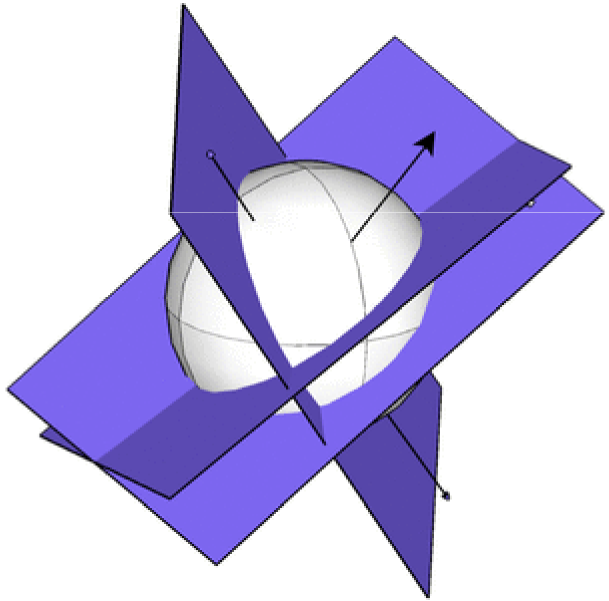
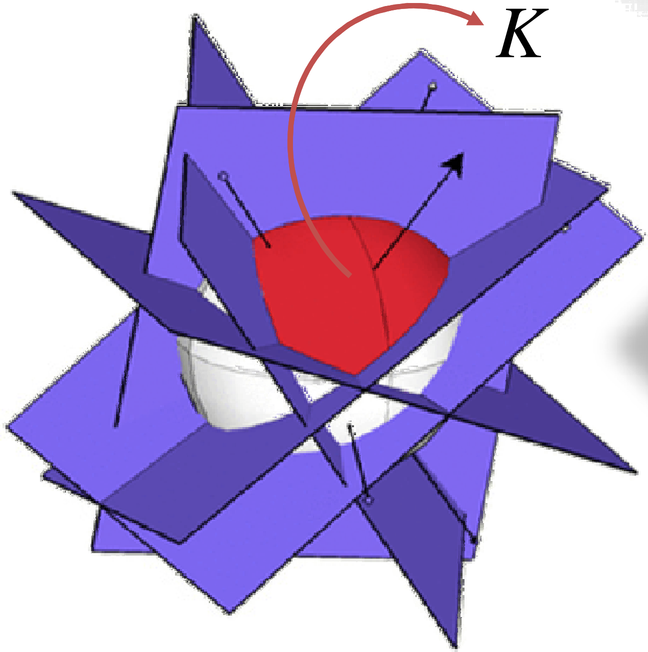
Robust 1-bit Compressed Sensing and Sparse Logistic Regression: A Convex Programming Approach

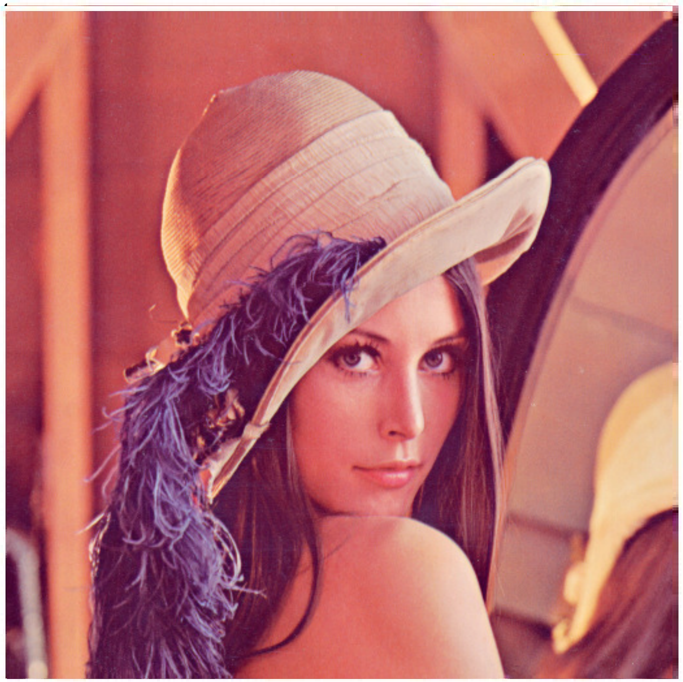
Yaniv Plan and Roman Vershynin



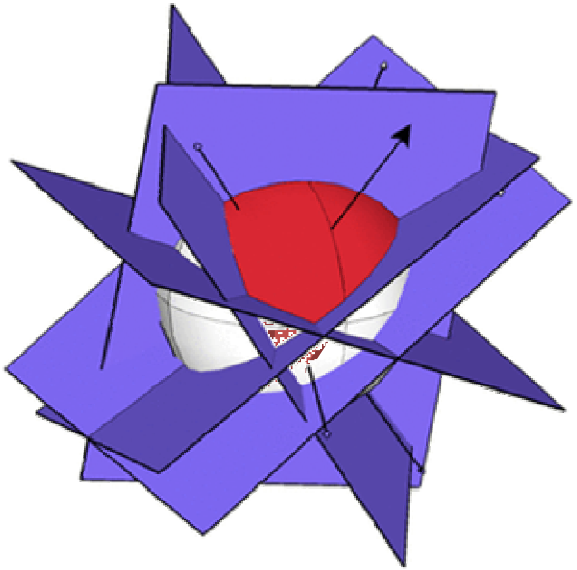
2014.03.05

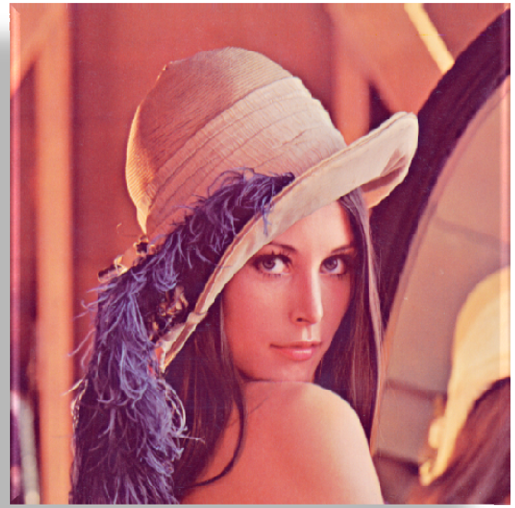
Presenter: Kuo-Shih Tseng





At least ***m*** cuts s.t. the image is **clear**?





m cuts ⟷ mean width ⟷ signal

$$m \geq C\delta^{-2} w(K)^2 \qquad w(K)^2 \leq Cs\log(2n/s)$$

$$m = O(s\log(n/s))$$

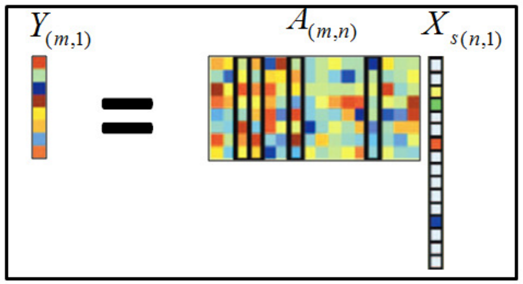
$$\left.\begin{array}{l} n = 512*512 \\ s/n = 2\% \end{array}\right\} m \geq 150$$

Outline

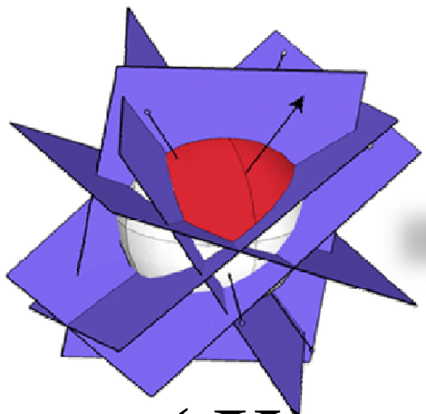
- What's compressed sensing
- What's 1-bit compressed sensing

- Mean width [w ↔ (n, s)] [m ↔ w]
- 1-bit compressed sensing

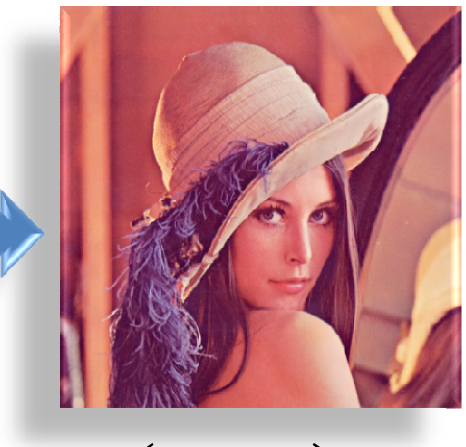
- Theorem 1.1



m

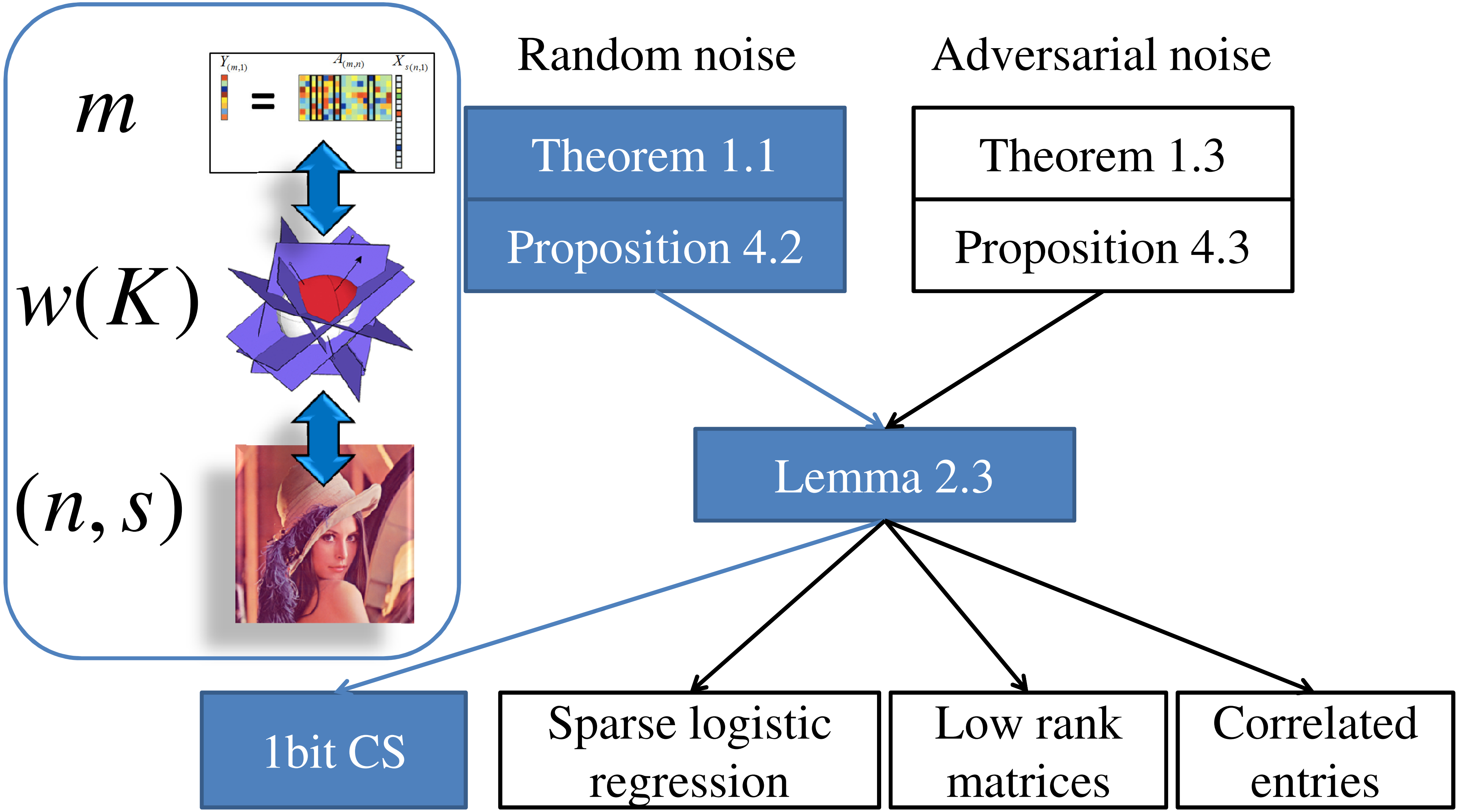


$w(K)$

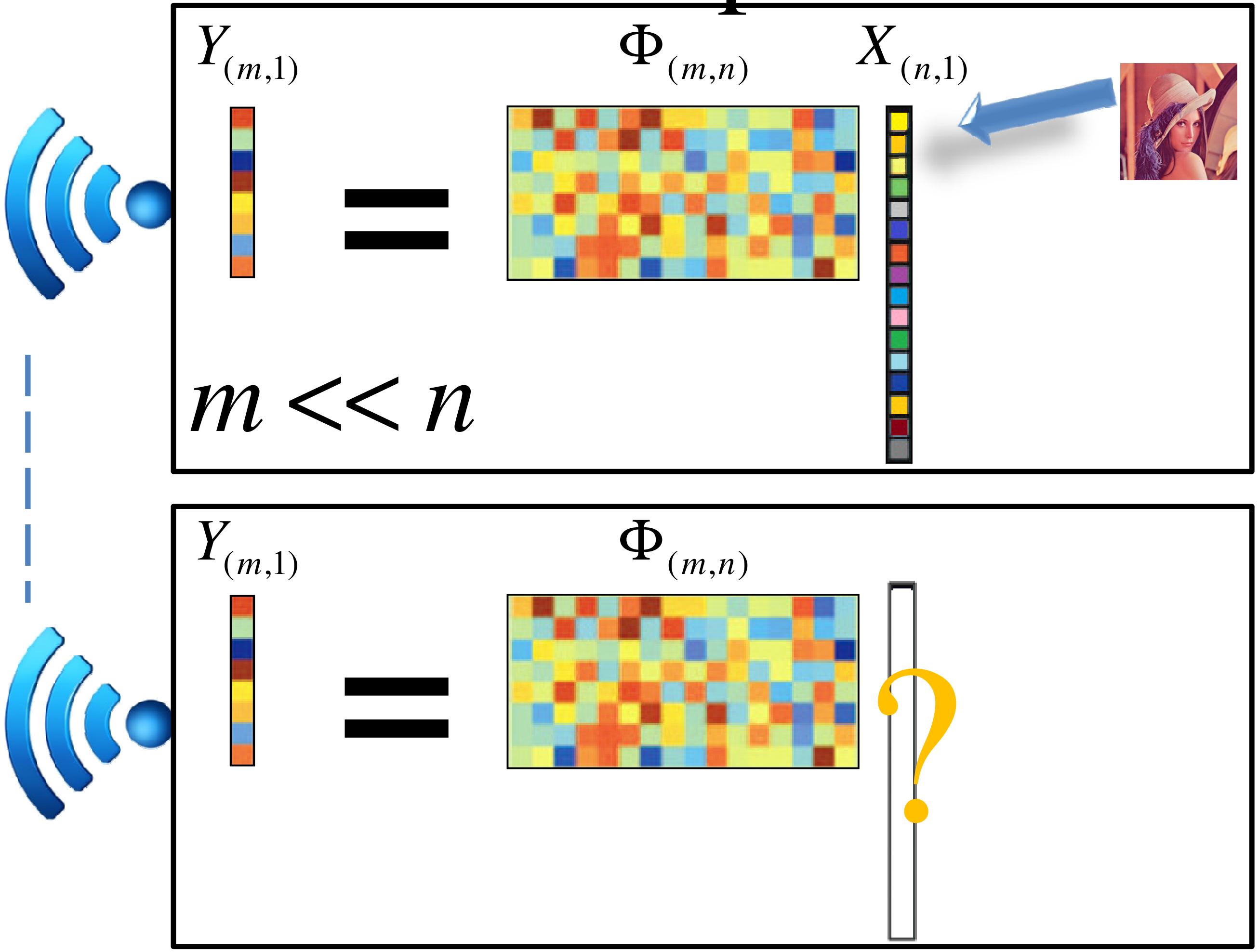


(n, s)

Outline


m
w(K)
(n, s)
Random noise
Adversarial noise
Theorem 1.1
Proposition 4.2
Theorem 1.3
Proposition 4.3
Lemma 2.3
1bit CS
Sparse logistic regression
Low rank matrices
Correlated entries

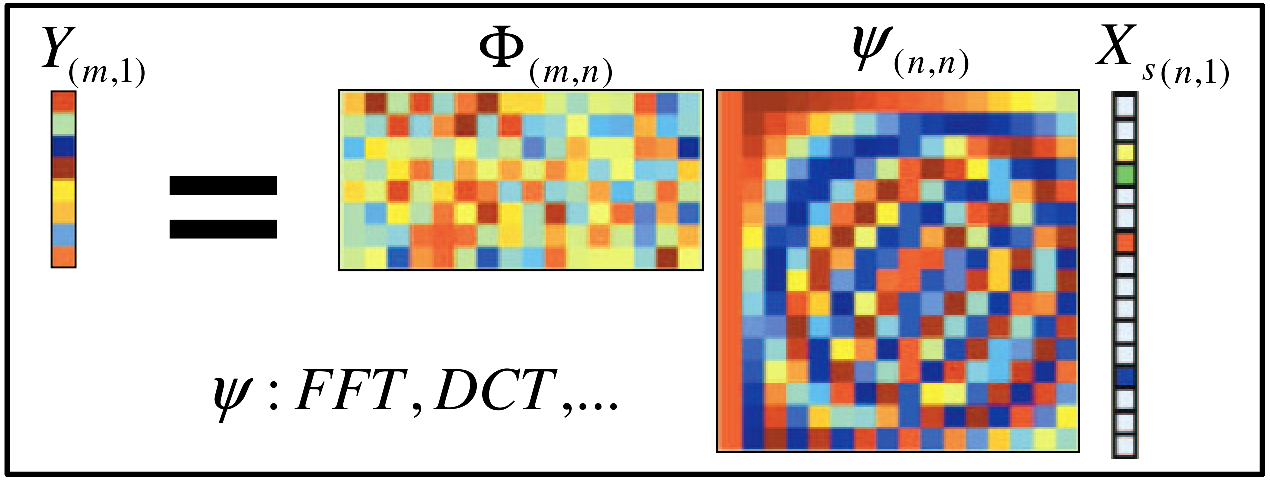
What's compressed sensing

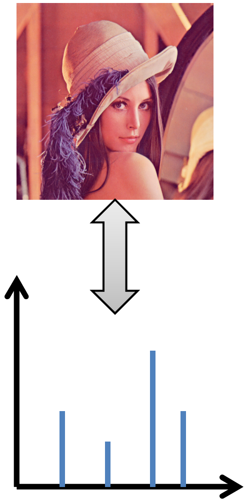


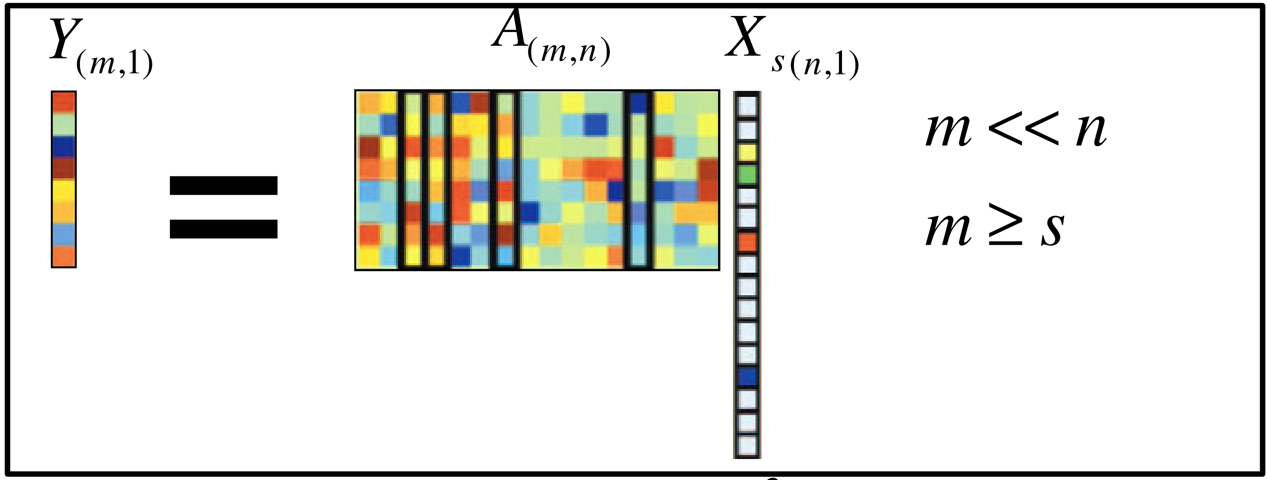
$$\text{Known}: Y_{(m,1)}, \Phi_{(m,n)}, \text{Find}: X_{(n,1)}$$

$$X = (\Phi^T \Phi)^{-1} \Phi^T Y \,?$$

What's compressed sensing

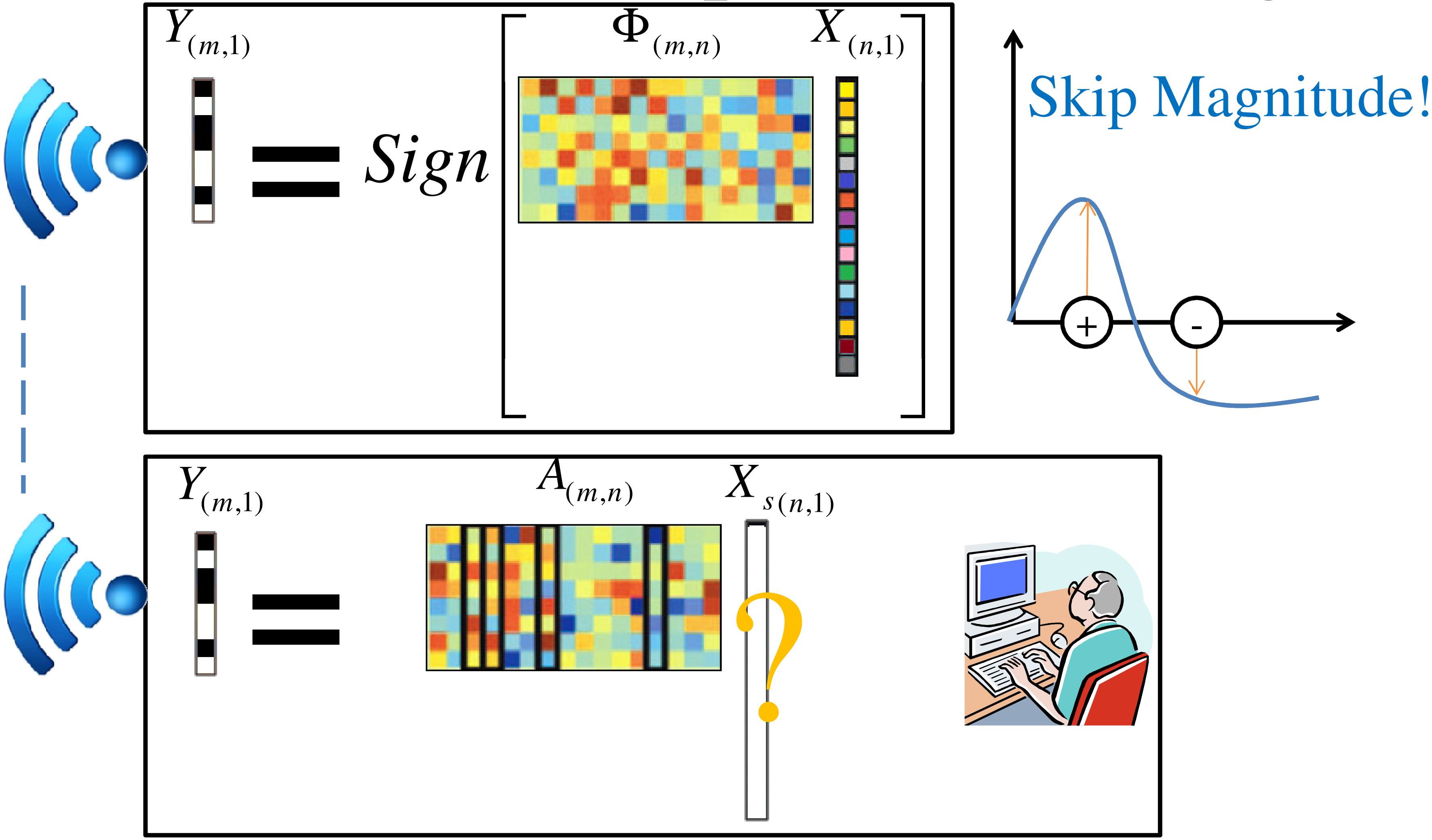






$$\hat{X} = \arg\min_{X} \|X\|_1 \text{ s.t. } \frac{\lambda}{2} \|Y - AX\|_2^2$$

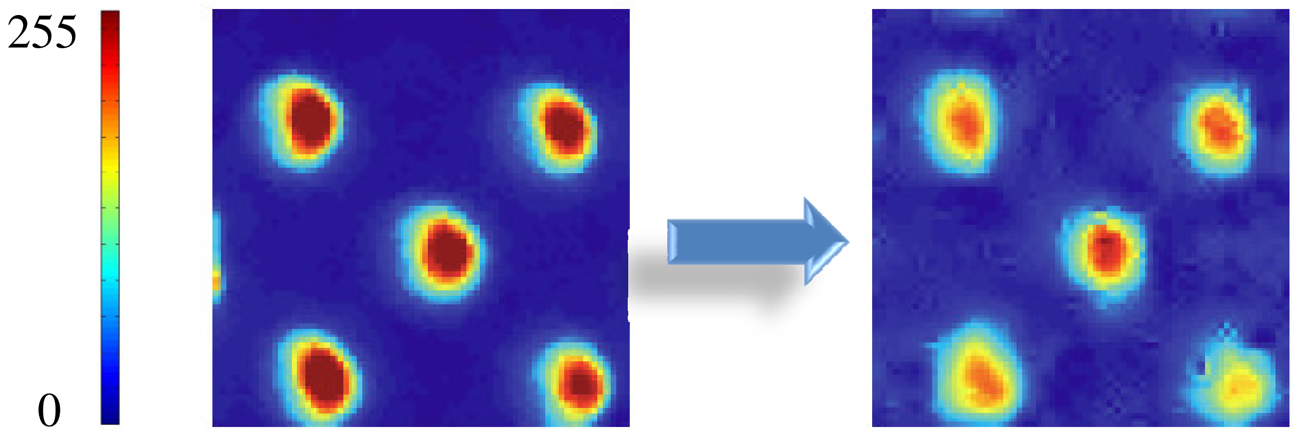
What’s 1-bit compressed sensing



At least ***m*** bits s.t. the image is **clear**?

What's 1-bit compressed sensing

- Dramatic compression
  - n→m (compressed sensing)
  - m bits (1-bit)
- Could we recover signal accurately?



4096 pixels, 8bits = 32 kb 4096 bits= 4 kb

P. T. Boufounos and R. G. Baraniuk. 1-bit Compressive Sensing, in Proceedings of Conference on Information Science and Systems (CISS), Princeton, NJ, March 2008.

What's 1-bit compressed sensing

- Model

CS

$$Y_{(m,1)} = A_{(m,n)} x_{(n,1)}$$

$$y_i = < a_i, x >, \quad i = 1,\ldots m$$

1-bit CS

$$y_i = \theta\left(< a_i, x >\right), \quad i = 1,\ldots m \ , \ y_i = \{+1,-1\}$$

$$-1 \le \theta(z) \le 1, \quad \theta(z): \text{Sign, logistic,...}$$

$$y_i = \theta\left(< a_i, x > + v_i\right), \quad i = 1,\ldots m, \ v_i : \text{noise}$$

convex program

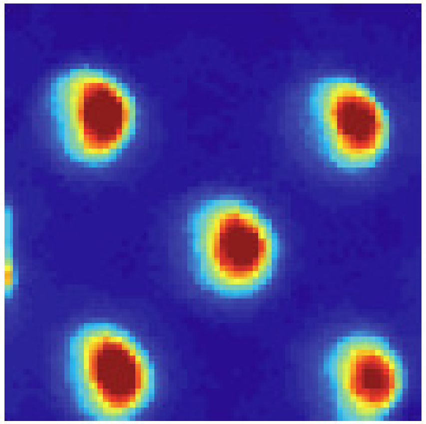
$$\max \sum_{i=1}^{m} y_i < a_i, x' >, \text{ subject to } x' \in K$$

(Convex)

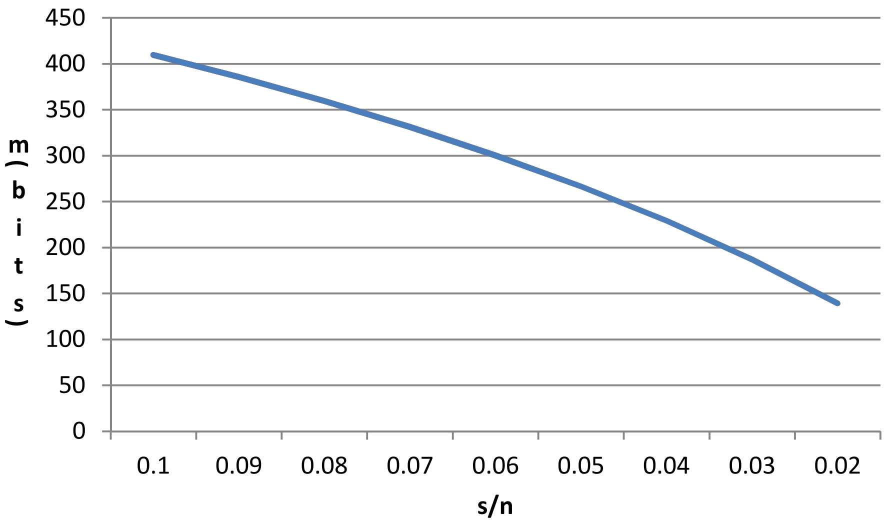
m=?

This paper

- If 1-bit measurement is *noisy*,
  - $m = O(s\log(n/s))$ to recover the signal X
  - *Accuracy* $m \geq C\delta^{-2} w(K)^2$ $\delta \downarrow \Rightarrow m \uparrow$
  - *Probability* $> 1 - 8\exp(-c\delta^2 m)$ $m \uparrow \Rightarrow P \uparrow$



4096 pixels, 8bits = 32 kb
s/n=0.1
1-bit CS→ 410 b



This paper—
3 surprising conclusions

$$y_i = \theta(< a_i, x > + v_i), \qquad v_i \sim N(0, \sigma^2)$$

1. The signal x can be estimated even if each measurement is flopped with probability nearly ½.
2. The signal x can be estimated even when the noise level σ eclipses the magnitude of the linear measurements. $< a_i, x >$
3. If the noise is big enough, nothing is lost by quantizing to 1-bit (minimax error).

This paper

$$y_i = \theta(< a_i, x >)$$

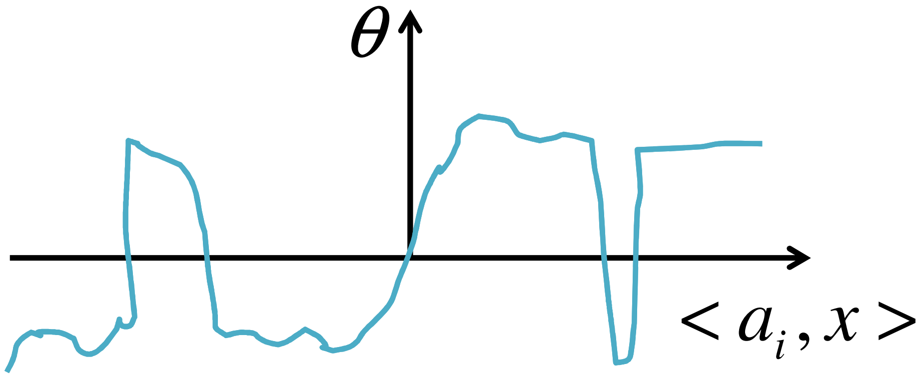
$$E\theta(g)g =: \lambda > 0$$

λ : correlation coefficient

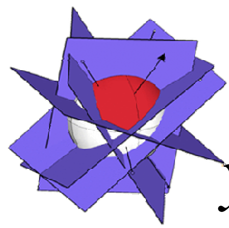
$\because a_i$ are standrard Gaussian random vectors & $\|\mathrm{x}\|_2 = 1$

g : standard Gaussian random vector

$$Ey_i < a_i, x > = E\theta(g)g = \lambda$$



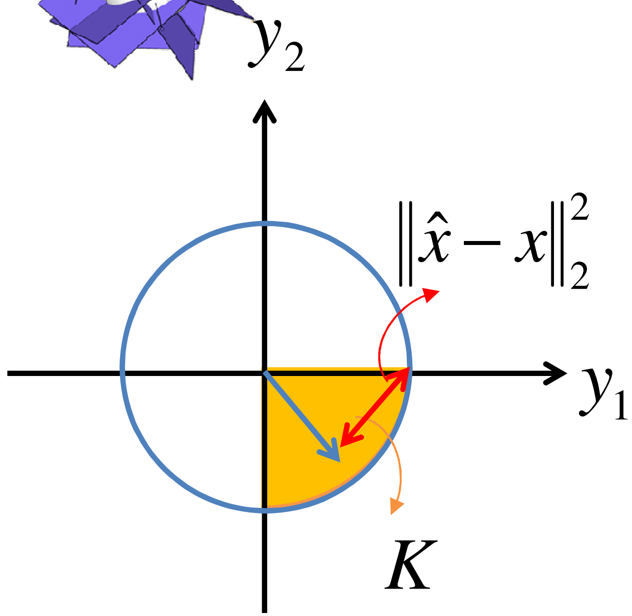
Mean width



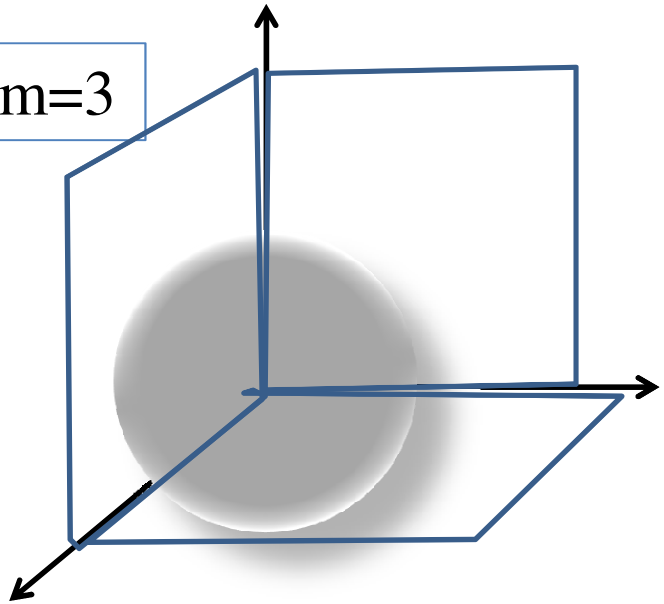
m=2

$$y_i = sign(< a_i, x >), \quad i = 1,2$$

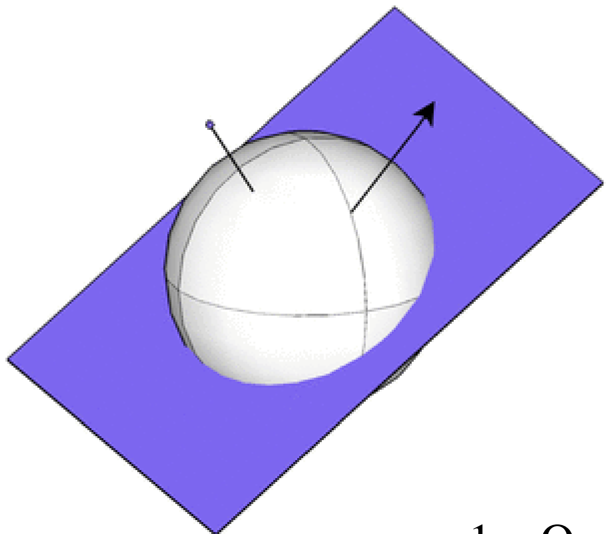
$$Y = sign\left(\begin{bmatrix} 0.6 \\ -0.8 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \Rightarrow \begin{cases} y_1 \geq 0 \\ y_2 < 0 \end{cases}$$



m=3

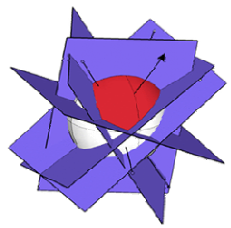


Impossible to exact recovery!



1. O point
2. R or L

Mean width



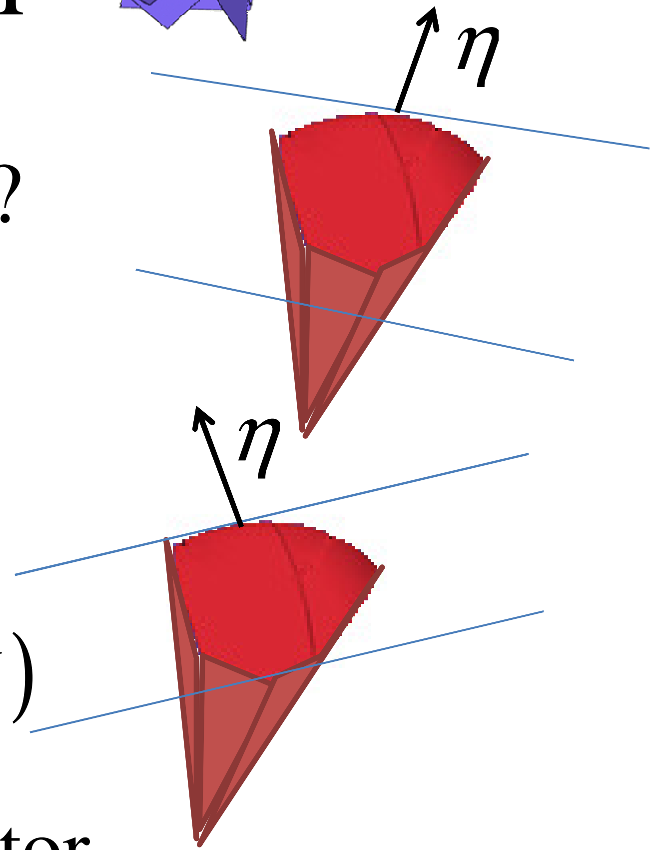
- How to measure the size of K?

$$\sup_{u\in K}\langle \eta, u\rangle - \inf_{v\in K}\langle \eta, v\rangle = \sup_{x\in K-K}\langle \eta, x\rangle$$

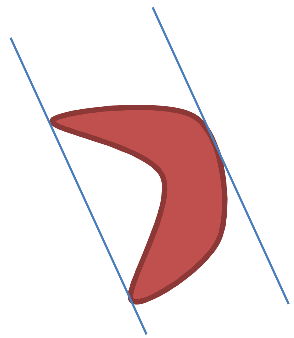
$$\tilde{w}(K) := E \sup_{x\in K-K}\langle \eta, x\rangle$$

$$w(K) := E \sup_{x\in K-K}\langle g, x\rangle, \quad g \sim N(0, I)$$

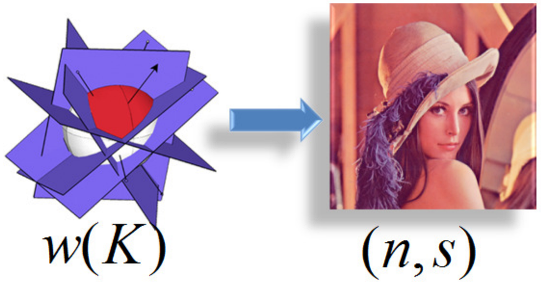
g : standard Gaussian random vector



- Proposition 2.1 (mean width)
  2) $w(K) = w(conv(K))$

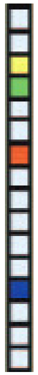


Mean width



- Sparse signal set

$$S_{n,s}=\{x\in R^n:\|x\|_0\le s,\|x\|_2\le 1\}$$



$S_{n,s}$ is the union of $\binom{n}{s}$ - dimension unit Euclidean balls.

$g_T\in R^T$: the restriction of g on to the coordinates in T

Given g in R^n, $\mathrm{T}\subset\{1,2....\mathrm{n}\}$

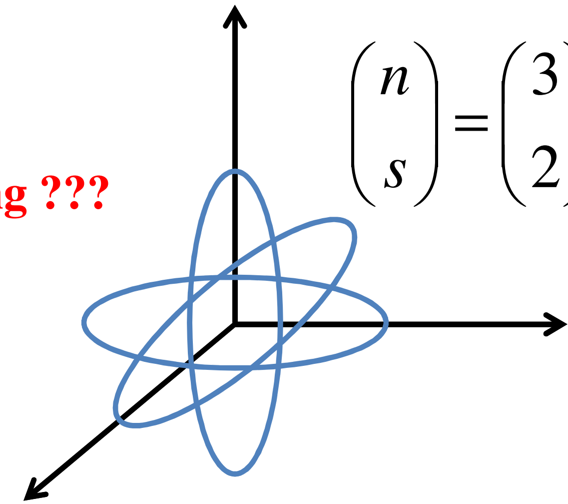
$$w(S_{n,s})=E\max_{|T|=s}\|g_T\|_2$$

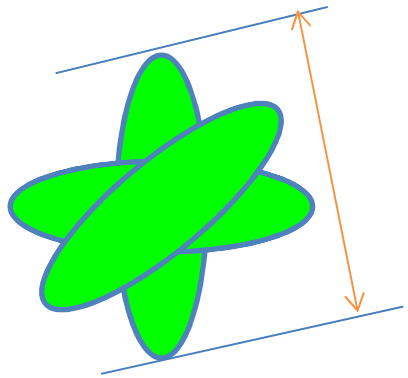
Meaning ???

$$\because \left|\langle g,x\rangle\right|\le\|g\|_2\|x\|_2$$

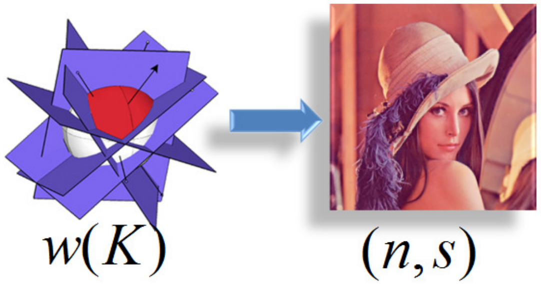
$$P\big(w(S_{n,s})\le f(n,s)\big)>99\%$$

$$\binom{n}{s}=\binom{3}{2}$$



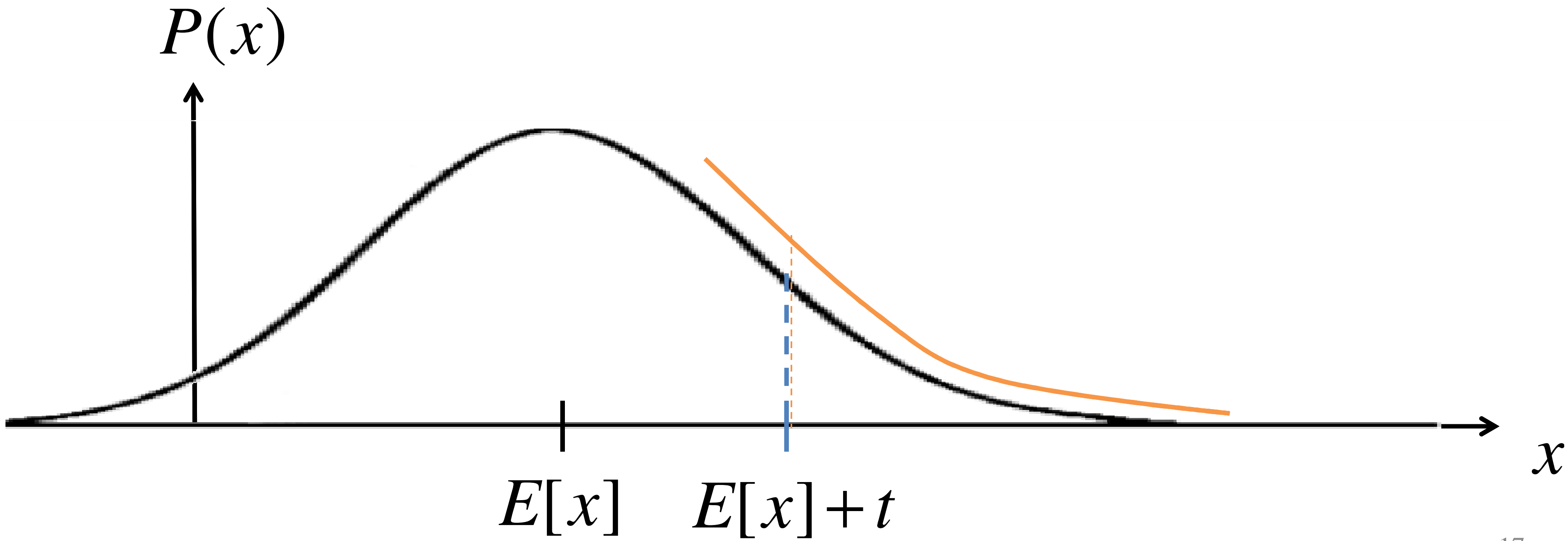


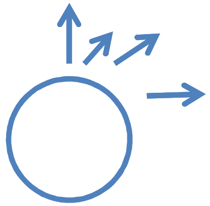
Mean width



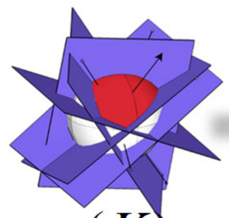
- Theorem 5.2 (Gaussian Concentration inequality)

$$P(x - E[x] \geq t) \leq \exp\left(-\frac{t^2}{2}\right), \textit{where } t > 0$$





Mean width



$w(K)$

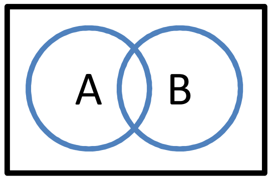
(n,s)

$$P(\|g_T\|_2 \geq E\|g_T\|_2 + t) \leq \exp\left(-\frac{t^2}{2}\right), \text{ for each T, } \quad T \subset \{1,2....n\}$$

$$\because E\|g_T\|_2 \leq \left(\|g_T\|_2^2\right)^{1/2} = \sqrt{s}$$ **($\because$ Jensen's inequality)**

$$P(\|g_T\|_2 \geq \sqrt{s} + t) \leq \exp\left(-\frac{t^2}{2}\right), \text{ for each T, } \quad T \subset \{1,2....n\}$$

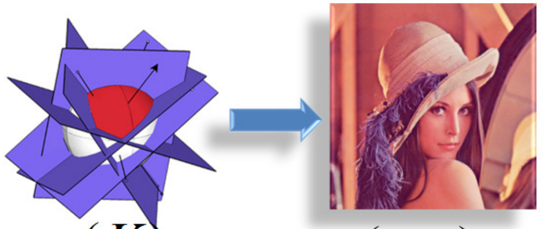
$$\because P(A \cup B) \leq P(A) + P(B)$$ **($\because$ union bound)**



$$P\left(\max_{|T|=s}\|g_T\|_2 \geq \sqrt{s} + t\right) \leq P(\|g_{T1}\|_2 \geq \sqrt{s} + t) + ... + P\left(\left\|g_{T\binom{n}{s}}\right\|_2 \geq \sqrt{s} + t\right)$$

$$\leq \binom{n}{s} \exp\left(-\frac{t^2}{2}\right)$$

Mean width



$w(K)$ (n,s)

$$P(\max_{|T|=s}\|g_T\|_2 \geq \sqrt{s}+t) \leq \binom{n}{s}\exp\left(-\frac{t^2}{2}\right)$$

$$\because \binom{n}{s} \leq \exp\left(s\log(en/s)\right)$$

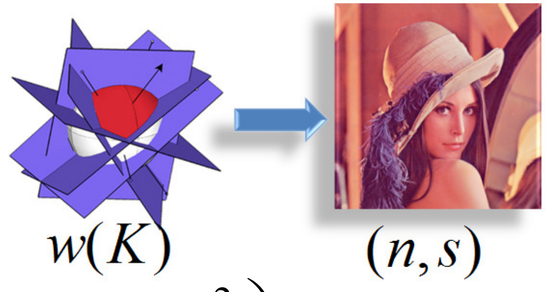
$$P(\max_{|T|=s}\|g_T\|_2 \geq \sqrt{s}+t) \leq \exp\left(s\log(en/s)\right)\exp\left(-\frac{t^2}{2}\right)$$

$$P(\max_{|T|=s}\|g_T\|_2 \geq \sqrt{s}+t) \leq \exp\left(s\log e + s\log(n/s) - \frac{t^2}{2}\right)$$

$$P(\max_{|T|=s}\|g_T\|_2 \leq \sqrt{s}+t) \geq 1-\exp\left(s\log e + s\log(n/s) - \frac{t^2}{2}\right)$$

f(n,s)? **High probability?**

Mean width



$$P(\max_{|T|=s}\|g_T\|_2 \le \sqrt{s}+t) \ge 1-\exp\left(s\log e + s\log(n/s) - \frac{t^2}{2}\right)$$

$$\frac{t^2}{2} \ge s\log e + s\log(n/s)$$

$$t^2 \ge 2s + 2s\log(n/s) \Rightarrow t \ge \sqrt{2s+2s\log(n/s)}$$

$$\sqrt{s}+t = \sqrt{s} + \sqrt{2s+2s\log(n/s)}$$

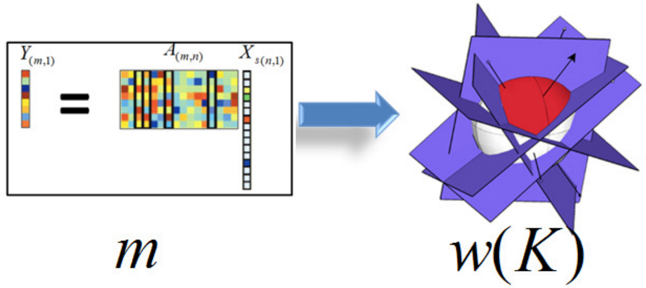
$$\because (a+b)^2 \le 2(a^2+b^2)$$

$$\left(\sqrt{s}+\sqrt{2s+2s\log(n/s)}\right)^2 \le 2(s+2s+2s\log(n/s)) \le Cs\log(n/s)$$

$$E\max_{|T|=s}\|g_T\|_2 \le \max_{|T|=s}\|g_T\|_2$$

$$w^2(S_{n,s}) \le Cs\log(2n/s)$$

Mean width



Theorem 1.1 (Fixed Signal Estimation, Random Noise): Let $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m$ be independent standard Gaussian random vectors in $\mathbb{R}^n$, and let K be a subset of the unit Euclidean ball in $\mathbb{R}^n$. Fix $\boldsymbol{x} \in K$ satisfying $\|\boldsymbol{x}\|_2 = 1$. Assume that the measurements $y_1, \ldots, y_n$ follow the model above.[4] Then for each $\beta > 0$, with probability at least $1 - 4\exp(-2\beta^2)$ the solution $\hat{\boldsymbol{x}}$ to the optimization problem (I.7) satisfies

Robust recovery

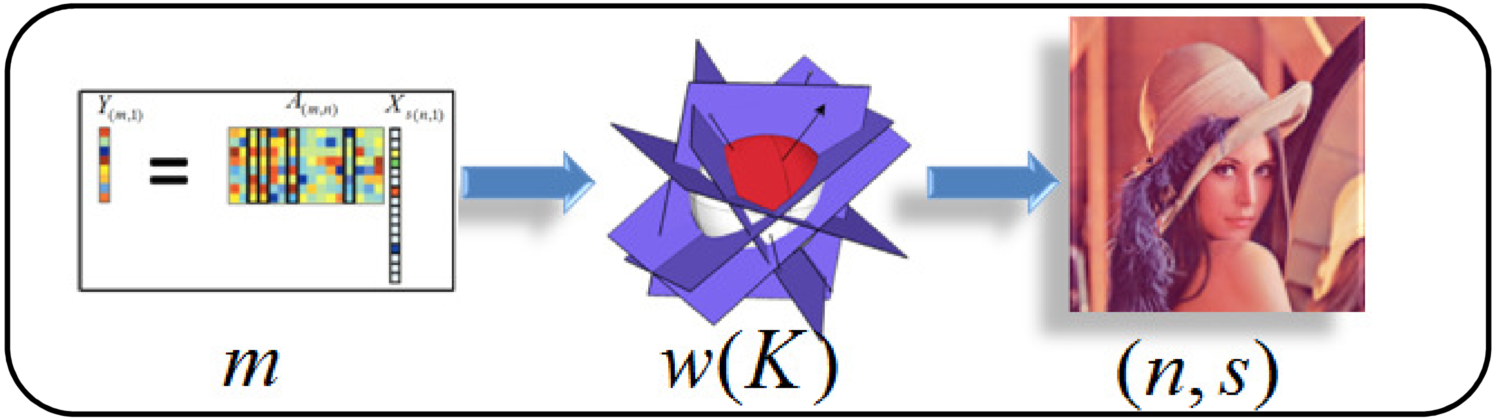
$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2^2 \leq \frac{8}{\lambda\sqrt{m}}(w(K) + \beta).$$

$m = O\left(w(K)^2\right)$

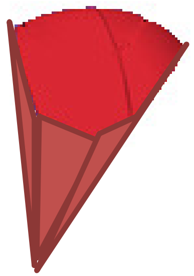
$m \geq C\delta^{-2} w(K)^2, P > 1 - 8\exp(-c\delta^2 m)$ satisfies $\|\hat{x} - x\|_2^2 \leq \delta / \lambda$

δ : Accuracy(defined by users)

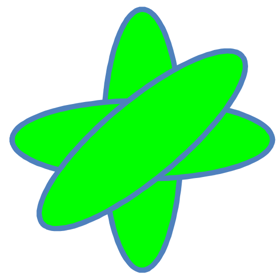
Mean width



$m \geq C\delta^{-2} w(K)^2 \qquad w(S_{n,s})^2 \leq Cs \log(2n/s)$

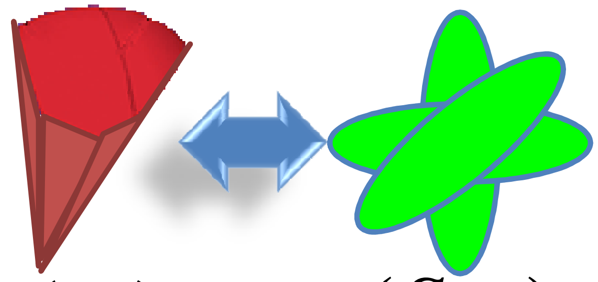


$K \equiv S_{n,s}$?



$$w(K) = w(conv(K))$$

Mean width



If $x \in S_{n,s}, \|x\|_1 \leq \sqrt{s}$ **(∵ Cauchy-Schwarz inequality)**

$$K_{n,s} = \{x \in R^n : \|x\|_2 \leq 1, \|x\|_1 \leq \sqrt{s}\} = B_2^n \cap \sqrt{s} B_1^n$$

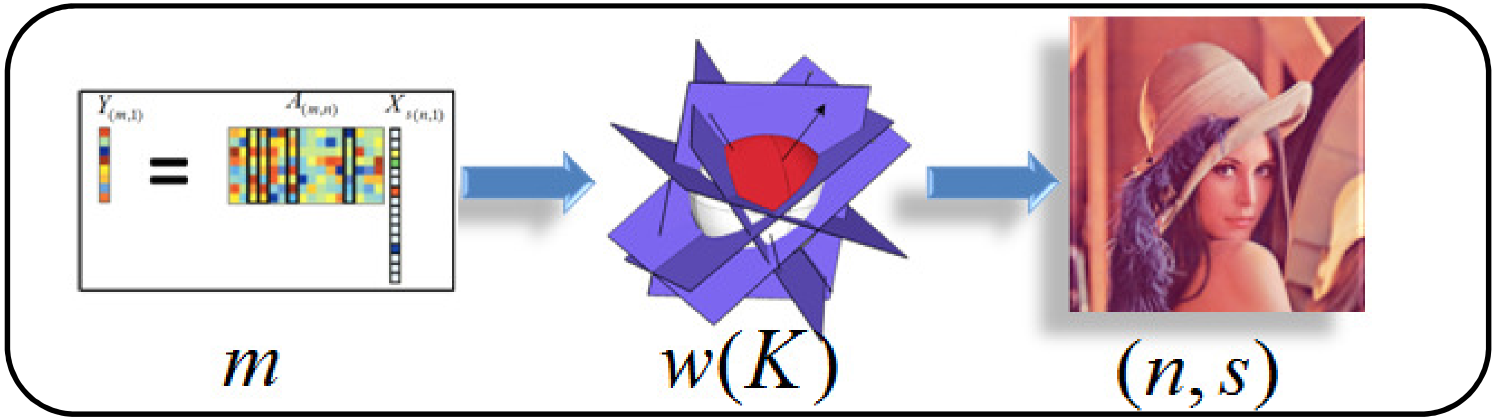
$conv(S_{n,s}) \subset K_{n,s} \subset 2conv(S_{n,s})$ **(∵ Lemma 3.1 [23])**

$$w(K_{n,s}) \leq 2w(conv(S_{n,s})) \leq C\sqrt{s\log(2n/s)}$$

$$\Rightarrow w(K_{n,s}) \leq C\sqrt{s\log(2n/s)}$$

$$\max \sum_{i=1}^{m} y_i < a_i, x' >, \text{ subject to } \|x'\|_2 \leq 1 \text{ and } \|x'\|_1 \leq \sqrt{s}$$

Mean width



$$m \geq C\delta^{-2} w(K)^2 \qquad w(K) \leq C\sqrt{s \log(2n/s)}$$

$$m \geq C\delta^{-2} s \log(2n/s)$$

$$\Rightarrow m = O(s \log(n/s))$$

1-bit compressed sensing

1. The signal x can be estimated even if each measurement is flopped with probability nearly ½.

$$y_i = \xi_i sign(\langle a_i, x \rangle)$$

$$P\{\xi_i = 1\} = p$$

$$E[\xi_1] = 1 \bullet p \bullet sign(z) + (-1) \bullet (1 - p) \bullet sign(z)$$

$$= 2(p - 1/2) sign(z)$$

$$\lambda = 2(p - 1/2) E|g|$$

$$m \geq C\delta^{-2}(p - 1/2)^{-2} s \log(2n / s)$$

1-bit compressed sensing

2. The signal x can be estimated even when the noise level σ eclipses the magnitude of the linear measurements $E\left|\langle a_i, x\rangle\right| = \sqrt{2/\pi}$

$$y_i = \theta\left(< a_i, x > + v_i\right), v_i \sim N(0, \sigma^2)$$

$$\lambda = \sqrt{\frac{2}{\pi\sigma^2}} \exp(-g^2 / 2\sigma^2) = \sqrt{\frac{2}{\pi(\sigma^2+1)}}$$

$$m \geq C\delta^{-2}(\sigma^2+1)s\log(2n/s)$$

$$\left\|\hat{x} - x\right\|_2^2 \leq C\sqrt{\frac{(\sigma^2+1)s\log(2n/s)}{m}}$$

1-bit compressed sensing

3. If the noise is big enough, nothing is lost by quantizing to 1-bit (minimax error)

$$y_i = < a_i, x > + v_i \qquad y_i = \theta\left(< a_i, x > + v_i\right), v_i \sim N(0, \sigma^2)$$

minimax error [24]

$$\delta = c\sigma\sqrt{\frac{s\log(2n/s)}{m}} \qquad \left\|\hat{x} - x\right\|_2^2 \le C\sqrt{\frac{(\sigma^2+1)s\log(2n/s)}{m}}$$

1) Up to a constant
2) $\sigma > 1$

Theorem 1.1

Theorem 1.1 (Fixed Signal Estimation, Random Noise): Let $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m$ be independent standard Gaussian random vectors in $\mathbb{R}^n$, and let K be a subset of the unit Euclidean ball in $\mathbb{R}^n$. Fix $\boldsymbol{x} \in K$ satisfying $\|\boldsymbol{x}\|_2 = 1$. Assume that the measurements $y_1, \ldots, y_n$ follow the model above.[4] Then for each $\beta > 0$, with probability at least $1 - 4\exp(-2\beta^2)$ the solution $\hat{\boldsymbol{x}}$ to the optimization problem (I.7) satisfies

Robust recovery

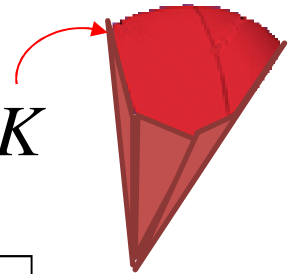
$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2^2 \leq \frac{8}{\lambda\sqrt{m}}(w(K) + \beta).$$

Theorem 1.1

$$f_x(x') = \frac{1}{m}\sum_{i=1}^{m} y_i \langle a_i, x' \rangle$$

$f_x(\hat{x}) \geq f_x(x)$, x is feasible set

$\hat{x}$: solution of $\max \sum_{i=1}^{m} y_i < a_i, x' >$, subject to $x' \in K$



Lemma 4.1 (Expectation): Fix $\boldsymbol{x}' \in \mathbb{R}^n$. Then

$$\mathbb{E} f_{\boldsymbol{x}}(\boldsymbol{x}') = \lambda \langle \boldsymbol{x}, \boldsymbol{x}' \rangle$$

and thus

$$\mathbb{E}[f_{\boldsymbol{x}}(\boldsymbol{x}) - f_{\boldsymbol{x}}(\boldsymbol{x}')] = \lambda(1 - \langle \boldsymbol{x}, \boldsymbol{x}' \rangle) \geq \frac{\lambda}{2} \|\boldsymbol{x} - \boldsymbol{x}'\|_2^2 .$$

Theorem 1.1

Proposition 4.2 (Concentration): For each $t > 0$, we have

$$\left\{\mathbb{P}\left\{\sup_{\boldsymbol{z}\in K-K} |f_{\boldsymbol{x}}(\boldsymbol{z}) - \mathbb{E} f_{\boldsymbol{x}}(\boldsymbol{z})| \geq 4w(K)/\sqrt{m} + t\right\}\right\}$$
$$\leq 4\exp(-mt^2/8).$$

$$0 \le f_x(\hat{x}) - f_x(x) = f_x(\hat{x}-x) \le E\left[f_x(\hat{x}-x)\right] + \frac{4w(K)}{\sqrt{m}} + t$$

$$\because P\left(\sup_{z\in K-K}\left|f_x(z) - Ef_x(z)\right| \le \frac{4w(K)}{\sqrt{m}} + t\right) \ge 1 - 4\exp(-mt^2/8)$$

$$f_x(z) - Ef_x(z) \le \frac{4w(K)}{\sqrt{m}} + t \Rightarrow f_x(z) \le Ef_x(z) + \frac{4w(K)}{\sqrt{m}} + t$$

$$f_x(\hat{x}-x) \le Ef_x(\hat{x}-x) + \frac{4w(K)}{\sqrt{m}} + t$$

Theorem 1.1

> *Lemma 4.1 (Expectation):* Fix $\boldsymbol{x}' \in \mathbb{R}^n$. Then
>
> $$\mathbb{E} f_{\boldsymbol{x}}(\boldsymbol{x}') = \lambda \langle \boldsymbol{x}, \boldsymbol{x}' \rangle$$
>
> and thus
>
> $$\mathbb{E}[f_{\boldsymbol{x}}(\boldsymbol{x}) - f_{\boldsymbol{x}}(\boldsymbol{x}')] = \lambda(1 - \langle \boldsymbol{x}, \boldsymbol{x}' \rangle) \geq \frac{\lambda}{2} \|\boldsymbol{x} - \boldsymbol{x}'\|_2^2 .$$

$$f_x(\hat{x} - x) \leq E\left[f_x(\hat{x} - x)\right] + \frac{4w(K)}{\sqrt{m}} + t$$

$$\leq -\frac{\lambda}{2}\|\hat{x} - x\|_2^2 + \frac{4w(K)}{\sqrt{m}} + t \quad \text{(∵ Lemma 4.1)}$$

choose $t = 4\beta / \sqrt{m}$

$$\|\hat{x} - x\|_2^2 \leq \frac{8}{\lambda\sqrt{m}}(w(K) + \beta) \quad , \; P > 1 - 4\exp(-2\beta)$$

Proposition 4.2

Lemma 5.1 (Symmetrization): Let $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_m$ be independent Rademacher random variables.[8] Then

$$\mu := \mathbb{E} \sup_{\boldsymbol{z} \in K-K} |f_{\boldsymbol{x}}(\boldsymbol{z}) - \mathbb{E} f_{\boldsymbol{x}}(\boldsymbol{z})| \leq 2\mathbb{E} \sup_{\boldsymbol{z} \in K-K} \frac{1}{m} \left| \sum_{i=1}^{m} \varepsilon_i y_i \langle \boldsymbol{a}_i, \boldsymbol{z} \rangle \right|. \quad \text{(V.1)}$$

Furthermore, we have the deviation inequality

$$\mathbb{P}\left\{ \sup_{\boldsymbol{z} \in K-K} |f(\boldsymbol{z}) - \mathbb{E} f(\boldsymbol{z})| \geq 2\mu + t \right\} \leq 4 \left\{ \sup_{\boldsymbol{z} \in K-K} \left| \sum_{i=1}^{m} \varepsilon_i y_i \langle \boldsymbol{a}_i, \boldsymbol{z} \rangle \right| > t/2 \right\}. \quad \text{(V.2)}$$

[19] M. Ledoux and M. Talagrand*, Probability in Banach Spaces: Isoperimetry and Processes. Berlin, Germany: Springer-Verlag,* 1991. [chapter 6.1]

Proposition 4.2

$$\sup_{z\in K-K}\frac{1}{m}\left|\sum_{i=1}^{m}\varepsilon_i y_i\langle a_i,z\rangle\right| \overset{dist}{=} \sup_{z\in K-K}\frac{1}{m}\left|\sum_{i=1}^{m}\langle a_i,z\rangle\right| \overset{dist}{=} \frac{1}{\sqrt{m}}\sup_{z\in K-K}\left|\langle g,z\rangle\right|$$

$$=\frac{1}{\sqrt{m}}\sup_{z\in K-K}\langle g,z\rangle \qquad (\text{V}.4)$$

$$E\sup_{z\in K-K}\left|f_x(z)-Ef_x(z)\right| \le \frac{2}{\sqrt{m}}E\sup_{z\in K-K}\langle g,z\rangle = \frac{2w(K)}{\sqrt{m}} \qquad (\text{V}.5)$$

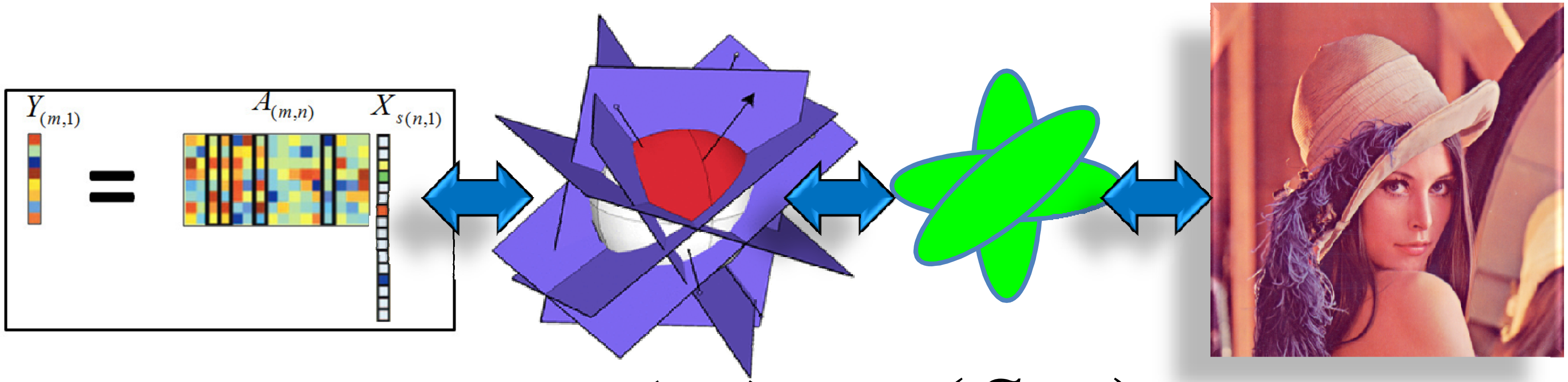
$$P\left\{\sup_{z\in K-K}\left|f_x(z)-Ef_x(z)\right| \ge \frac{4w(K)}{\sqrt{m}}+t\right\} \le 4P\left\{\frac{1}{\sqrt{m}}\sup_{z\in K-K}\langle g,z\rangle > \frac{t}{2}\right\}$$

$$\because P\left\{\sup_{z\in K-K}\langle g,z\rangle \ge w(K)+r\right\} \le \exp(-\frac{r^2}{2}),\text{choose } r=t\sqrt{m}/2$$

$$P\left\{\sup_{z\in K-K}\left|f_x(z)-Ef_x(z)\right| \ge \frac{4w(K)}{\sqrt{m}}+t\right\} \le 4\exp(-t^2 m/8)$$

Conclusions

- If 1-bit measurement is *noisy*,
  - $m = O(s \log(n/s))$ to recover the signal X
  - *Accuracy:* $m \geq C\delta^{-2} w(K)^2$
  - *Probability* $> 1 - 8\exp(-c\delta^2 m)$



m $\quad w(K) \quad w(S_{n,s}) \quad (n, s)$

Reference

1. Yaniv Plan and Roman Vershynin, ”**Robust 1-bit compressed sensing and sparse logistic regression: A convex programming approach**” IEEE Transactions on Information Theory, 2013.
2. http://dsp.rice.edu/1bitCS/ (slides & papers)
3. Richard G. Baraniuk, “**Compressive Sensing**,” IEEE Signal Processing Magazine, 2007.
4. Qaisar S., Bilal R.M., Wafa Iqbal, Naureen M. and Lee S., “**Compressive Sensing: From Theory to Applications A Survey**,” Journal of Communications and Networks, 2013.