# Randomized Sketches of Convex Programs with Sharp Guarantees

With the help of tikzducks and others

August 29, 2018

# Overview: a blessing...

- Convex Optimization is fundamental in engineering, statistics, and other disciplines.

# Overview: a blessing...

- Convex Optimization is fundamental in engineering, statistics, and other disciplines.
- Attractive property: convex programs can be solved to global optimality (in practice, $\epsilon$-close to the global optimum)

# Overview: a blessing...

- Convex Optimization is fundamental in engineering, statistics, and other disciplines.
- Attractive property: convex programs can be solved to global optimality (in practice, $\epsilon$-close to the global optimum)
- Plethora of theoretical (e.g. convergence, optimality conditions) and practical results (e.g. algorithms, accelerated methods).

# ...and a Curse (of Dimensionality)

While many convex problems can be solved in polynomial time, *not all polynomials are practical*.

# ...and a Curse (of Dimensionality)

While many convex problems can be solved in polynomial time, *not all polynomials are practical*.

## An example: linear regression

Noisy linear measurements:

$$y_i = \langle a_i, x^* \rangle + \eta_i, \ i = 1, \ldots, n.$$

Convex program to find least squares estimate:

$$\text{Minimize } \|Ax - y\|_2^2$$

Solving via $QR$ factorization requires ($\sim \mathcal{O}(nd^2)$)

# More generally: statistical estimation

- Parameter estimation + prior information about parameter
- Low dimensional spaces: sparse vectors, low rank matrices, etc.
- The convex optimization way: relax constraints.

## Examples

1. $s$-sparse vectors $\Rightarrow \|x\|_1 \leq s$ (also known as *basis pursuit*)
2. $\text{rank}(A) \leq r \Rightarrow \|A\|_* \leq r$ (nuclear norm regularization)

Problem: ambient space of relaxation can be too large

# Random projections

Random projections go back at least as far as 1984: Johnson and Lindenstrauss showed that we can project a set of $m$ points from $\mathbb{R}^d$ in a subspace of dimension $\Theta\left(\frac{\log m}{\epsilon^2}\right)$ without distorting the distances between them more than $\epsilon$.

**General idea**: project to low dimensional subspace and solve the problem efficiently there to obtain $\widehat{x}$.

# Sketching for quadratic programs

Convex program:

$$x^\star \in \arg\min_{x \in \mathcal{C}} \overbrace{\|Ax - y\|_2^2}^{f(x)}, \quad A \in \mathbb{R}^{n \times d}, y \in \mathbb{R}^n.$$

**Sketched** program:

$$\widehat{x} \in \arg\min_{x \in \mathcal{C}} \underbrace{\|S(Ax - y)\|_2^2}_{g(x)}, \quad S \in \mathbb{R}^{m \times n}, \; \underbrace{m < \min\{n, d\}}_{\text{ideally}}$$

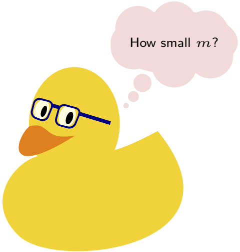
# Sketching for quadratic programs

Convex program:

$$x^\star \in \arg\min_{x \in \mathcal{C}} \overbrace{\|Ax - y\|_2^2}^{f(x)}, \quad A \in \mathbb{R}^{n \times d}, y \in \mathbb{R}^n.$$

**Sketched** program:

$$\widehat{x} \in \arg\min_{x \in \mathcal{C}} \underbrace{\|S(Ax - y)\|_2^2}_{g(x)}, \quad S \in \mathbb{R}^{m \times n}, \ \underbrace{m < \min\{n, d\}}_{\text{ideally}}$$



# Sketching matrices

Different choices of sketching matrix $S$:

- **subgaussian** sketch: rows $s_i$ are i.i.d. subgaussian, i.e.

$$\mathbb{P}\left(|\langle s_i, u\rangle| \geq t\right) \leq 2e^{-\frac{t^2}{2\sigma^2}}, \; \forall t > 0.$$

  Satisfied by Gaussian or Rademacher vectors

- **randomized orthogonal** sketch: $s_i$ are i.i.d. orthonormal rows satisfying $s_i = \sqrt{n} D H^\top p_i$, where $D$ a random diagonal matrix, $H$ the Hadamard matrix, and $p_i$ chosen uniformly from the canonical basis of $\mathbb{R}^n$.

Will examine subgaussian case in detail.

# Key quantities

## Tangent cone $\mathcal{K}$

Given a constraint set $\mathcal{C} \subseteq \mathbb{R}^d$, the cone of all feasible directions from the optimum $x^\star \in \mathcal{C}$ is defined as

$$\mathcal{K} := \operatorname{clconv}\left\{z \in \mathbb{R}^d \;\middle|\; z = t(x - x^\star),\ t \geq 0,\ x \in \mathcal{C}\right\}$$

Since the objective function is $\|Ax - y\|_2^2$, we need to examine the **transformed** cone:

$$A\mathcal{K} := \{Az \in \mathbb{R}^n \mid z \in \mathcal{K}\}$$

# Key quantities

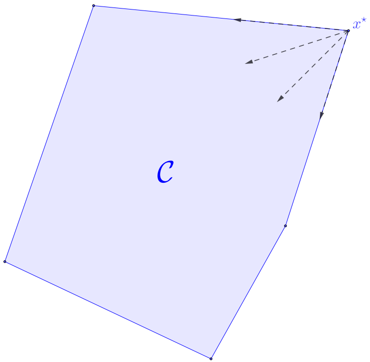



Figure: rays of $\mathcal{K}$ shown in black

# Key quantities

## Gaussian width

Given a set $S \subseteq \mathbb{R}^n$, we define its **Gaussian width** as

$$\mathbb{W}(S) := \mathbb{E}_g \left[ \sup_{z \in S} |\langle g, z \rangle| \right], \quad g \sim \mathcal{N}(0, I_n)$$

**Interpretation**: in a probabilistic scenario, sets with large gaussian width will exhibit more degrees of freedom.

Quantity of interest: $\boxed{\mathbb{W}(A\mathcal{K}) := \mathbb{E}_g \left[ \sup_{z \in A\mathcal{K} \cap \mathcal{S}^{n-1}} |\langle g, z \rangle| \right]}$

# Main results - in a nutshell

- $m$ should not depend on the ambient dimension, but rather the **statistical** dimension!

# Main results - in a nutshell

- $m$ should not depend on the ambient dimension, but rather the **statistical** dimension!
- if the sketching matrix is structured, expect to pay a small price for the randomness lost

# Main results - in a nutshell

- $m$ should not depend on the ambient dimension, but rather the **statistical** dimension!
- if the sketching matrix is structured, expect to pay a small price for the randomness lost

$\delta$-optimality: $\boxed{f(\hat{x}) \leq (1+\delta)^2 f(x^\star)}$

# Main result - subgaussian sketch

## Theorem 1

Pick $S \in \mathbb{R}^{m \times n}$ according to the subgaussian model. There exist universal constants $c_0, c_1, c_2 > 0$ such that, $\forall \delta \in (0, 1)$, when

$$m \geq \frac{c_0}{\delta^2} \mathbb{W}^2(A\mathcal{K}),$$

the sketched solution is $\delta$-optimal with probability at least $1 - c_1 e^{-c_2 m \delta^2}$.

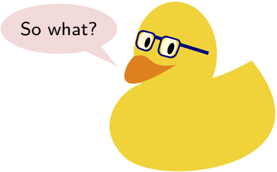
# Main result - subgaussian sketch

**Theorem 1**

Pick $S \in \mathbb{R}^{m \times n}$ according to the subgaussian model. There exist universal constants $c_0, c_1, c_2 > 0$ such that, $\forall \delta \in (0,1)$, when

$$m \geq \frac{c_0}{\delta^2} \mathbb{W}^2(A\mathcal{K}),$$

the sketched solution is $\delta$-optimal with probability at least $1 - c_1 e^{-c_2 m \delta^2}$.



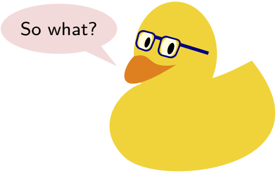
# Main result - subgaussian sketch

**Theorem 1**

Pick $S \in \mathbb{R}^{m \times n}$ according to the subgaussian model. There exist universal constants $c_0, c_1, c_2 > 0$ such that, $\forall \delta \in (0,1)$, when

$$m \geq \frac{c_0}{\delta^2} \mathbb{W}^2(A\mathcal{K}),$$

the sketched solution is $\delta$-optimal with probability at least $1 - c_1 e^{-c_2 m \delta^2}$.




- always true: $\mathbb{W}^2(A\mathcal{K}) \leq n$
- for many sets $A\mathcal{K}$, **much** smaller! (e.g. low rank matrices)

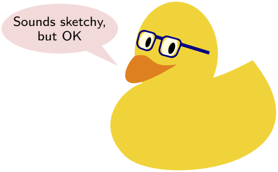
# Main result - subgaussian sketch

**Theorem 1**

Pick $S \in \mathbb{R}^{m \times n}$ according to the subgaussian model. There exist universal constants $c_0, c_1, c_2 > 0$ such that, $\forall \delta \in (0,1)$, when

$$m \geq \frac{c_0}{\delta^2} \mathbb{W}^2(A\mathcal{K}),$$

the sketched solution is $\delta$-optimal with probability at least $1 - c_1 e^{-c_2 m \delta^2}$.




- always true: $\mathbb{W}^2(A\mathcal{K}) \leq n$
- for many sets $A\mathcal{K}$, **much** smaller! (e.g. low rank matrices)

# Unconstrained least squares

Consider the sketched problem:

$$\inf_{x \in \mathbb{R}^n} \|S(Ax - y)\|_2^2$$

Intuition: complexity of the problem depends on $\mathrm{rank}(A)$

**How big is** $\mathbb{W}(A\mathcal{K})$?

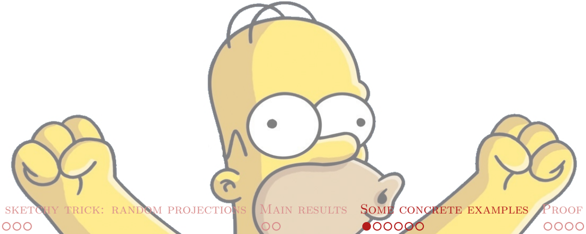
# Unconstrained least squares

Consider the sketched problem:

$$\inf_{x \in \mathbb{R}^n} \|S(Ax - y)\|_2^2$$

Intuition: complexity of the problem depends on $\mathrm{rank}(A)$

**How big is $\mathbb{W}(A\mathcal{K})$?** Less than $\sqrt{\mathrm{rank}(A)}$!



# Unconstrained least squares

## Corollary 2(a) in [4]

Consider the problem $\inf_{x \in \mathbb{R}^n} \|Ax - y\|_2^2$ and its sketched version $\inf_{x \in \mathbb{R}^n} \|S(Ax - y)\|_2^2$, $S \in \mathbb{R}^{m \times n}$. If $m \geq c_0 \frac{\mathrm{rank}(A)}{\delta^2}$, the sketched solution satisfies

$$\mathbb{P}\left(f(\widehat{x}) \leq (1+\delta)^2 f(x^\star)\right) \geq 1 - c_1 \exp\left(-c_2 m \delta^2\right).$$

**Proof sketch**: Write $Au = \sum_{i=1}^{\mathrm{rank}(A)} \lambda_i a_i$, where $\{a_i\}_{i=1}^{\mathrm{rank}(A)}$ is an orthonormal basis of $\mathrm{im}(A)$. Plug into the definition of $\mathbb{W}(A\mathcal{K})$ and apply Theorem 1.

# Unconstrained least squares

## Corollary 2(a) in [4]

Consider the problem $\inf_{x\in\mathbb{R}^n} \|Ax-y\|_2^2$ and its sketched version $\inf_{x\in\mathbb{R}^n} \|S(Ax-y)\|_2^2$, $S\in\mathbb{R}^{m\times n}$. If $m \geq c_0 \frac{\mathrm{rank}(A)}{\delta^2}$, the sketched solution satisfies

$$\mathbb{P}\left(f(\widehat{x}) \leq (1+\delta)^2 f(x^\star)\right) \geq 1 - c_1 \exp\left(-c_2 m\delta^2\right).$$

**Proof sketch**: Write $Au = \sum_{i=1}^{\mathrm{rank}(A)} \lambda_i a_i$, where $\{a_i\}_{i=1}^{\mathrm{rank}(A)}$ is an orthonormal basis of $\mathrm{im}(A)$. Plug into the definition of $\mathbb{W}(A\mathcal{K})$ and apply Theorem 1.

Solving sketched problem via QR: $\mathcal{O}(ndm + md^2)$

# $\ell_1$-constrained least squares

Consider one of the equivalent formulations of the LASSO, given by

$$x^\star \in \arg\min_{\|x\|_1 \leq R} \|Ax - y\|_2^2$$

Case of interest: $x^\star$ sparse ($|\{i \in [d] : x_i \neq 0\}| = k \ll d$), unique.

**Key quantity**: $\ell_1$-restricted eigenvalue

$$\gamma_k^-(A) \triangleq \inf_{\substack{\|z\|_2 = 1 \\ \|z\|_1 \leq 2\sqrt{k}}} \|Az\|_2^2 \tag{1}$$

# Calculating $\mathbb{W}(A\mathcal{K})$ - 1

- Tangent cone: $\mathcal{K} = \{\Delta : \langle \Delta_S, \mathrm{sign}(x_S^*) \rangle + \|\Delta_{S^c}\|_1 \leq 0\}$, where $S$ support of $x^\star$. By C-S, gives

$$\|\Delta_{S^c}\|_1 \leq \|\Delta_S\|_2 \underbrace{\|\mathrm{sign}(x_S^\star)\|_2}_{=\sqrt{k}} \tag{2}$$

- Chain of (in)equalities:

$$\|\Delta\|_1 = \|\Delta_S\|_1 + \|\Delta_{S^c}\|_1 \leq \sqrt{k}\,\|\Delta_S\|_2 + \underbrace{\sqrt{k}\,\|\Delta_S\|_2}_{\text{from (2)}} \leq 2\sqrt{k}\,\|\Delta\|_2$$

- $\underbrace{\|A\Delta\|_2 = 1}_{A\Delta \in \mathbb{S}^{n-1}} \Rightarrow \|\Delta\|_2 \leq \frac{1}{\sqrt{\gamma_k^-(A)}}$, from definition of $\gamma_k^-(A)$.

# Calculating $\mathbb{W}(A\mathcal{K})$ - 2

Combine all of the above to obtain

$$\begin{aligned}\mathbb{E}\left[|\langle A\Delta, g\rangle|\right] = \mathbb{E}\left[\left|\langle \Delta, A^\top g\rangle\right|\right] &\overset{\text{(Hölder)}}{\leq} \mathbb{E}\left[\|\Delta\|_1 \left\|A^\top g\right\|_\infty\right] \\ &\leq \frac{2\sqrt{k}\mathbb{E}\left[\left\|A^\top g\right\|_\infty\right]}{\sqrt{\gamma_k^-(A)}} = \frac{2\sqrt{k}}{\sqrt{\gamma_k^-(A)}}\mathbb{E}\left[\max_{i\in[d]} \left|a_i^\top g\right|\right]\end{aligned}$$

Observe: $a_i^\top g \sim \mathcal{N}(0, \|a_i\|_2^2)$ so by a standard argument [2, Eq. (3.13)]

$$\mathbb{W}(A\mathcal{K}) \leq 6\sqrt{\frac{k \log d}{\gamma_k^-(A)}} \max_{i\in[d]} \|a_i\|_2$$

# Calculating $\mathbb{W}(A\mathcal{K})$ - 3

Finally: $A\mathcal{K} \subseteq \mathrm{Im}(A) \Rightarrow$ bound from unconstrained LS still valid!

$$\mathbb{W}(A\mathcal{K}) \leq \min\left\{\mathrm{rank}(A), 2\sqrt{\frac{k\log d}{\gamma_k^-(A)}}\max_{i\in[d]}\|a_i\|_2\right\}$$

## Corollary 3a in [4]

Consider a sub-gaussian sketch applied to the $\ell_1$-constrained least squares. The solution is $\delta$-optimal for

$$m \geq \frac{c_0}{\delta^2}\min\left\{\sqrt{\mathrm{rank}(A)}, 6\sqrt{\frac{k\log d}{\gamma_k^-(A)}}\max_{i\in[d]}\|a_i\|_2\right\}.$$

# Proving Theorem 1 - Overview

Proof of Theorem 1 consists of 2 conceptually simple steps:

1. A deterministic lemma relating $f(\widehat{x})$ to $f(x^\star)$, which depends on two random quantities, using convex optimality conditions

# Proving Theorem 1 - Overview

Proof of Theorem 1 consists of 2 conceptually simple steps:

1. A deterministic lemma relating $f(\widehat{x})$ to $f(x^\star)$, which depends on two random quantities, using convex optimality conditions
2. Bounding said quantities with high probability, using concentration inequalities and empirical process theory

**Key quantities:**

$$
\begin{aligned}
Z_1(A\mathcal{K}) &:= \inf_{v \in A\mathcal{K} \cap \mathbb{S}^{n-1}} \frac{1}{m} \|Sv\|_2^2 \\
Z_2(A\mathcal{K}) &:= \sup_{v \in A\mathcal{K} \cap \mathbb{S}^{n-1}} \left| \left\langle u, \left( \frac{1}{m} S^\top S - I_n \right) v \right\rangle \right|, \; u \in \mathbb{S}^{n-1}.
\end{aligned} \tag{3}
$$

# From $f(\widehat{x})$ to $f(x^\star)$

Step (1a): define $e = \widehat{x} - x^\star$, rewrite $f(\widehat{x})$ + triangle ineq.:

$$f(\widehat{x}) \leq \left(1 + \frac{\|Ae\|_2}{\|Ax^\star - y\|_2}\right)^2 f(x^\star).$$

Step (1b): write optimality conditions for original and sketched problem to obtain

$$\frac{1}{2m}\|SAe\|_2^2 \leq \left| \left\langle Ax^\star - y, \underbrace{\left(\frac{1}{m}S^\top S - I_n\right)}_{=:Q} Ae \right\rangle \right|$$

$e$ belongs to $\mathcal{K}$ since both original and sketched problem have same constraints, leading to

$$\frac{1}{2} Z_1(A\mathcal{K})\, \|Ae\|_2 \leq \|Ax^\star - y\|_2\, Z_2(A\mathcal{K}).$$

# From $f(\widehat{x})$ to $f(x^\star)$

We've proved the following deterministic lemma:

**Lemma 1**

For any choice of sketching matrix $S \in \mathbb{R}^{m \times n}$, we have

$$f(\widehat{x}) \leq \left(1 + 2\frac{Z_2(A\mathcal{K})}{Z_1(A\mathcal{K})}\right)^2 f(x^\star).$$

# Controlling $Z_1, Z_2$

To show that the ratio in Lemma 1 is small w.h.p, sufficient to bound $Z_1$ from below and $Z_2$ from above.
Proof depends on following result:

**Proposition 1 (follows Theorem D from [3])**

Let $\{s_i\}_{i=1}^m$ i.i.d from an isotropic $\sigma$-subgaussian distr. $\exists c_1, c_2 > 0$ such that for any $\mathcal{Y} \subseteq \mathbb{S}^{n-1}$:

$$\mathbb{P}\left(\sup_{y \in \mathcal{Y}} \left| y^\top \left(\frac{1}{m} S^\top S - I_n\right) y \right| \leq c_1 \frac{\mathbb{W}(\mathcal{Y})}{\sqrt{m}} + \delta \right) \geq 1 - e^{-\frac{c_2 m \delta^2}{\sigma^4}}.$$

# Step (2a) - controlling $Z_1$

Controlling $Z_1$ is easy - apply Prop. 1 with $\mathcal{Y} = A\mathcal{K} \cap \mathbb{S}^{n-1}$:

$$\forall v \in A\mathcal{K} \cap \mathbb{S}^{n-1} : \frac{1}{m}\|Sv\|_2^2 - \cancelto{1}{v^\top v} \geq -c_1 \frac{\mathbb{W}(A\mathcal{K})}{\sqrt{m}} - \delta$$

$$\Rightarrow \inf_{v \in A\mathcal{K} \cap \mathbb{S}^{n-1}} \frac{1}{m}\|Sv\|_2^2 \geq 1 - c_1 \frac{\mathbb{W}(A\mathcal{K})}{\sqrt{m}} - \delta$$

Set $m \geq \left(\frac{c_1 \mathbb{W}(A\mathcal{K})}{\delta}\right)^2$, relabel as required. $\square$

# Step (2b) - controlling $Z_2$

Controlling $Z_2(A\mathcal{K})$ is more involved: $Z_2$ involves two vectors $u, v$, which dictate a decomposition of $Z_2$ s.t. Prop. 1 is applicable:

$$Z_2(A\mathcal{K}) \leq \sup_{v \in \mathcal{V}_+} \left|u^\top Q v\right| + \sup_{v \in \mathcal{V}_-} \left|u^\top Q v\right|$$

For $v \in \mathcal{V}_+$, wlog:

$$\left|u^\top Q v\right| \leq \underbrace{\frac{1}{2}\left|(u+v)^\top Q (u+v)\right|}_{\text{only nontrivial term}} + \underbrace{\frac{1}{2}\left|u^\top Q u\right|}_{\mathcal{Y} \leftarrow \{u\}} + \underbrace{\frac{1}{2}\left|v^\top Q v\right|}_{\mathcal{Y} \leftarrow \mathcal{V}_+}$$

$$Q := \frac{S^\top S}{m} - I_n$$

$$V_+ = \{v \in A\mathcal{K} : \langle u, v\rangle \geq 0\}$$

$$V_- = \{v \in A\mathcal{K} : \langle u, v\rangle < 0\}$$

# Step (2b) - controlling the nontrivial term

Define $\mathcal{U}_+ := \left\{ \frac{u+v}{\|u+v\|} \;\middle|\; v \in \mathcal{V}_+ \right\}$ and apply Prop. 1 with $\mathcal{Y} = \mathcal{U}_+$:

$$\sup_{v \in \mathcal{V}_+} \frac{\left|(u+v)^\top Q (u+v)\right|}{\underbrace{\|u+v\|^2}_{\leq 4}} \leq c_1 \frac{\mathbb{W}(\mathcal{U}_+)}{\sqrt{m}} + \delta.$$

Split $v$ into a $u$-parallel and a $u$-orthogonal component, $\Pi(v)$, to obtain

$$\underbrace{\mathbb{E}\left[\sup_{v \in \mathcal{V}_+} \left|\left\langle g, \frac{u+v}{\|u+v\|}\right\rangle\right|\right]}_{=\mathbb{W}(\mathcal{U}_+)} \leq \underbrace{\mathbb{E}\left[|\langle g, u\rangle|\right]}_{=\sqrt{\frac{2}{\pi}}} + \underbrace{\mathbb{E}\left[\sup_{v \in \mathcal{V}_+} |\langle g, \Pi(v)\rangle|\right]}_{\leq ?}$$

# Step (2b) - controlling the nontrivial term

For the last term, note that

$$\mathbb{E}\left[\langle g, \Pi(v) - \Pi(v')\rangle^2\right] = (\Pi(v) - \Pi(v'))^\top \underbrace{\mathbb{E}\left[gg^\top\right]}_{=I}(\Pi(v) - \Pi(v'))$$
$$\leq \|v - v'\|^2 = \mathbb{E}\left[(\langle g, v - v'\rangle)^2\right]$$

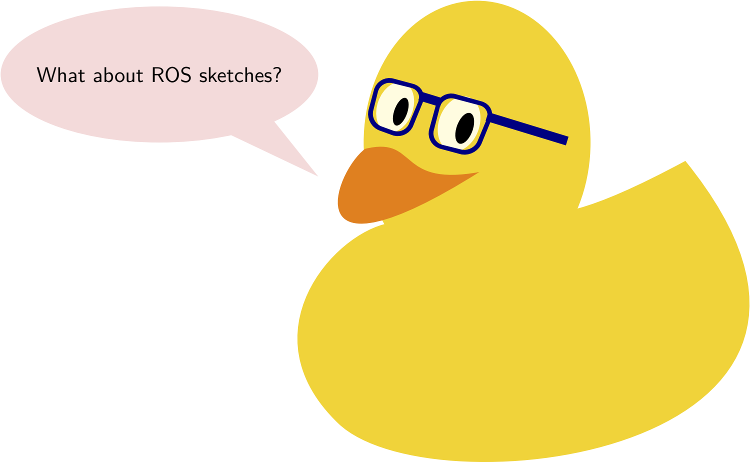
Using the Sudakov-Fenrique inequality[1], we deduce that

$$\mathbb{E}\left[\sup_{v \in \mathcal{V}_+} |\langle g, \Pi(v)\rangle|\right] \leq \mathbb{E}\left[\sup_{v \in \mathcal{V}_+} |\langle g, v\rangle|\right] = \mathbb{W}(\mathcal{V}_+) \leq \mathbb{W}(A\mathcal{K} \cap \mathbb{S}^{n-1})$$

which completes the bounding of $\mathbb{W}(\mathcal{U}_+)$.

[1] see e.g. [7, Theorem 7.2.11]

# The duck says...



# What about ROS sketches?

In a nutshell: faster computations, not so sharp guarantees for $m$.
Different quantities at play:

# What about ROS sketches?

In a nutshell: faster computations, not so sharp guarantees for $m$.
Different quantities at play:

- *S-Gaussian width*: $\boxed{\mathbb{W}_S(A\mathcal{K}) := \mathbb{E}_{g,S}\left[\sup_{z \in A\mathcal{K} \cap \mathcal{S}^{n-1}} \frac{1}{\sqrt{m}} |\langle g, SAz \rangle|\right]}$
  - ▷ can be close to $\mathbb{W}(A\mathcal{K})$.

# What about ROS sketches?

In a nutshell: faster computations, not so sharp guarantees for $m$.
Different quantities at play:

- *S-Gaussian width*: $\boxed{\mathbb{W}_S(A\mathcal{K}) := \mathbb{E}_{g,S}\left[\sup_{z \in A\mathcal{K} \cap \mathcal{S}^{n-1}} \frac{1}{\sqrt{m}} |\langle g, SAz \rangle|\right]}$
  - ▷ can be close to $\mathbb{W}(A\mathcal{K})$.
- *Rademacher complexity*: $\boxed{\mathbb{R}(A\mathcal{K}) := \mathbb{E}_{\varepsilon}\left[\sup_{z \in A\mathcal{K} \cap \mathcal{S}^{n-1}} |\langle \varepsilon, z \rangle|\right]}$
  - ▷ bounded above by a constant times $\mathbb{W}(A\mathcal{K})$.

# What about ROS sketches?

In a nutshell: faster computations, not so sharp guarantees for $m$.
Different quantities at play:

- *S-Gaussian width*: $\boxed{\mathbb{W}_S(A\mathcal{K}) := \mathbb{E}_{g,S}\left[\sup_{z\in A\mathcal{K}\cap\mathcal{S}^{n-1}} \frac{1}{\sqrt{m}}|\langle g, SAz\rangle|\right]}$
    - ▷ can be close to $\mathbb{W}(A\mathcal{K})$.
- *Rademacher complexity*: $\boxed{\mathbb{R}(A\mathcal{K}) := \mathbb{E}_{\varepsilon}\left[\sup_{z\in A\mathcal{K}\cap\mathcal{S}^{n-1}} |\langle \varepsilon, z\rangle|\right]}$
    - ▷ bounded above by a constant times $\mathbb{W}(A\mathcal{K})$.

Caveat: $\frac{m}{\log m} = \Omega\left\{\left(\mathbb{R}^2(A\mathcal{K}) + \log n\right)\mathbb{W}_S^2(A\mathcal{K})\right\}$, best $\mathcal{K}$-agnostic bound obtained in paper

# Sharpening bounds for ROS sketches

Obtaining sharper bounds requires tangent cone $\mathcal{K}$ to be more structured. **General pattern** for sharpening bounds (not covered here): reduction to finite maxima.

Example - $\mathcal{K}$ is a subspace:

- $Z_1(A\mathcal{K}) \geq 1 - \delta, Z_2(A\mathcal{K}) \leq \frac{3}{2}\delta$, as long as $m = \Omega(\log^4 n \cdot \operatorname{rank}(A))$
- **Proof sketch**: relax $A\mathcal{K} \cap \mathbb{S}^{n-1}$ to $A\mathcal{K} \cap \mathbb{B}_2^n$, take $\epsilon$-net, and appeal to a JL property for finite sets of points from [1]

# Disclaimer

- We described a randomized method to $(1+\delta)^2$ approximate the optimal *cost* with high probability
- We did **not** give any guarantees about approximating the optimal *solution*, $x^\star$

Addressed in other works, e.g. Iterative Hessian sketch [5], Newton sketch [6].

# Disclaimer

- We described a randomized method to $(1+\delta)^2$ approximate the optimal *cost* with high probability
- We did **not** give any guarantees about approximating the optimal *solution*, $x^\star$

Addressed in other works, e.g. Iterative Hessian sketch [5], Newton sketch [6].

Takeaway: *randomized methods work!* (pro{v, b}ably)

Thank you!

# References I

[1] Felix Krahmer and Rachel Ward. New and improved johnson–lindenstrauss embeddings via the restricted isometry property. *SIAM Journal on Mathematical Analysis*, 43(3):1269–1281, 2011.

[2] Michel Ledoux and Michel Talagrand. *Probability in Banach Spaces: isoperimetry and processes*. Springer Science & Business Media, 2013.

[3] Shahar Mendelson, Alain Pajor, and Nicole Tomczak-Jaegermann. Reconstruction and subgaussian operators in asymptotic geometric analysis. *Geometric and Functional Analysis*, 17(4):1248–1282, 2007.

[4] Mert Pilanci and Martin J Wainwright. Randomized sketches of convex programs with sharp guarantees. *IEEE Transactions on Information Theory*, 61(9):5096–5115, 2015.

[5] Mert Pilanci and Martin J Wainwright. Iterative hessian sketch: Fast and accurate solution approximation for constrained least-squares. *The Journal of Machine Learning Research*, 17(1):1842–1879, 2016.

# References II

[6] Mert Pilanci and Martin J Wainwright. Newton sketch: A near linear-time optimization algorithm with linear-quadratic convergence. *SIAM Journal on Optimization*, 27(1):205–245, 2017.

[7] Roman Vershynin. *High Dimension Probability: with Applications in Data Science*. Cambridge University Press, 2018.