Sampling and high-dimensional convex geometry

Roman Vershynin

University of Michigan

SampTA 2013
Bremen, Germany, June 2013
Geometry of sampling problems

Signals live in high dimensions; sampling is often random.

Geometry of sampling problems

Signal recovery problem.

Signal: \( x \in \mathbb{R}^n \). Unknown.

Sampling (measurement) map: \( \Phi : \mathbb{R}^n \rightarrow \mathbb{R}^m \). Known.

Measurement vector: \( y = \Phi(x) \in \mathbb{R}^m \). Known.

Goal: recover \( x \) from \( y \).
Geometry of sampling problems

Prior information (model): $x \in K$, where $K \subset \mathbb{R}^n$ is a known signal set.

Examples of $K$:

Discrete $L_p$, i.e. $\ell_p^n$, for $0 < p < \infty$

{band-limited functions}

{s-sparse signals in $\mathbb{R}^n$}

{block-sparse signals}

{low-rank matrices}

etc.
Geometry of sampling problems

**Linear sampling:** \( y = Ax \), where \( A \) is an \( m \times n \) matrix.

\[ \begin{array}{c}
\uparrow \\
m
\end{array} \quad \begin{array}{c}
y \\
= \\
\downarrow \\
m
\end{array} \quad \begin{array}{c}
A \\
\end{array} \quad \begin{array}{c}
x \\
\end{array} \]

**Non-linear sampling:** later.
Convexity

Signal set $K$ may be non-convex. Then convexity: $K \mapsto \text{conv}(K)$.

**Question.** What do convex sets look like in high dimensions?

This is the main question of **asymptotic convex geometry**
$=$ geometric functional analysis.
Asymptotic convex geometry

Main message of asymptotic convex geometry:

\[ K \approx \text{bulk} + \text{outliers}. \]

Bulk = round ball, makes up most volume of \( K \).
Outliers = few long tentacles, contain little volume.

V. Milman’s heuristic picture of a convex body in high dimensions
Concentration of volume

“The volume of an isotropic convex set is concentrated in a round ball.”

Isotropy assumption: $X \sim \text{Unif}(K)$ satisfies $\mathbb{E} X = 0$, $\mathbb{E} XX^\top = I_n$.

$\mathbb{E} \|X\|_2^2 = n$, so the radius of that ball is $\sqrt{n}$.

**Theorem (concentration of volume)** [Paouris ’06]

$$\mathbb{P}\{\|X\|_2 > t\sqrt{n}\} \leq \exp(-ct\sqrt{n}), \quad t \geq 1.$$  

A simpler proof: [Adamczak-Litvak-Oleszkiewicz-Pajor-Tomczak ’13]
Concentration of volume

Theorem (thin shell) [Klartag ’08]

\[ ||X||_2 - \sqrt{n} \ll \sqrt{n} \quad \text{with high probability.} \]

Question. What is the best bound?

Currently best known: \( \lesssim n^{1/3} \) [Guedon-E.Milman ’11].

Thin Shell Conjecture: \( \lesssim \) absolute constant.
This would imply the Hyperplane Conjecture [Eldan-Klartag ’11].
Round sections

A random subspace $E$ misses the outliers, **passes through the bulk** of $K$.

**Theorem (Dvoretzky’s theorem)**

Consider a random subspace $E$ in $\mathbb{R}^n$ of dimension $\sim \log n$. Then $K \cap E \approx \text{round ball}$ with high probability.

**Remark.** For many convex sets $K$, the **dimension** of $E$ is much larger than $\log n$. See Milman’s version of Dvoretzky’s theorem.
Sections of arbitrary dimension

$K \subset \mathbb{R}^n$ any convex set.

**Theorem ($M^*$ estimate)** [Milman ’81, Pajor-Tomczak ’85, Mendelson-Pajor-Tomczak ’07]

Consider a random subspace $E$ in $\mathbb{R}^n$ of codimension $m$. Then

$$\text{diam}(K \cap E) \leq \frac{C \cdot w(K)}{\sqrt{m}}$$

with high probability.

Here $w(K)$ is the mean width of $K$. 
Mean width

\[ w(K) := \mathbb{E} \sup_{x \in K - K} \langle g, x \rangle, \quad \text{where } g \sim N(0, I_n). \]

Note: \[ \|g\|_2 \sim \sqrt{n}, \] so

\[ w(K) = \sqrt{n} \cdot \text{width of } K \text{ in a random direction.} \]
Example: the $\ell_1$ ball

$$K = \text{conv}(\pm e_i) = \{x : \|x\|_1 \leq 1\} = B^n_1.$$  

The standard picture:

$$\text{Vol}(K)^{1/n} \sim \text{Vol}(\bullet)^{1/n} \sim \frac{1}{n}.$$  

Hence a more accurate picture is this:
Example: the $\ell_1$ ball

Mean width: $w(K) := \mathbb{E} \sup_{x \in K-K} \langle g, x \rangle = \sqrt{n} \cdot \mathbb{E} \left[ \text{width of } K \text{ in random direction} \right]$.

$w(K) \sim \sqrt{\log n}$, $w(\bullet) \sim 1$.

Hence mean width sees the bulk, ignores the outliers.
Signal recovery (following [Mendelson-Pajor-Tomczak '07])

Back to the **signal recovery problem** (linear sampling):

Recover signal $x \in K \subset \mathbb{R}^n$ from $y = Ax \in \mathbb{R}^m$.

If $m \leq n$, problem is **ill-posed**.

What do we know? $x$ belongs to both $K$ and the affine subspace

$$E_x := \{x' : Ax' = y\} = x + \ker(A).$$

Both $K$ and $E_x$ are **known**.

If $\text{diam}(K \cap E_x) \leq \varepsilon$ then $x$ can be recovered with error $\varepsilon$. 
Signal recovery

Assume $K$ is convex and origin-symmetric. Then diam is largest for $x = 0$.

**Conclusion.** Let $E = \ker(A)$. If $\text{diam}(K \cap E) \leq \varepsilon$ then any $x \in K$ can be recovered from $y = Ax$ with error $\varepsilon$.

The recovery can be done by a **convex program**

\[
\text{Find } x' \in K \text{ such that } Ax' = y.
\]

“Find a signal consistent with the model ($K$) and with measurements ($y$).”
Remaining question: when is \( \text{diam}(K \cap E) \leq \varepsilon \)?

\( E = \ker(A) \). If \( A \) is a random matrix then \( E \) is a random subspace. So the \( M^* \) estimate applies and yields

\[
\text{diam}(K \cap E) \leq \frac{C \omega(K)}{\sqrt{m}} \quad \text{with high probability.}
\]

Equate with \( \varepsilon \) and obtain the **sample size** (\# of measurements):

\[
m \sim \varepsilon^{-2} \omega(K)^2.
\]
Conclusion (Signal recovery)

Let $K$ be a convex and origin-symmetric signal set in $\mathbb{R}^n$. Let $A$ be an $m \times n$ random matrix. If $m \sim w(K)^2$ then one can accurately recover any signal $x \in K$ from $m$ random linear measurements given as $y = Ax \in \mathbb{R}^m$.

The recovery is done by the **convex program**

$$\text{Find } x' \in K \text{ such that } Ax' = y$$

i.e. “find $x'$ consistent with the model ($K$) and measurements ($y$”).

**Other natural recovery programs:**

1. $\max \langle Ax', y \rangle$ subject to $x' \in K$

i.e. “find $x \in K$ most correlated with the measurements”.

2. find $x' \in K$ closest to $A^*y$

i.e. the *metric projection* of $A^*y$ onto $K$. 
Remarks.

1. If the signal set $K$ is not convex, then convexify;

   $w(\text{conv}(K)) = w(K)$.

2. If the signal set $K$ is not origin-symmetric, then symmetrize;

   $w(K - K) \leq 2w(K)$.

3. Mean width can be efficiently estimated. Randomized linear program:

   $w(K) \approx \sup_{x \in K - K} \langle g, x \rangle, \quad g \sim N(0, I_n)$. 
Information theory viewpoint

The sample size $m = w(K)^2$ is an effective dimension of $K$, the amount of information in $K$.

**Conclusion:** One can effectively recover any signal in $K$ from $w(K)^2$ random linear measurements.

**Example 1.** $K = B_1^n = \text{conv}(\pm e_i)$. $w(K)^2 \sim \log n$.

So, one can effectively recover any signal in $B_1^n$ from $\log n$ random linear measurements.

**Remark.** $\log n$ bits are required to specify even a vertex of $K$. So the number of measurements is optimal.
Information theory viewpoint

Example 2. \( K = \{ s\text{-sparse unit vectors in } \mathbb{R}^n \} \). \( w(K)^2 \sim s \log n \).

**Conclusion:** One can effectively recover any \( s \)-sparse signal from \( s \log n \) random linear measurements.

This is a well-known result in **compressed sensing**. (Warning: exact recovery is not explained by this geometric reasoning.)

**Remark.** \( \log \binom{n}{s} \sim s \log n \) bits are required to specify the sparsity pattern.
Remark. The mean width is related to other information-theoretic quantities, in particular to the metric entropy:

\[ c \cdot \max_{t > 0} t \sqrt{\log N(K, t)} \leq w(K) \leq C \cdot \int_0^\infty \sqrt{\log N(K, t)} \, dt. \]

Sudakov’s and Dudley’s inequalities.
Non-linear sampling

Claim.
One can recover a signal $x \in K$ from non-linear measurements $y = \Phi(x)$, and even without fully knowing the nature of non-linearity $\Phi$. 
Non-linear sampling

Non-linear measurements are given as

$$y = \theta(Ax).$$

Here $A$ is an $m \times n$ random matrix (known), $\theta : \mathbb{R} \rightarrow \mathbb{R}$ is a link function (possibly unknown).

$\theta$ is applied to each coordinate of $Ax$, i.e. the measurements are

$$y_i = \theta(\langle a_i, x \rangle), \quad i = 1, \ldots, m.$$
Non-linear sampling

Examples.

1. **Quantization.** Single-bit compressed sensing: $\theta(t) = \text{sign}(t)$.
   [Plan-Vershynin ’11]

2. **Generalized Linear Models (GLM) in Statistics.** $\theta(t)$ is arbitrary, perhaps unknown.
   In particular, logistic regression.
   [Plan-Vershynin ’12]
Probabilistic Heuristics

How is reconstruction from non-linear measurements possible?

Probabilistic heuristic. 1. Linear.

Let $a \sim N(0, I_n)$. Then

$$\mathbb{E}\langle a, x \rangle\langle a, y \rangle = \langle x, y \rangle \text{ for all } x, y \in \mathbb{R}^n.$$ 

Both sides are linear in $y$, so

$$\mathbb{E}\langle a, x \rangle a = x \text{ for all } x \in \mathbb{R}^n.$$ 

Law of Large Numbers $\Rightarrow$

$$\frac{1}{m} \sum_{i=1}^{m} \langle a_i, x \rangle a_i \approx x \text{ if } m \text{ is large.}$$
Probabilistic Heuristics

\[ \frac{1}{m} \sum_{i=1}^{m} \langle a_i, x \rangle a_i \approx x \quad \text{if } m \text{ is large.} \]

One can interpret this as:

\( \{ a_i \}_{i=1}^{m} \) is a frame in \( \mathbb{R}^n \). “Gaussian frame”.

One can reconstruct \( x \) from linear measurements \( y = Ax \), where

\[ y_i = \langle a_i, x \rangle, \quad i = 1, \ldots, m \]

**Question.** How large does \( m \) need to be?

- Without any prior knowledge about \( x \), \( m \sim n \) suffices.
- If \( x \in K \), then \( m \sim w(K)^2 \), the effective dimension of \( K \).

Reconstruction: metric projection of \( \frac{1}{m} \sum_{i=1}^{m} \langle a_i, x \rangle a_i = A^*y \) onto \( K \).
Probabilistic Heuristics

Probabilistic heuristic. 2. Non-linear.

Let \( a \sim N(0, I_n) \). Then

\[
\mathbb{E} \theta(\langle a, x \rangle) \langle a, y \rangle = \lambda \langle x, y \rangle \quad \text{for all unit } x, y \in \mathbb{R}^n
\]

where \( \lambda = \mathbb{E} \theta(g)g =: 1, \ g \in N(0, 1) \). Still linear in \( y \Rightarrow \)

\[
\mathbb{E} \theta(\langle a, x \rangle) a = x \quad \text{for all unit } x \in \mathbb{R}^n.
\]

Law of Large Numbers \( \Rightarrow \)

\[
\frac{1}{m} \sum_{i=1}^{m} \theta(\langle a_i, x \rangle) a_i \approx x \quad \text{if } m \text{ is large}.
\]
Probabilistic Heuristics

\[
\frac{1}{m} \sum_{i=1}^{m} \theta(\langle a_i, x \rangle) a_i \approx x \quad \text{if } m \text{ is large.}
\]

One can interpret this as:

\[{a_i}_{i=1}^{m} \text{ is a non-linear frame in } \mathbb{R}^n.\]

One can reconstruct \(x\) from non-linear measurements \(y = \theta(Ax)\), where

\[y_i = \theta(\langle a_i, x \rangle), \quad i = 1, \ldots, m\]

For the reconstruction, one does not need to know \(\theta\)!

(at least, in principle).
Single-bit measurements

Let us work out the particular example of single-bit measurements,

\[ y_i = \text{sign}(\langle a_i, x \rangle), \quad i = 1, \ldots, m. \]

**Geometric interpretation:**
\[ y = \text{vector of orientations} \] of \( x \) with respect to \( m \) random hyperplanes (with normals \( a_i \)).

In stochastic geometry: **random hyperplane tessellation**.
Single-bit measurements

We know \( y = \text{sign}(Ax) \), i.e. \( y_i = \langle a_i, x \rangle \) \( \forall i \implies \) we know the cell \( \ni x \).

If \( \text{diam(cell)} \leq \varepsilon \) \( \forall \text{ cell} \), then we could recover \( x \) with error \( \varepsilon \).

Recovery would be done by the convex program

\[
\text{Find } x' \in K \text{ such that } \text{sign}(Ax') = y.
\]

It remains to find \( m \) so that all cells have diameter \( \leq \varepsilon \).
Question (Pizza cutting). Given a set $K \subset \mathbb{R}^n$, how many random hyperplanes does it take in order to cut $K$ in pieces of diameter $\leq \varepsilon$?

Non-trivial even for $K = S^{n-1}$. 
**Theorem (Pizza cutting)** [Plan-V ’12]. Consider a set $K \subset S^{n-1}$ and $m$ random hyperplanes. Then, with high probability,

$$
\text{diam}(\text{cell}) \leq \left[ \frac{C \, w(K)}{\sqrt{m}} \right]^{1/3} \quad \forall \text{ cell}.
$$

This is similar to the $M^*$ estimate for $\text{diam}(K \cap \text{random subspace})$, except for the exponent $1/3$. **Optimal exponent = ?**

**Corollary (Single-bit recovery).** One can accurately and effectively recover any signal $x \in K$ from $m \sim w(K)^2$ random measurements $y = \text{sign}(Ax) \in \{-1, 1\}^m$. 
General non-linear measurements

Similar recovery results hold for general link function \( \theta(\cdot) \),

\[
y = \theta(Ax).
\]

Recovery of \( x \) can be done by the convex program

\[
\max \langle y, Ax' \rangle \quad \text{subject to} \quad x' \in K.
\]
i.e. “find \( x \in K \) most correlated with the measurements”.

**Remark.** The solver does not need to know the nature of non-linearity given by \( \theta \). It may be unknown or unspecified.

[Plan-V '12]
Summary: geometric view of signal recovery

Recover signal $x \in K$ from $m$ random measurements/samples

$$y = Ax \quad \text{(linear)}, \quad y = \theta(Ax) \quad \text{(non-linear)}.$$  

Convex sets look like this: $K \approx \text{bulk} + \text{outliers}$

The size of the bulk determines the accuracy of recovery.

Accurate and efficient recovery is possible if

$$m \sim w(K)^2 = \text{the effective dimension of } K.$$  

Here $w(K)$ is the mean width of $K$, a computable quantity.
Further reading:

- **Elementary introduction** into modern convex geometry [Ball ’97]
- **Lectures** in geometric functional analysis [Vershynin ’11]
- **Notes** on isotropic convex bodies
  [Brazitikos-Giannopoulos-Valettas-Vritsiou ’13+]
- **Textbook** in geometric functional analysis
  [Artstein-Giannopoulos-Milman-Vritsiou ’13+]
- Further on non-linear recovery: **single-bit matrix completion**
  [Davenport-Plan-van den Berg-Wootters ’12]

Thank you!