The *Vector in Subspace* Problem

Bo'az Klartag

Tel-Aviv University

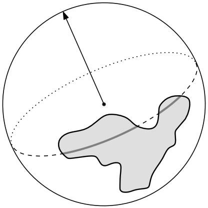
Workshop at Oberwolfach, Germany, May 2011

Joint work with Oded Regev.

Sampling Sets by Subspaces

Suppose $A \subset S^{n-1}$ is an arbitrary subset with $\sigma(A) = \varepsilon$.
Randomly select a subspace $E \subset \mathbb{R}^n$ of dimension k.

We intersect the fixed set $A \subset S^{n-1}$ with the random subspace $E \in G_{n,k}$.



- *What can we say about the distribution of*

$$\sigma_E(A \cap E)$$

(σ_E is the uniform probability measure on $S^{n-1} \cap E$).

Clearly,

$$\mathbb{E}\sigma_E(A \cap E) = \sigma(A).$$

Sampling Theorem

A particular case of our theorem, important for applications:

Theorem (K., Regev '10)

Suppose that $A \subseteq S^{n-1}$ satisfies $\sigma(A) \geq C\exp(-cn^{1/3})$. Suppose that $E \in G_{n,n/2}$ is a random subspace. Then,

$$\mathbb{P}\left\{\left|\frac{\sigma_E(A\cap E)}{\sigma(A)} - 1\right| \geq \frac{1}{10}\right\} \leq C\exp(-cn^{1/3}).$$

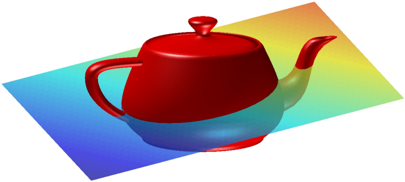
Here, $c, C > 0$ are universal constants.

- The theorem is optimal (i.e., you can't improve the $\exp(-n^{1/3})$'s).
- Tradeoff between parameters (size of A, dimension of E, probability estimate, deviation from one).

Lecture Outline

The rest of the talk is divided into two parts:

1. Related results, an application to computer science, comparison with Dvoretzky's theorem.
2. Proof of the theorem: Uses martingale bounds, the spherical Radon transform, and estimates for distribution of polynomials on the sphere.



Previous Results

Fix a subset $A \subseteq S^{n-1}$, denote $\varepsilon = \sigma(A)$.
Suppose that $E \in G_{n,k}$ is a random subspace.

- Raz '99:

$$\mathbb{P}\left\{\left|\frac{\sigma_E(A \cap E)}{\sigma(A)} - 1\right| \geq \frac{1}{10}\right\} \leq \frac{C}{\varepsilon} \exp\left(-c\varepsilon^2 k\right).$$

- Improved by V. Milman, Wagner '03:

$$\mathbb{P}\left\{\left|\frac{\sigma_E(A \cap E)}{\sigma(A)} - 1\right| \geq \frac{1}{10}\right\} \leq C \exp\left(-c\varepsilon^2 k\right).$$

- These two bounds are useless when $\varepsilon \leq 1/\sqrt{n}$.
  Surprisingly, the true dependence on ε is only logarithmic:

$$\mathbb{P}\left\{\left|\frac{\sigma_E(A \cap E)}{\sigma(A)} - 1\right| \geq \frac{1}{10}\right\} \leq C \exp\left(-c\frac{k}{\log^2 \varepsilon}\right)$$

meaningful estimate when $\varepsilon \geq \exp(-n^c)$.

Application to Computer Science

Our main motivation comes from *Communication Complexity*.

The "Vector in Subspace" Problem

Suppose Alice has a vector $x \in S^{n-1}$. Bob has a subspace $E \in G_{n,n/2}$. We can guarantee that

$$\text{either } x \in E \text{ or } x \in E^{\perp}.$$

Their goal is to decide which possibility holds, communicating the least possible number of bits between them.

- What does it mean for a computer to "have a vector"? Say, suppose that Alice has a genie in the basement ("an oracle"), which immediately answers any finite question about the vector $x \in S^{n-1}$ (e.g., what is the k^{th} digit of the i^{th} coordinate). The genie can perform any computation instantaneously.

Communication Complexity

This question is not about computing power. In some sense, the problem is: How many "bits of communication" are there in the statement "$x \in E$", or in knowing $d(x, E)$ up to an error of 0.01.

The term "information" is usually used in science in the context of entropy of random variables (Boltzmann, Shannon). We will therefore avoid this word, and say "communication complexity".

- Alice and Bob are allowed to use randomness, as long as they give the right answer with probability greater than $2/3$, for **any** $x \in S^{n-1}$ and for **any** $E \in G_{n,n/2}$.

Theorem (Raz '99)

There is a protocol that uses $C\sqrt{n}$ bits.

Theorem (K., Regev '10 – the first non-trivial lower bound)

Any protocol requires the exchange of at least $cn^{1/3}$ bits.

Communication Complexity

- There is still a gap between $cn^{1/3}$ and $Cn^{1/2}$. Some ideas will be discussed later.
- Apparently, our lower bound has theoretical significance, as it shows the advantages of **quantum communication.**

A sketch of Raz's $\sqrt{n}$-protocol

Alice and Bob generate $e^{5\sqrt{n}}$ random points $x_1, x_2, \ldots \in S^{n-1}$, known to both of them. ("looks strange, but it's possible")

Alice sends Bob the index i of the vector x_i closest to x. Bob announces that "$x \in E$" iff

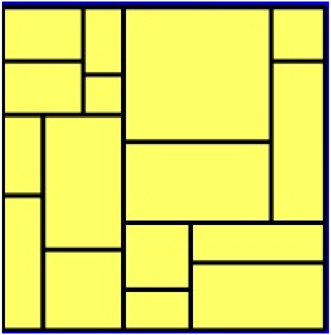
$$d(x_i, E) < d(x_i, E^{\perp}).$$

Not difficult to see that Bob is correct with prob. at least 95%.

What is a protocol?

What is a (deterministic) protocol of comm. complexity L? (i.e., L bits are exchanged between Alice and Bob)?

1. It induces a partition of the space $S^{n-1} \times G_{n,n/2}$ into 2^L combinatorial rectangles $A \times B$.
2. Each rectangle is marked with the decision: "$x \in E$" or "$x \in E^\perp$".



The Rectangle Bound

A standard technique for obtaining lower bounds

Find prob. measures μ_1, μ_2 on $S^{n-1} \times G_{n,n/2}$, such that

$$\mu_1\left(\left\{(x,E); x \in E\right\}\right) = 1, \qquad \mu_2\left(\left\{(x,E); x \in E^{\perp}\right\}\right) = 1$$

and such that for most rectangles $A \times B$,

$$\left|\frac{\mu_1(A \times B)}{\mu_2(A \times B)} - 1\right| \leq \frac{1}{10}$$

Set μ_0 to be the uniform measure on $S^{n-1} \times G_{n,n/2}$.
The measure μ_1 is uniform on $\{(x,E); x \in E\}$.
Our **sampling theorem** yields: If $\mu_0(A \times B) \geq \exp(-cn^{1/3})$,

$$0.9 \leq \frac{\mu_0(A \times B)}{\mu_1(A \times B)} \leq 1.1$$

Similarly for μ_2, which is uniform on $\{(x,E); x \perp E\}$.

Concluding the Computer Science application

- To conclude, assume by contradiction that the comm. complexity
$$L \leq cn^{1/3}.$$
The chances that the protocol will announce "$x \in E$" are roughly the same, no matter if the inputs to Alice and Bob are drawn according to μ_0, μ_1 or μ_2.

The main point for theoretical computer science, perhaps, is that the lower bound is a power of n, and not logarithmic in n.

- How to improve the lower bound from $cn^{1/3}$ to $Cn^{1/2}$? The sampling theorem is optimal. Perhaps it is true that when $\mu_0(A \times B) \geq \exp(-\sqrt{n})$,
$$\frac{\mu_1(A \times B) + \mu_2(A \times B)}{2} \geq 0.9\mu_0(A \times B).$$
We do not know.

Comparison with Dvoretzky's Theorem

Suppose we have a norm $\|\cdot\|$ on $\mathbb{R}^n$ with $\int_{S^{n-1}} \|x\| d\sigma(x) = 1$. Denote

$$\mathcal{A} = \left\{ x \in S^{n-1}; \left| \|x\| - 1 \right| \geq \frac{1}{10} \right\}.$$

Set $b = \sup_{x \in S^{n-1}} \|x\|$.

Theorem (Milman's version of Dvoretzky's theorem, '71)

Suppose $E \in G_{n,k}$ is a random subspace, where $k \leq cn/b^2$. Then,

$$\mathbb{P}\left\{ E \cap \mathcal{A} = \emptyset \right\} \geq 1 - Ce^{-ck}.$$

The subspace E usually escapes the "bad directions" in $\mathcal{A}$.

- In fact, according to Milman '71, Litvak, Milman, Schechtman '98: Assuming $b \geq 2$,

$$c \exp\left(-C \frac{n}{b^2} \right) \leq \sigma(\mathcal{A}) \leq C \exp\left(-c \frac{n}{b^2} \right).$$

Spherical Sets

- Therefore, the Dvoretzky-type theorem implies:

$$k \le c \log \frac{1}{\sigma(\mathcal{A})} \qquad \Rightarrow \qquad E \cap \mathcal{A} = \emptyset$$

  with probability at least $1 - C\exp(-ck)$ of selecting E. This uses special properties of $\mathcal{A}$ ("convexity of the norm").

- Our theorem says that for *any* subset $\mathcal{A} \subset S^{n-1}$,

$$k \ge C \log^2 \frac{1}{\sigma(\mathcal{A})} \qquad \Rightarrow \qquad \left| \frac{\sigma_E(\mathcal{A} \cap E)}{\sigma(\mathcal{A})} - 1 \right| \le \frac{1}{10},$$

  with probability at least $9/10$ of selecting $E \in G_{n,k}$.

Question

What happens between $\log \frac{1}{\sigma(\mathcal{A})}$ and $\log^2 \frac{1}{\sigma(\mathcal{A})}$?

Analysis of an example

Consider the following example. Let $\frac{1}{\sqrt{n}} \ll t \ll 1$ be a small parameter. Set

$$\mathcal{A}_t = \{x \in S^{n-1}; |x_1| \geq t\}.$$

Denote $R = \log \frac{1}{\sigma(\mathcal{A}_t)} \sim ct^2 n$, so $1 \ll R \ll n$.
Suppose $E \in G_{n,k}$ is a random subspace, $k \leq n/2$.

Dvoretzky-type regime

When $k \leq R$, with high probability $\mathcal{A}_t \cap E = \emptyset$.

The sampling regime

When $k \geq R^2$, usually $|\sigma_E(\mathcal{A}_t \cap E)/\sigma(\mathcal{A}_t) - 1| \leq 1/2$.

Both estimates are tight. So, what happens when $R \leq k \leq R^2$?

Analysis of an example, continued

- One computes that only when $k \geq R$, the distribution of

$$\log \frac{\sigma_E(\mathcal{A}_t \cap E)}{\sigma(\mathcal{A}_t)}$$

is approximately gaussian, with mean zero (only slightly negative), and with variance R^2/k.

1. First regime, $k \leq R = t^2 n$. With high prob. $\mathcal{A}_t \cap E = \emptyset$.
2. Intermediate regime $R \leq k \leq R^2$: Large fluctuations,

$$Var\left(\log \frac{\sigma_E(\mathcal{A}_t \cap E)}{\sigma(\mathcal{A}_t)}\right) \approx \frac{R^2}{k} \gg 1.$$

3. Only when $k \geq R^2$, we have good concentration, as the variance R^2/k is a small number.

Ideas of Proof

Theorem (K., Regev '10)

Let $A \subseteq S^{n-1}$. Denote $R = \log \frac{2}{\sigma(A)}$. Suppose that $E \in G_{n,k}$ is a random subspace. Then, for any $0 < t < 1$,

$$\mathbb{P}\left\{ \left| \frac{\sigma_E(A \cap E)}{\sigma(A)} - 1 \right| \geq t \right\} \leq C \exp\left(-c \frac{t^2 k}{R^2} \right).$$

Here, $c, C > 0$ are universal constants.

- The function $E \mapsto \sigma_E(A \cap E)$ is far from being Lipschitz, so hard to use standard concentration of measure.
- It seems to us that smoothing techniques don't help much in this respect.
- In the range $k = n - o(n)$, more precise estimates exist.

The case $k = n - 1$

Begin with the case where $k = n - 1$. Thus, suppose $H \subset \mathbb{R}^n$ is a random **hyperplane**.

Theorem (K., Regev '10)

Denote $R = \log \frac{2}{\sigma(A)}$. *Then, for* $0 < t < 1$,

$$\mathbb{P}\left\{ \left| \frac{\sigma_H(A \cap H)}{\sigma(A)} - 1 \right| \geq t \right\} \leq C \exp\left(-c \frac{tn}{R} \right).$$

- Exponential tail, standard deviation CR/n. Recall that for $k = n/2$ the tail was gaussian with std. dev. $CR/\sqrt{n}$.
- Bound is tight, as shown in the example above.

The proof relies on the **Radon Transform.** For $f : S^{n-1} \to \mathbb{R}$, and $\theta \in S^{n-1}$ set

$$\mathcal{R}(f)(\theta) = \int_{S^{n-1} \cap \theta^\perp} f(x) d\sigma_{\theta^\perp}(x).$$

Radon Transform

An equivalent formulation of the hyperplane-sampling theorem:

$$\sigma\left\{\theta \in S^{n-1} ; |\mathcal{R}(f)(\theta) - 1| \geq t\right\} \leq C \exp\left(-c\frac{tn}{R}\right),$$

where $f = 1_A/\sigma(A)$, $R = \log(2/\sigma(A))$.

- Again, concentration of Lipschitz functions seems irrelevant, tail is exponential and not gaussian.

Take a test-set $B \subset S^{n-1}$. Equivalently, we need to prove

$$\left|\int_B \mathcal{R}(f)(\theta)\frac{d\sigma(\theta)}{\sigma(B)} - 1\right| \leq C\frac{R\log\frac{2}{\sigma(B)}}{n}$$

assuming RHS is smaller than $1/2$ (i.e., “quantiles grow logarithmically”).

Some Harmonic Analysis

We arrived at an equivalent symmetric statement:

Theorem (for any non-negative functions f, g on the sphere)

$$\left| \int_{S^{n-1}} \mathcal{R}(f) g d\sigma - 1 \right| \leq C \frac{RT}{n}$$

whenever $RT \leq cn$, where $\int f = \int g = 1$, and

$$R = \log\left(2\|f\|_\infty\right), \quad T = \log\left(2\|g\|_\infty\right).$$

- The Radon transform commutes with rotations. Therefore it is diagonal in the basis of spherical harmonics.

The eigenvalues λ_k of $\mathcal{R}$, corresponding to spherical harmonics of degree k, are approximately

$$1, 0, -\frac{1}{n}, 0, \frac{1}{n^2}, 0, -\frac{1}{n^3}, 0, \ldots$$

Harmonic Analysis

- Note how quickly $|\lambda_{2k}|$ decays! The Radon transform does a lot of **smoothing**. It resembles the smoothing done by the heat kernel (for time $t \approx \log n$).

Therefore, for any $f, g : S^{n-1} \to \mathbb{R}$ with $\int f = \int g = 1$,

$$\left| \int_{S^{n-1}} \mathcal{R}(f) g d\sigma - 1 \right| \leq \sum_{k=1}^{\infty} |\lambda_{2k}| \|f_{2k}\|_2 \|g_{2k}\|_2$$
$$\lesssim \sum_{k=1}^{n} \left(\frac{Ck}{n} \right)^k \|f_{2k}\|_2 \|g_{2k}\|_2$$

where $f = \sum_k f_k$ and $g = \sum_k g_k$ are decompositions into spherical harmonics. To conclude, it is enough to show that

$$\|f_{2k}\|_2 \leq \left(C \frac{\log(2\|f\|_\infty)}{k} \right)^k,$$

and similarly for g.

Distribution of Polynomials over the Sphere

Thus, in order to prove the theorem for **hyperplanes**, all that remains is to prove

Lemma

Suppose $f : S^{n-1} \to \mathbb{R}$, $\|f\|_1 = 1$, $\|f\|_\infty = M$. Then for any spherical harmonic φ_d of degree $d \le \log M$ with $\|\varphi_d\|_2 = 1$,

$$\left| \int_{S^{n-1}} \varphi_d f d\sigma \right| \le \left(C \frac{\log M}{d} \right)^{d/2}.$$

The extremal case, up to factor 2, is when $f = 1_A / \sigma(A)$.

1. Suppose $d = 1$. Then φ_1 is a linear functional on S^{n-1}, which has a sub-gaussian tail, so we get at most $C\sqrt{\log M}$.
2. Roughly, we need to show that the tail distribution of φ_d is of the form $C \exp(-ct^{2/d})$.

Kahn-Kalai-Linial or Needle Decomposition

- How can you prove that for any polynomial $p : \mathbb{R}^n \to \mathbb{R}$ of degree d with $\|p\|_2 = 1$,

$$\sigma \left\{ \theta \in S^{n-1} ; |p(\theta)| \geq t \right\} \leq C \exp(-ct^{2/d})$$

Two amusing non-direct methods:

Option 1

Use log-Sobolev inequality on the sphere (Bakry-Émery '85, Rothaus '86) and hyper-contractivity of heat semigroup (Gross '75). This is the approach suggested by Kahn-Kalai-Linial '88. Only for spherical harmonics. "Quick, mysterious proof".

Option 2

Apply Needle Decomposition on the sphere and use Remez-type Inequality, as in Gromov-Milman '86, Kannan-Lovász-Simonovits '95, Bobkov '00, Carbery-Wright '01, Nazarov-Sodin-Volberg '03. "A bit messy, but clear."

Iterating the hyperplane theorem

We completed the proof of

Theorem

Let $A \subset S^{n-1}$. Suppose $H \subset \mathbb{R}^n$ is a random hyperplane. Denote $R = \log \frac{2}{\sigma(A)}$. Then, for $0 < t < 1$,

$$\mathbb{P}\left\{\left|\frac{\sigma_H(A \cap H)}{\sigma(A)} - 1\right| \geq t\right\} \leq C \exp\left(-c\frac{tn}{R}\right).$$

We still need to analyze $\sigma_E(A \cap E)$ for a random k-dimensional subspace E. Select a flag of random subspaces

$$\mathbb{R}^n = H_0 \supset H_1 \supset H_2 \supset \cdots \supset H_{n-k}$$

where $\dim(H_i) = n - i$. Consider the martingale

$$X_\ell = \sigma_{H_\ell}(A \cap H_\ell).$$

Martingale Inequalities

Clearly,

$$\mathbb{E}(X_\ell | H_1, \ldots, H_{\ell-1}) = X_{\ell-1}.$$

Furthermore, by the hyperplane-sampling theorem,

$$\mathbb{P}\left(\left| \frac{X_\ell}{X_{\ell-1}} - 1 \right| \geq t \right) \leq C \exp\left(-c \frac{(n-\ell)t}{\log(1/X_{\ell-1})} \right).$$

- We need to estimate large deviations of X_{n-k}/X_0. Use:

Theorem (Bernstein's Inequality '37)

Suppose $\mathbb{E}(S_\ell | S_1, \ldots, S_{\ell-1}) = S_{\ell-1}$, *and*

$$\forall t, \quad \mathbb{P}\left(|S_\ell - S_{\ell-1}| \geq t \mid S_1, \ldots, S_{\ell-1}\right) \leq 2\exp\left(-t/R\right).$$

Then, for any $|t| \leq \sqrt{n}R$,

$$\mathbb{P}\left(|S_n - S_0| > t\right) \leq C \exp\left(-ct^2/(nR^2) \right).$$

Odds and Ends

A few remarks on the proof:

- We cannot apply Bernstein's theorem as is. Yet, a straightforward adaptation of the proof yields what we need.
- The main message: The logarithmic increments $\log X_\ell - \log X_{\ell-1}$ have an exponential tail. Therefore $\log X_\ell$ has a sub-gaussian tail, up to $\sqrt{\ell}$ standard deviations.

Question about proof strategy

Why do we use harmonic analysis for hyperplane-sampling, and then iterate to get subspace-sampling? Can't you do harmonic analysis directly on $G_{n,k}$?

Partial Answer: Yes, you can. Our straightforward attempt provided an inferior estimate (in the CS problem, only $Cn^{1/4}$ in place of $cn^{1/3}$). The main difficulty: We don't know enough about the range of the Radon transform in $G_{n,k}$.

Thank you!