# Randomness in reduced order modeling

Akil Narayan[1]

[1]Department of Mathematics, and Scientific Computing and Imaging (SCI) Institute
University of Utah

February 7, 2020
ICERM


# Randomness is your friend

Many things that are difficult to accomplish with deterministic optimization/algorithms can be accomplished* with randomization.

# Randomness is your friend

Many things that are difficult to accomplish with deterministic optimization/algorithms can be accomplished* with randomization.

*: with "high" probability

# Randomness is your friend

Many things that are difficult to accomplish with deterministic optimization/algorithms can be accomplished* with randomization.

*: with "high" probability

We'll consider three examples of this in ROM:

- RBM for elliptic PDE's
- Sparse approximation
- Measure atomization/discretization

# Randomness is your friend

Many things that are difficult to accomplish with deterministic optimization/algorithms can be accomplished* with randomization.

*: with "high" probability

We'll consider three examples of this in ROM:

- RBM for elliptic PDE's
- Sparse approximation
- ~~Measure atomization/discretization~~

# Why is randomness helpful?

Intuition is straightforward and simplistic: Let $X$ be a random variable.

Let $(X_m)_{m \geqslant 1}$ be iid copies of $X$. Law of large numbers:

$$S := \sum_{m=1}^{M} X_m \to \mathbb{E}[X].$$

# Why is randomness helpful?

Intuition is straightforward and simplistic: Let $X$ be a random variable.

Let $(X_m)_{m \geqslant 1}$ be iid copies of $X$. Law of large numbers:

$$S := \sum_{m=1}^{M} X_m \to \mathbb{E}[X].$$

Furthermore, this convergence is quantitative through the Central limit theorem:

$$S(M) - \mathbb{E}[X] \sim \mathcal{N}\left(0, \frac{\sigma^2(X)}{M}\right).$$

In other words, $S$ concentrates around $\mathbb{E}[X]$.

# Why is randomness helpful?

Intuition is straightforward and simplistic: Let $X$ be a random variable.

Let $(X_m)_{m \geqslant 1}$ be iid copies of $X$. Law of large numbers:

$$S := \sum_{m=1}^{M} X_m \to \mathbb{E}[X].$$

Furthermore, this convergence is quantitative through the Central limit theorem:

$$S(M) - \mathbb{E}[X] \sim \mathcal{N}\left(0, \frac{\sigma^2(X)}{M}\right).$$

In other words, $S$ concentrates around $\mathbb{E}[X]$.

This statement is quite powerful:

- $S$ provides an estimator for $\mathbb{E}[X]$, without knowing $\mathbb{E}[X]$.
- Convergence is essentially *independent* of distribution of $X$.
- Convergence rate is *independent* of *dimension* of $X$.

# Examples of concentration

Concentration in general plays an important role in computing estimates:

- Monte Carlo (CLT) estimates
- Chebyshev inequalities (bounds on mass away from the mean)
- Hoeffding inequalities (bounds on deviation of iid sums from the mean)
- Chernoff bounds (bounds on deviation of spectrum)
- Concentration of measure (bounds on deviation of random functions)

Today: We'll see a particular Chernoff bound in action.

# Chernoff bound applications

We will see how randomization and Chernoff bounds can be applied to:

- RBM for elliptic PDE's
- Sparse approximation
- Measure atomization/discretization

Before discussing ROM, let's present the Chernoff bound.

# Matrix law of large numbers

Let $\boldsymbol{G} \in \mathbb{R}^{N \times N}$ be a Gramian matrix that is an iid sum of symmetric rank-1 matrices.

I.e., let $\boldsymbol{X} \in \mathbb{R}^N$ have distribution $\mu$ on $\mathbb{R}^N$, and define

$$\boldsymbol{G} := \frac{1}{M} \sum_{m=1}^{M} \boldsymbol{X}_m \boldsymbol{X}_m^T,$$

where $\{\boldsymbol{X}_m\}_{m \geqslant 1}$ are iid copies of $\boldsymbol{X}$.

Chernoff bounds make quantitative statements about the spectrum of $\boldsymbol{G}$ that depend on the distribution of $\boldsymbol{X}$.

# Matrix law of large numbers

Let $\boldsymbol{G} \in \mathbb{R}^{N \times N}$ be a Gramian matrix that is an iid sum of symmetric rank-1 matrices.

I.e., let $\boldsymbol{X} \in \mathbb{R}^N$ have distribution $\mu$ on $\mathbb{R}^N$, and define

$$\boldsymbol{G} := \frac{1}{M} \sum_{m=1}^{M} \boldsymbol{X}_m \boldsymbol{X}_m^T,$$

where $\{\boldsymbol{X}_m\}_{m \geqslant 1}$ are iid copies of $\boldsymbol{X}$.

Chernoff bounds make quantitative statements about the spectrum of $\boldsymbol{G}$ that depend on the distribution of $\boldsymbol{X}$.

For large $M$, we expect that

$$(G)_{j,k} \overset{M \uparrow \infty}{\longrightarrow} \mathbb{E}[X_j X_k].$$

# Matrix law of large numbers

Let $\boldsymbol{G} \in \mathbb{R}^{N \times N}$ be a Gramian matrix that is an iid sum of symmetric rank-1 matrices.

I.e., let $\boldsymbol{X} \in \mathbb{R}^N$ have distribution $\mu$ on $\mathbb{R}^N$, and define

$$\boldsymbol{G} := \frac{1}{M} \sum_{m=1}^{M} \boldsymbol{X}_m \boldsymbol{X}_m^T,$$

where $\{\boldsymbol{X}_m\}_{m \geqslant 1}$ are iid copies of $\boldsymbol{X}$.

Chernoff bounds make quantitative statements about the spectrum of $\boldsymbol{G}$ that depend on the distribution of $\boldsymbol{X}$.

For large $M$, we expect that

$$(G)_{j,k} \overset{M \uparrow \infty}{\longrightarrow} \mathbb{E}[X_j X_k].$$

For simplicity, in all that follows we assume that the components of $\boldsymbol{X}$ are

- uncorrelated,
- of unit variance,

so that

$$\boldsymbol{G} \overset{M \uparrow \infty}{\longrightarrow} \boldsymbol{I}$$

# Matrix Chernoff bounds

The proximity of $\boldsymbol{G}$ to $\boldsymbol{I}$, as a function of $M$, is determined by

$$K := \sup_{\boldsymbol{X}} \|\boldsymbol{X}\|_2,$$

which is assumed finite.

## Matrix Chernoff bounds

The proximity of $\boldsymbol{G}$ to $\boldsymbol{I}$, as a function of $M$, is determined by

$$K := \sup_{\boldsymbol{X}} \|\boldsymbol{X}\|_2,$$

which is assumed finite.

**Theorem ([Cohen, Davenport, Leviatan 2012])**

*Assume that*

$$\frac{M}{\log M} \gtrsim \frac{K}{\delta^2} \log\left(\frac{1}{\epsilon}\right).$$

*Then,*

$$\Pr\left[(\sigma_{\min}(\boldsymbol{G}) < 1-\delta) \bigcup (\sigma_{\max}(\boldsymbol{G}) > 1+\delta)\right] \leqslant \epsilon.$$

## Matrix Chernoff bounds

The proximity of $\boldsymbol{G}$ to $\boldsymbol{I}$, as a function of $M$, is determined by

$$K := \sup_{\boldsymbol{X}} \|\boldsymbol{X}\|_2,$$

which is assumed finite.

**Theorem ([Cohen, Davenport, Leviatan 2012])**

*Assume that*

$$\frac{M}{\log M} \gtrsim \frac{K}{\delta^2} \log\left(\frac{1}{\epsilon}\right).$$

*Then,*

$$\Pr\left[(\sigma_{\min}(\boldsymbol{G}) < 1-\delta) \bigcup (\sigma_{\max}(\boldsymbol{G}) > 1+\delta)\right] \leqslant \epsilon.$$

What can we do with $\boldsymbol{G}$? Form least-squares approximations using $\boldsymbol{X}$.
Remarks:

- The $\delta^{-2}$ dependence is "CLT-like".
- $K$ is the only thing that depends on the distribution of $X$.

# The induced distribution

It turns out that $K$ can be quite large (or infinite) for practical situations.

A fix for this utilizes *importance sampling*. In particular, define

$$\mathrm{d}\rho(x) := \left( \frac{1}{N} \sum_{n=1}^{N} x_n^2 \right) \mathrm{d}\mu(x),$$

where $\mu$ is the distribution of $\boldsymbol{X}$.

$\rho$ is a probability measure on $\mathbb{R}^N$, and is frequently called the *induced distribution*.

# A (more) optimal Chernoff bound

In practical scenarios, the induced distribution $\rho$ can also be sampled from without too much effort.

More importantly, we can get a (much) better Chernoff bound here.

Let $(\boldsymbol{Y}_m)_{m\geqslant 1} \in \mathbb{R}^N$ be iid samples from $\rho$. We need to weight the Gramian so that we produce an unbiased estimate:

$$\boldsymbol{F} := \frac{1}{M}\sum_{m=1}^{M} w_m \boldsymbol{Y}_m \boldsymbol{Y}_m^T, \qquad\qquad w_m := \frac{\mathrm{d}\mu}{\mathrm{d}\rho}(\boldsymbol{Y}_m)$$

# A (more) optimal Chernoff bound

In practical scenarios, the induced distribution $\rho$ can also be sampled from without too much effort.

More importantly, we can get a (much) better Chernoff bound here.

Let $(\boldsymbol{Y}_m)_{m\geqslant 1} \in \mathbb{R}^N$ be iid samples from $\rho$. We need to weight the Gramian so that we produce an unbiased estimate:

$$\boldsymbol{F} := \frac{1}{M}\sum_{m=1}^{M} w_m \boldsymbol{Y}_m \boldsymbol{Y}_m^T, \qquad\qquad w_m := \frac{\mathrm{d}\mu}{\mathrm{d}\rho}(\boldsymbol{Y}_m)$$

This results in the (better) Chernoff bound

$$\Pr\left[(\sigma_{\min}(\boldsymbol{F}) < 1-\delta)\bigcup(\sigma_{\min}(\boldsymbol{F}) > 1+\delta)\right] \leqslant \epsilon,$$

with the much more reasonable assumption

$$\frac{M}{\log M} \gtrsim \frac{N}{\delta^2}\log\left(\frac{1}{\epsilon}\right).$$

This Chernoff bound will be a seed for achieving model reduction.

# Example 1: RBM (for elliptic problems)

# Reduced basis methods

For the parameterized problem,

$$-\nabla \cdot \left( \sum_{j=1}^{\infty} \mu_j a_j(x) \nabla u \right) = b,$$

with $\mu \in [-1,1]^{\infty}$, recall that RBM (essentially) iteratively computes

$$\arg\max_{\mu} \left\| u(\mu) - \mathcal{P}_{j-1}(u(\mu)) \right\|,$$

# Reduced basis methods

For the parameterized problem,

$$-\nabla \cdot \left( \sum_{j=1}^{\infty} \mu_j a_j(x) \nabla u \right) = b,$$

with $\mu \in [-1,1]^{\infty}$, recall that RBM (essentially) iteratively computes

$$\arg\max_{\mu} \left\| u(\mu) - \mathcal{P}_{j-1}(u(\mu)) \right\|,$$

If (any truncation of) $\mu$ is high-dimensional, this is an expensive optimization, even if the objective is easy to evaluate.

There's a bigger problem: the $\arg\max$ is typically taken over a discrete $\mu$ grid. If $\mu$ is high-dimensional, how can we certify error without densely sampling?

# Reduction feasibility

Some analysis gives us a strategy to proceed: if the $\{a_j\}_{j=1}^d$ satisfies an $\ell^p$ summability condition,

$$\sum_{j=1}^{\infty} \|a_j\|_{L^\infty}^p < \infty, \qquad\qquad p < 1,$$

then there is an $N$-dimensional *downward-closed* polynomial space $P_N$ in the variable $\mu$ such that

$$\sup_{\mu} \left\| u(\mu) - \mathrm{Proj}_{P_N} u(\mu) \right\| \leqslant N^{-s}, \qquad\qquad s := \frac{1}{p} - \frac{1}{2}.$$

There are constructive algorithms to essentially identify $P_N$, [Cohen, Devore, Schwab 2011].

In particular, once $P_N$ is identified, this approximation can be obtained by $\mu$-least-squares approximation.

# Polynomial meshes

I.e., if we can certify accuracy on a "*polynomial* grid", we can probably obtain accuracy.

Let $\mu$ be a random variable with distribution $\nu$.

# Polynomial meshes

I.e., if we can certify accuracy on a "*polynomial* grid", we can probably obtain accuracy.

Let $\mu$ be a random variable with distribution $\nu$. Let $\boldsymbol{X} = (X_n(\nu))_{n=1}^N$ denote a $\mathrm{d}\nu$-orthonormal basis for $P_N$. Define the induced distribution $\rho = \rho(\nu, \boldsymbol{X})$ based on this, sample $\{\boldsymbol{Y}_m\}_{m\geqslant 1}^M$ from $\rho$, and use this to discretize the $\arg\max$ procedure in RBM.

Let $u_N(\mu)$ denote the resulting $N$-degree of freedom RBM surrogate.

# Polynomial meshes

I.e., if we can certify accuracy on a "*polynomial* grid", we can probably obtain accuracy.

Let $\mu$ be a random variable with distribution $\nu$. Let $\boldsymbol{X} = (X_n(\nu))_{n=1}^N$ denote a $\mathrm{d}\nu$-orthonormal basis for $P_N$. Define the induced distribution $\rho = \rho(\nu, \boldsymbol{X})$ based on this, sample $\{\boldsymbol{Y}_m\}_{m\geqslant 1}^M$ from $\rho$, and use this to discretize the $\arg\max$ procedure in RBM.

Let $u_N(\mu)$ denote the resulting $N$-degree of freedom RBM surrogate.

If

$$\frac{M}{\log M} \gtrsim \frac{N}{\delta^2} \log\left(\frac{1}{\epsilon}\right),$$

then the least-squares $P_N$-polynomial approximation $v_N(\mu) \in P_N$ to $u_N$ satisfies

$$\mathbb{E}\left[v_N(\mu) - u(\mu)\right]^2 \lesssim N^{-2s} + U^2 \epsilon \frac{1+\delta}{1-\delta},$$

where $U$ is the uniform bound $U = \sup_\mu \|u(\mu)\|$.

# Polynomial meshes

I.e., if we can certify accuracy on a "*polynomial* grid", we can probably obtain accuracy.

Let $\mu$ be a random variable with distribution $\nu$. Let $\boldsymbol{X} = (X_n(\nu))_{n=1}^N$ denote a $\mathrm{d}\nu$-orthonormal basis for $P_N$. Define the induced distribution $\rho = \rho(\nu, \boldsymbol{X})$ based on this, sample $\{\boldsymbol{Y}_m\}_{m\geqslant 1}^M$ from $\rho$, and use this to discretize the $\arg\max$ procedure in RBM.

Let $u_N(\mu)$ denote the resulting $N$-degree of freedom RBM surrogate.

If

$$\frac{M}{\log M} \gtrsim \frac{N}{\delta^2} \log\left(\frac{1}{\epsilon}\right),$$

then the least-squares $P_N$-polynomial approximation $v_N(\mu) \in P_N$ to $u_N$ satisfies

$$\mathbb{E}\left[v_N(\mu) - u(\mu)\right]^2 \lesssim N^{-2s} + U^2\epsilon\frac{1+\delta}{1-\delta},$$

where $U$ is the uniform bound $U = \sup_\mu \|u(\mu)\|$. Without randomization, such a rigorous bound is practically infeasible.

# Example 2: Sparse (polynomial) approximation

# Underdetermined systems

Let $\boldsymbol{x}_0$ be a signal (vector) in $\mathbb{R}^N$. If we have $M \geqslant N$ linear measurements of $\boldsymbol{x}_0$:

$$\boldsymbol{b} := \boldsymbol{A}\boldsymbol{x}_0,$$

then there is (usually) a unique solution $\boldsymbol{x}^*$ that minimizes the $\ell^2$ discrepancy:

$$\boldsymbol{x}^* := \underset{\boldsymbol{z}\in\mathbb{R}^N}{\arg\min} \left\|\boldsymbol{A}\boldsymbol{z} - \boldsymbol{b}\right\|_2 .$$

And (usually), $\boldsymbol{x}^* = \boldsymbol{x}_0$.

# Underdetermined systems

Let $\boldsymbol{x}_0$ be a signal (vector) in $\mathbb{R}^N$. If we have $M \geqslant N$ linear measurements of $\boldsymbol{x}_0$:

$$\boldsymbol{b} := \boldsymbol{A}\boldsymbol{x}_0,$$

then there is (usually) a unique solution $\boldsymbol{x}^*$ that minimizes the $\ell^2$ discrepancy:

$$\boldsymbol{x}^* := \operatorname*{arg\,min}_{\boldsymbol{z}\in\mathbb{R}^N} \|\boldsymbol{A}\boldsymbol{z} - \boldsymbol{b}\|_2 .$$

And (usually), $\boldsymbol{x}^* = \boldsymbol{x}_0$. The situation is (far) more complicated if $M < N$.

This is a particularly salient concern for MOR: $\boldsymbol{x}$ may be a high-dimensional model, but we may only have a small number of measurements.

# Compressive sampling

How can we make this problem well-posed?

Suppose that $\boldsymbol{x}_0$ is $s$-*sparse*, i.e., the number of non-zero terms is at most $s \ll N$.

We can consider the optimization problem,

$$\min \|\boldsymbol{z}\|_0 \quad \text{such that} \quad \boldsymbol{A}\boldsymbol{z} = \boldsymbol{b}.$$

This problem is well-posed under mild conditions.

# Compressive sampling

How can we make this problem well-posed?

Suppose that $\boldsymbol{x}_0$ is $s$-*sparse*, i.e., the number of non-zero terms is at most $s \ll N$.

We can consider the optimization problem,

$$\min \|\boldsymbol{z}\|_0 \quad \text{such that} \quad \boldsymbol{A}\boldsymbol{z} = \boldsymbol{b}.$$

This problem is well-posed under mild conditions.

Unfortunately, it's also NP-hard. A (fairly naive) relaxation of this problem is

$$\min \|\boldsymbol{z}\|_1 \quad \text{such that} \quad \boldsymbol{A}\boldsymbol{z} = \boldsymbol{b}.$$

This is a *convex* problem, and hence it is computationally practical to solve.

# Compressive sampling

How can we make this problem well-posed?

Suppose that $\boldsymbol{x}_0$ is $s$-*sparse*, i.e., the number of non-zero terms is at most $s \ll N$.

We can consider the optimization problem,

$$\min \|\boldsymbol{z}\|_0 \quad \text{such that} \quad \boldsymbol{A}\boldsymbol{z} = \boldsymbol{b}.$$

This problem is well-posed under mild conditions.

Unfortunately, it's also NP-hard. A (fairly naive) relaxation of this problem is

$$\min \|\boldsymbol{z}\|_1 \quad \text{such that} \quad \boldsymbol{A}\boldsymbol{z} = \boldsymbol{b}.$$

This is a *convex* problem, and hence it is computationally practical to solve.

If $\boldsymbol{x}_0$ is sparse, does the $\ell^1$ minimization problem recover the sparse solution?

# Null space and restricted isometry properties

The matrix $\boldsymbol{A}$ satisfies the (robust) *null space property* (NSP) with constant $c$ and sparsity $s$ if

$$\|\boldsymbol{k}_S\|_1 \leqslant c\,\|\boldsymbol{k}_{S^c}\|_1\,, \tag{1}$$

holds for every $\boldsymbol{k} \in \ker(\boldsymbol{A})$, and every subset $S \subset [N]$ with cardinality at most $s$.

Needless to say this is a rather difficult condition to verify directly.

But: (1) is a necessary and sufficient condition so that $\ell^1$ minimization and $\ell^0$ minimization are equivalent. [Cohen, Devore 2009]

# Null space and restricted isometry properties

The matrix $\boldsymbol{A}$ satisfies the (robust) *null space property* (NSP) with constant $c$ and sparsity $s$ if

$$\|\boldsymbol{k}_S\|_1 \leqslant c\,\|\boldsymbol{k}_{S^c}\|_1\,, \tag{1}$$

holds for every $\boldsymbol{k} \in \ker(\boldsymbol{A})$, and every subset $S \subset [N]$ with cardinality at most $s$.

Needless to say this is a rather difficult condition to verify directly.

But: (1) is a necessary and sufficient condition so that $\ell^1$ minimization and $\ell^0$ minimization are equivalent. [Cohen, Devore 2009]

There is a stronger condition to ensure that $\ell^1$ minimization can compute sparse solutions, the *restricted isometry property* (RIP).

$\boldsymbol{A}$ satisfies the RIP with constant $\epsilon$ and sparsity $s$ if

$$(1-\epsilon)\|\boldsymbol{x}\|_2 \leqslant \|\boldsymbol{A}\boldsymbol{x}\|_2 \leqslant (1+\epsilon)\|\boldsymbol{x}\|_2,$$

for every $s$-sparse vector $\boldsymbol{x}$.

This condition may also seem difficult to verify, but it contains $\ell^2$ norms!

# RIP and sparse approximation

The virtue of the RIP is that:

$$\text{RIP} \implies \text{NSP} (\iff \ell^1 \equiv \ell^0)$$

and the RIP is much easier to verify. [Candes, Tao 2005]

# RIP and sparse approximation

The virtue of the RIP is that:

$$\text{RIP} \implies \text{NSP} (\iff \ell^1 \equiv \ell^0)$$

and the RIP is much easier to verify. [Candes, Tao 2005]

In particular, suppose that $\boldsymbol{B} \in \mathbb{R}^{P \times N}$ with $P \geqslant N$ satisfies

$$1 - \delta \leqslant \sigma_{\min}(\boldsymbol{B}), \qquad \sigma_{\max}(\boldsymbol{B}) \leqslant 1 + \delta.$$

Now, form $\boldsymbol{A}$ from $\boldsymbol{B}$ by uniformly at random subsampling $M$ rows from $\boldsymbol{B}$.

# RIP and sparse approximation

The virtue of the RIP is that:

$$\text{RIP} \implies \text{NSP} (\iff \ell^1 \equiv \ell^0)$$

and the RIP is much easier to verify. [Candes, Tao 2005]

In particular, suppose that $\boldsymbol{B} \in \mathbb{R}^{P \times N}$ with $P \geqslant N$ satisfies

$$1 - \delta \leqslant \sigma_{\min}(\boldsymbol{B}), \qquad \sigma_{\max}(\boldsymbol{B}) \leqslant 1 + \delta.$$

Now, form $\boldsymbol{A}$ from $\boldsymbol{B}$ by uniformly at random subsampling $M$ rows from $\boldsymbol{B}$.

Then $\boldsymbol{A}$ satisfies the $(s, \epsilon)$ RIP "with high probability" if

$$M \gtrsim K \log\left(\frac{1}{\epsilon}\right) \frac{1}{1-\delta^2} s \log^3(s) \log N,$$

where $K$ is the maximum row norm of $\boldsymbol{B}$.[Rauhut 2010]

# RIP and sparse approximation

The virtue of the RIP is that:

$$\text{RIP} \implies \text{NSP} (\iff \ell^1 \equiv \ell^0)$$

and the RIP is much easier to verify. [Candes, Tao 2005]

In particular, suppose that $\boldsymbol{B} \in \mathbb{R}^{P \times N}$ with $P \geqslant N$ satisfies

$$1 - \delta \leqslant \sigma_{\min}(\boldsymbol{B}), \qquad \sigma_{\max}(\boldsymbol{B}) \leqslant 1 + \delta.$$

Now, form $\boldsymbol{A}$ from $\boldsymbol{B}$ by uniformly at random subsampling $M$ rows from $\boldsymbol{B}$.

Then $\boldsymbol{A}$ satisfies the $(s, \epsilon)$ RIP "with high probability" if

$$M \gtrsim K \log\left(\frac{1}{\epsilon}\right) \frac{1}{1-\delta^2} s \log^3(s) \log N,$$

where $K$ is the maximum row norm of $\boldsymbol{B}$.[Rauhut 2010]

The problems: (i) $K$ can be very large, and (ii) sometimes $P$ must be (extremely) large before $\delta$ is small.

# The major point

If $\boldsymbol{B}$ is a matrix with "nearly" orthonormal columns, and maximum row norm $K$, then forming $\boldsymbol{A}$ with

$$M \sim Ks$$

subsampled rows yields an RIP matrix.

Hence, if $\boldsymbol{b}$ contains measurements from a sparse vector $\boldsymbol{x}_0$, then (with high probability) the solution to

$$\min \|\boldsymbol{z}\|_1 \quad \text{such that} \quad \boldsymbol{A}\boldsymbol{z} = \boldsymbol{b},$$

is the sparse vector $\boldsymbol{x}_0$.

# The major point (optimized)

From the Chernoff bound: Forming $\boldsymbol{A}$ with

$$M \sim 1s$$

subsampled rows yields an RIP matrix, if:

- we form $\boldsymbol{B}$ by taking $P \sim N \log N$ samples from the induced distribution
- we use the appropriate biasing weights to rescale $\boldsymbol{A}$.

Hence with $M \sim s$ samples, we can guarantee recovery of sparse vectors with sparse measurements.[Adcock, Brugiapaglia, Razi, N 2020]

# The major point (optimized)

From the Chernoff bound: Forming $\boldsymbol{A}$ with

$$M \sim 1s$$

subsampled rows yields an RIP matrix, if:

- we form $\boldsymbol{B}$ by taking $P \sim N \log N$ samples from the induced distribution
- we use the appropriate biasing weights to rescale $\boldsymbol{A}$.

Hence with $M \sim s$ samples, we can guarantee recovery of sparse vectors with sparse measurements.[Adcock, Brugiapaglia, Razi, N 2020]

This type of guarantee is extremely difficult to achieve in general without randomization.

# Randomness is your friend

Many things that cannot be accomplished with deterministic methods can be accomplished* with randomization.

*: with "high" probability

We looked at two examples of this in ROM:

- RBM for elliptic PDE's
- Sparse approximation

There are *many* more examples.