce the problem to computing extrema of a Gaussian process. The particular bound we employ next is the following. Note that any vector $w \in \mathcal{W}_t$ consists (besides zero values) of exactly t entries $+1$ and same number of -1 's, and so we decompose $w = w_- + w_+$ accordingly. There are only $\binom{K}{t}$ choices for w_+ , as they correspond to a t -subset of the support of β_0 , and to each such choice there are $(M-1)^t$ choices of w_- . Thus, for each w_- we can use Chernoff bound to lower bound $\min_{w_+} \|Gw_+ + v\|_2$ in terms of $\|v\|_2^2$ (Lemma 1), where $v = Gw_- + \sqrt{n}\mathbf{1}$. We then use Lemma 2 for $u = 0$ to lower-bound $\min_{w_-} \|Gw_-\|_2$.

Remark 6 (Comparison to bounds of [1]). The new bound uniformly outperforms the bounds in [1] (see Section III-C). In [1] two bounds were proposed, denoted by p_t and q_t . p_t is based on “error-exponent” type analysis and is the one evaluated in Section III-C. q_t similar to this paper requires extreme value analysis of certain Gaussian process. The bound of this work significantly tightens q_t via using a combination of Chernoff bound and Gordon’s lemma instead of the standard information-density based change of measure.

C. Discussion and Plots

We evaluate the bound from Thm. 3 for $k = 100$ and $\text{PUPE} = 10^{-3}$ on Fig. 1. We plot and compare with one converse bound and three achievability bounds.

a) *Converse bounds:* We compare ourselves with two converse bounds from [1]. For large user-density we use that any code with parameters (M, μ, ϵ) must

satisfy: $(1 - \epsilon)\rho \leq \frac{1}{2} \log(1 + P_{tot}) + \mu h(\epsilon)$, where P_{tot} is the total power in each channel use, $P_{tot} = 2\mu \mathcal{E} \log_2 M$ and ρ is the total spectral efficiency $\rho = \mu \log_2 M$. This bound follows from Fano's inequality, the sum-rate for the GMAC, and the rate-distortion function for M -ary source reconstructed with Hamming-like distortion. Notice that our upper bound meets tightly this converse bound when $\mu > 0.008$. For small user density we use $\log_2 M \leq -\log_2 Q(\sqrt{\frac{P_{tot}}{\mu}} + Q^{-1}(1 - \epsilon))$ which follows from the fact that each user only transmits finitely many bits and then standard bounds connecting probability of error and packet size $\log_2 M$ [22, Theorem 2]. Our upper bound has the vertical behavior of the converse bound and lessens its achievable distance to it.

b) Achievability Bounds: TDMA/FDMA, TIN and Achievability from [1]: In [1] three upper bounds were discussed, one based on orthogonalization (TDMA/FDMA), one by treating interference as noise (TIN) and an original bound from [1]. Notice that our bound improves upon all of them for all but extremely small values of μ .

Vertical Slope for small μ : Both the converse and the achievability bounds strongly suggest that the optimal E_b/N_0 is invariant for small values of μ . In other words, for μ less certain critical value optimal coding systems are able to decode multiple users as if each of them were alone in the frame, i.e. as if $\mu \sim 0$. That is, there is a perfect cancellation of the multi-user interference. Note that it is a significant improvement to orthogonalization, since each user gets a whole frame for itself, not just $\frac{1}{K}$ portion of it. See also the relevant discussion in the regression section about this effect from the statistical physics point of view. Finally, let us point that there is a statistical-physics replica prediction for the predicted optimal PUPE and MAC-relevant definition of sparsity in [23]. Nevertheless we do not plot as computing it requires numerical evaluations over a 2^k -dimensional space.

IV. SPARSE LINEAR REGRESSION

Let $\mu, \epsilon \in (0, 1)$, $M, b > 0$ and for $n \in \mathbb{N}$, $K = \mu n$ and $p = M\mu n$. Consider n noisy linear samples $Y = X\beta + Z \in \mathbb{R}^n$ of an (unknown) u.a.r. chosen K -sparse binary vector $\beta \in \{0, 1\}^p$. Assume $X \in \mathbb{R}^{n \times p}$ has iid $\mathcal{N}(0, \frac{b^2}{n})$ entries and Z has iid $\mathcal{N}(0, 1)$ entries. The task of interest is given (Y, X) to produce an estimator K -sparse binary vector $\hat{\beta} = \hat{\beta}(Y, X)$ of β with expected Hamming error at most ϵ , that is

$$\mathbb{E}[d_H(\hat{\beta}, \beta)]/2K \leq \epsilon. \quad (16)$$

Our goal will be to identify the minimum possible signal level $b = b(M, \mu, \epsilon)$ so that for some $\hat{\beta}$ the criterion (16) is satisfied asymptotically as $n \rightarrow +\infty$. Note that exactly because $\hat{\beta}$ is constrained to be K -sparse, $d_H(\hat{\beta}, \beta)/2K$ simultaneously equals to both the False Discovery Rate $\text{FDR}(\hat{\beta})$ as well as the complement of the True Positive Rate $1 - \text{TPR}(\hat{\beta})$ of the estimator $\hat{\beta}$. For this reason we refer to (16) as FDR of the estimator. Similar to the previous section we can equivalently focus on minimizing $\mathcal{E}_R = b^2/2 \log_2 M$ the energy per-bit in the regression problem. As explained in [1]

this problem can be understood as a random-access version of the MAC problem.

A. Result

Theorem 4 (Bound-Regression). *We use γ and ψ from (7), (12) respectively. Fix M, μ and ϵ . For any $b = b(M, \mu, \epsilon)$ such that for all $\theta \in [\epsilon, 1]$ one of the following two statements hold, either*

$$\sqrt{2\theta b^2 \mu + 1} \sqrt{\frac{1}{\mu M}} - \sqrt{\frac{\mu b^2}{M}} \tilde{\gamma} > 1, \quad (17)$$

with $\tilde{\gamma} \triangleq (M - 1)\gamma(\frac{\theta}{M-1}) + \gamma(\theta)$, or, (see also Remark 5)

$$(M - 1)\mu h(\frac{\theta}{M - 1}) + \mu h(\theta) <$$

$$\max_{\lambda \geq 0} \left\{ \frac{1}{2} \ln(1 + 2b^2 \theta \mu \lambda) + \lambda \frac{\psi(b, \theta, \mu)^2}{1 + 2b^2 \theta \mu \lambda} - \lambda \right\}. \quad (18)$$

there exist binary K -sparse $\hat{\beta} = \hat{\beta}(Y, X)$ s.t. $\frac{1}{2K} \mathbb{E}[d_H(\hat{\beta}, \beta)] \leq \epsilon + o(1)$.

The proof is on the same lines as in Theorem 3. Specifically, our ‘‘Chernoff+Gordon’’ bound yields (18). Moreover, for the problem described in this section, we strengthen the result by combining it with a bound that follows by a direct application of Gordon's lemma, which as explained next leads to (17). Recall from Section II that neither of the two methods (Lemma 1 or 2) universally dominates each other, thus we need both in Theorem 4.

Proof sketch: Set $\mathcal{B}_K = \{\beta' \in \{0, 1\}^p : \|\beta'\|_0 = K\}$ the set of binary K -sparse vectors. For a decoder, we use the maximum likelihood estimator $\hat{\beta}$:

$$\hat{\beta} = \underset{\beta' \in \mathcal{B}_K}{\operatorname{argmin}} \|Y - X\beta'\|_2^2 \quad (19)$$

and prove that for any b satisfying the theorem's conditions there is a $c > 0$ such that $\mathbb{P}[d_H(\hat{\beta}, \beta) = 2t] \leq e^{-nc}$ for all $t \geq \epsilon K$. Similar to the proof of Theorem 3 we condition $\beta = \beta_0$ for some fixed β_0 and then it suffices to show that for all $t = \theta K \geq \epsilon K$:

$$\mathbb{P}[\min_{w \in \mathcal{W}_t} \|bGw + \sqrt{n}\mathbf{1}\|_2 \leq n] \leq e^{-nc}, \quad (20)$$

under (18) or (17). Here, $G \in \mathbb{R}^{n \times p}$ has i.i.d. $\mathcal{N}(0, 1)$ entries and $\mathcal{W}_t \triangleq \{\beta_0 - \beta' : d_H(\beta', \beta_0) = 2t\}$. First, the proof under (17) proceeds by high probability lower bounding the extrema of the Gaussian Process $\min_{w \in \mathcal{W}_t} \frac{1}{n} \|bGw + \sqrt{n}\mathbf{1}\|_2$ via a direct application of Gordon's Lemma. This requires a little more effort compared to the proof of Lemma 2 due to the additional constraints in $\mathcal{W}_t$, but it can be shown that the minimization above is lower bounded with exponentially high probability by the LHS of (17) scaled by n ; we omit the details for brevity. Next, the proof under (18) is shown by repeating the argument in the proof of Theorem 3, mutatis mutandis, with the only difference that now there are exactly $\binom{p-K}{t}$ choices for w_- and for each, $\binom{K}{t}$ choices of w_+ .

B. Comparison to replica prediction and converse

Consider the following one-dimensional setting : $\beta \sim \text{Bern}(1/M)$, we observe $Y = \sqrt{E}\beta + Z$, $Z \sim \mathcal{N}(0, 1)$. Denote $I_1(E) = I(\beta; Y)$ the mutual information between signal and observation. We want

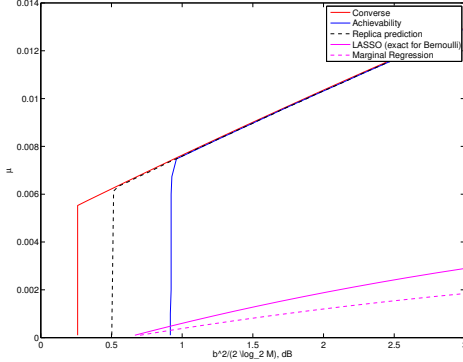


Fig. 2. Comparing bounds on $\mu \sim p/n$ vs b^2 tradeoff for FDR = 10^{-3} and $M = 2^{100}$.

to produce an estimator $\hat{\beta}(Y) \in \{0, 1\}$ with minimum probability of error (PE) $\mathbb{P}[\hat{\beta} \neq \beta]$ subject to $\mathbb{P}[\hat{\beta} = 1] = 1/M$. Via standard arguments the minimal PE $\epsilon_1^*(E)$ satisfies $\sqrt{E} = Q^{-1}(\frac{1}{M-1}\epsilon_1^*) + Q^{-1}(\epsilon_1^*)$. Returning to sparse regression, we recall that statistical-physics methods, e.g. [24], (non-rigorously) predict an *asymptotic decoupling* phenomenon. Namely, that inferring β from the observations (Y, X) behaves asymptotically as the one dimensional task described above with $E = \eta^*(\mu)b^2$, where $\eta^* \in [0, 1]$ is the minimizer of the free energy potential: $\eta^* = \operatorname{argmin}_{\eta \in [0, 1]} \mu MI_1(\eta b^2) + \frac{1}{2}(\eta - 1 - \ln \eta)$. Combining the above, the replica method predicts that the optimal FDR (LHS of (16)) for the sparse regression setting with fixed b^2 asymptotically converges to $\epsilon_1^*(\eta^*(\mu)b^2)$. We compare the replica prediction against our bounds on Fig. 2. In terms of this prediction, the existence of the vertical line in the plot corresponds to the fact that as b increases (under fixed μ) the efficiency η experiences a step transition to 1. Thus for sufficiently large b the replica FDR is independent of μ and is given by $\epsilon_1^*(b^2)$, which in turn means that the minimal b required for a certain level of FDR is constant in μ . (We note that discontinuous jump of η^* to 1 is unrelated to sparsity, cf. [24, Fig. 10c].) Fig. 2 also depicts the converse (impossibility) bound, which is a combination of two simple bounds. First, if genie reveals to decoder $K - 1$ entries of β , then we get list-decoding problem (with list size K) over n channel uses of an AWGN channel, for which we have from [22] a firm non-asymptotic bound: $b \geq Q^{-1}(1/M) + Q^{-1}(\epsilon)$ (almost coincides with $\epsilon_1^*(b^2)$). Using $I(Y; \beta|X) \leq \frac{n}{2} \ln(1+b^2)$ and comparing it with the rate-distortion function of the K -out-of- p source with $2\epsilon K$ Hamming-distortion gives the second bound. (We omit straightforward expressions.)

Finally, we plot the exact performance of LASSO (subject to (16), where FDR is set equal to 10^{-3}) which for iid inputs is known exactly [18], [25], [26]. (Weaker LASSO bounds are in [10], [13], [27].)

REFERENCES

- [1] Y. Polyanskiy, "A perspective on massive random-access," in *Proc. 2017 IEEE Int. Symp. Inf. Theory (ISIT)*, Aachen, Germany, Jun. 2017.
- [2] X. Chen, T.-Y. Chen, and D. Guo, "Capacity of Gaussian many-access channels," *arXiv preprint arXiv:1607.01048*, 2016.

- [3] G. Reeves and H. D. Pfister, "The replica-symmetric prediction for compressed sensing with gaussian matrices is exact," in *Information Theory (ISIT), 2016 IEEE International Symposium on*. IEEE, 2016, pp. 665–669.
- [4] S. Kowshik, K. Andreev, A. Frolov, and Y. Polyanskiy, "Energy-efficient random access for the quasi-static fading mac," in *Proc. ISIT'2019*. IEEE, 2019.
- [5] S. Kowshik and Y. Polyanskiy, "Quasi-static fading mac with many users and finite payload," in *Proc. ISIT'2019*. IEEE, 2019.
- [6] R. Calderbank and A. Thompson, "CHIRRUP: a practical algorithm for unsourced multiple access," *arXiv preprint arXiv:1811.00879*, 2018.
- [7] A. Fengler, G. Caire, P. Jung, and S. Haghighatshoar, "Massive MIMO Unsourced Random Access," *arXiv preprint arXiv:1901.00828*, 2019.
- [8] V. K. Amalladinne, A. Vem, D. K. Soma, K. R. Narayanan, and J.-F. Chamberland, "A coupled compressive sensing scheme for uncoordinated multiple access," *arXiv preprint arXiv:1809.04745*, 2018.
- [9] Y. Gordon, *On Milman's inequality and random subspaces which escape through a mesh in $\mathbb{R}^n$* . Springer, 1988.
- [10] M. Rudelson and R. Vershynin, "Sparse reconstruction by convex relaxation: Fourier and gaussian measurements," in *Information Sciences and Systems, 2006 40th Annual Conference on*. IEEE, 2006, pp. 207–212.
- [11] M. Stojnic, "Various thresholds for ℓ_1 -optimization in compressed sensing," *arXiv preprint arXiv:0907.3666*, 2009.
- [12] —, "Block-length dependent thresholds in block-sparse compressed sensing," *arXiv preprint arXiv:0907.3679*, 2009.
- [13] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, "The convex geometry of linear inverse problems," *Foundations of Computational Mathematics*, vol. 12, no. 6, pp. 805–849, 2012.
- [14] D. Amelunxen, M. Lotz, M. B. McCoy, and J. A. Tropp, "Living on the edge: A geometric theory of phase transitions in convex optimization," *arXiv preprint arXiv:1303.6672*, 2013.
- [15] M. Stojnic, "A framework to characterize performance of lasso algorithms," *arXiv preprint arXiv:1303.7291*, 2013.
- [16] —, "Upper-bounding ℓ_1 -optimization weak thresholds," *arXiv preprint arXiv:1303.7289*, 2013.
- [17] C. Thrampoulidis, S. Oymak, and B. Hassibi, "Regularized linear regression: A precise analysis of the estimation error," in *Proceedings of The 28th Conference on Learning Theory*, 2015.
- [18] C. Thrampoulidis, E. Abbasi, and B. Hassibi, "Precise error analysis of regularized m -estimators in high dimensions," *IEEE Transactions on Information Theory*, vol. 64, no. 8, pp. 5592–5628, 2018.
- [19] C. Thrampoulidis, W. Xu, and B. Hassibi, "Symbol error rate performance of box-relaxation decoders in massive mimo," *IEEE Transactions on Signal Processing*, vol. 66, no. 13, pp. 3377–3392, 2018.
- [20] C. Thrampoulidis, I. Zadik, and Y. Polyanskiy, "A simple bound on the ber of the map decoder for massive mimo systems," *accepted to ICASSP 2019, arXiv preprint arXiv:1903.03949*, 2019.
- [21] A. El Gamal and Y.-H. Kim, *Network information theory*. Cambridge university press, 2011.
- [22] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Minimum energy to send k bits with and without feedback," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 4880–4902, Aug. 2011.
- [23] J. Barbier and F. Krzakala, "Replica analysis and approximate message passing decoder for superposition codes," *arXiv preprint arXiv:1403.8024*, 2014.
- [24] D. Guo and S. Verdú, "Randomly spread CDMA: Asymptotics via statistical physics," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 1983–2010, 2005.
- [25] M. Bayati and A. Montanari, "The lasso risk for gaussian matrices," *IEEE Transactions on Information Theory*, vol. 58, no. 4, pp. 1997–2017, 2012.
- [26] W. Su, M. Bogdan, E. Candès *et al.*, "False discoveries occur early on the lasso path," *The Annals of Statistics*, vol. 45, no. 5, pp. 2133–2150, 2017.
- [27] S. Oymak, C. Thrampoulidis, and B. Hassibi, "The squared-error of generalized lasso: A precise analysis," *arXiv preprint arXiv:1311.0830*, 2013.