Distributed Compressed Sensing with One-Bit Measurements and Hard Thresholding

Lars Palzer

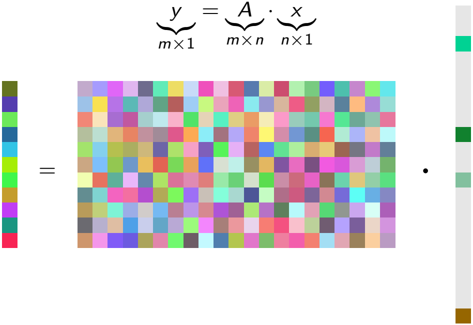
Joint work with Johannes Maly

2nd TUM-EURECOM Workshop on Communications and Machine Learning
May 17, 2018

Compressed Sensing

Compressed Sensing

$$\underbrace{y}_{m\times 1} = \underbrace{A}_{m\times n} \cdot \underbrace{x}_{n\times 1}$$

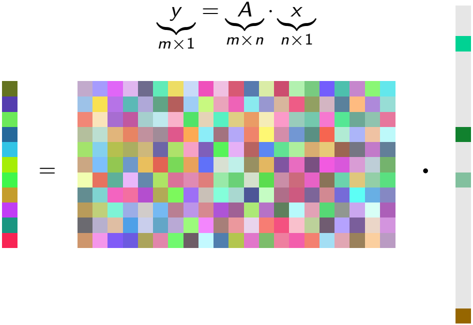


- Reconstruct "undersampled" signal $x \in \mathbb{R}^n$ from m linear measurements
- Basic assumption: $\|x\|_0 \leq s \ll n$

S. Foucart, H. Rauhut, "A Mathematical Introduction to Compressive Sensing", 2013.

Compressed Sensing

$$\underbrace{y}_{m\times 1} = \underbrace{A}_{m\times n} \cdot \underbrace{x}_{n\times 1}$$



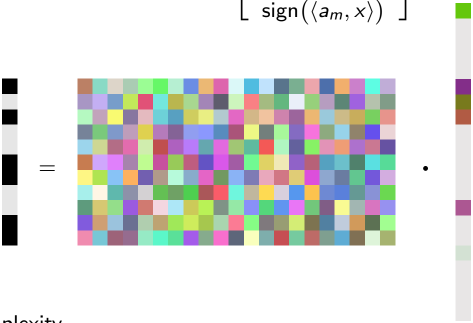
- Reconstruct "undersampled" signal $x \in \mathbb{R}^n$ from m linear measurements
- Basic assumption: $\|x\|_0 \leq s \ll n$
- Tractable uniform recovery if $m \geq C \cdot s \ln(en/s)$
- Tractable non-uniform recovery if $m > 2s \ln(en/s)$

S. Foucart, H. Rauhut, "A Mathematical Introduction to Compressive Sensing", 2013.

1-Bit Compressed Sensing

1-Bit Compressed Sensing

$$y = \operatorname{sign}\big(A \cdot x\big) = \begin{bmatrix} \operatorname{sign}\big(\langle a_1, x\rangle\big) \\ \vdots \\ \operatorname{sign}\big(\langle a_m, x\rangle\big) \end{bmatrix}$$



- \+ Very low complexity
- — Very coarse quantization
- — Amplitude information lost

Signal Approximation

Theorem (Foucart)

If A satisfies $RIP_1(2s, \delta)$, i.e.,

$$(1-\delta)\,\|z\|_{2} \leq \|Az\|_{1} \leq (1+\delta)\,\|z\|_{2}$$

for all $2s$-sparse vectors $z \in \mathbb{R}^n$, every ℓ_2-normalized s-sparse vector $x \in \mathbb{R}^n$ observed via $y = \mathrm{sign}(A \cdot x)$ has

$$\left\| x - \mathbb{H}_s(A^{\mathsf{T}} y) \right\|_2 \lesssim \sqrt{\delta},$$

where $\mathbb{H}_s(\cdot)$ keeps only the s largest (in magnitude) entries.

S. Foucart, "Flavors of compressive sensing", Int. Conf. Approx. Th., 2016

Y. Plan, R. Vershynin, "One-bit compressed sensing by linear programming", Comm. Pure Appl. Math., 2013

L. Jacques, J. N. Laska, P. T. Boufounos, and R. G. Baraniuk, "Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors," IEEE Trans. Inf. Theory, 2013.

Signal Approximation

Theorem (Foucart)

If A satisfies $RIP_1(2s, \delta)$, i.e.,

$$(1-\delta)\,\|z\|_2 \le \|Az\|_1 \le (1+\delta)\,\|z\|_2$$

for all $2s$-sparse vectors $z \in \mathbb{R}^n$, every ℓ_2-normalized s-sparse vector $x \in \mathbb{R}^n$ observed via $y = \operatorname{sign}(A \cdot x)$ has

$$\left\| x - \mathbb{H}_s(A^{\mathsf{T}} y) \right\|_2 \lesssim \sqrt{\delta},$$

where $\mathbb{H}_s(\cdot)$ keeps only the s largest (in magnitude) entries.

- $\mathrm{RIP}_1(2s, \delta)$ satisfied w.h.p. if entries of A are iid Gaussian and $m \gtrsim \delta^{-2} s \ln(en/s)$

S. Foucart, "Flavors of compressive sensing", Int. Conf. Approx. Th., 2016

Y. Plan, R. Vershynin, "One-bit compressed sensing by linear programming", Comm. Pure Appl. Math., 2013

L. Jacques, J. N. Laska, P. T. Boufounos, and R. G. Baraniuk, "Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors," IEEE Trans. Inf. Theory, 2013.

Signal Approximation

Theorem (Foucart)

If A satisfies RIP$_1(2s, \delta)$, i.e.,

$$(1-\delta)\|z\|_2 \leq \|Az\|_1 \leq (1+\delta)\|z\|_2$$

for all $2s$-sparse vectors $z \in \mathbb{R}^n$, every ℓ_2-normalized s-sparse vector $x \in \mathbb{R}^n$ observed via $y = \mathrm{sign}(A \cdot x)$ has

$$\left\| x - \mathbb{H}_s(A^{\mathsf{T}} y) \right\|_2 \lesssim \sqrt{\delta},$$

where $\mathbb{H}_s(\cdot)$ keeps only the s largest (in magnitude) entries.

- RIP$_1(2s, \delta)$ satisfied w.h.p. if entries of A are iid Gaussian and $m \gtrsim \delta^{-2} s \ln(en/s)$

Tractable *and* **accurate** *sparse approximation possible from 1-bit measurements!*

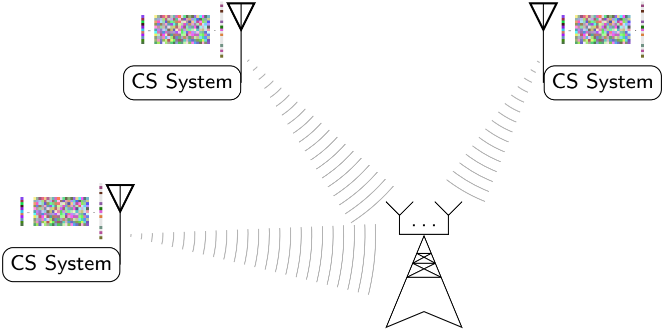
S. Foucart, "Flavors of compressive sensing", Int. Conf. Approx. Th., 2016

Y. Plan, R. Vershynin, "One-bit compressed sensing by linear programming", Comm. Pure Appl. Math., 2013

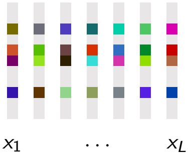
L. Jacques, J. N. Laska, P. T. Boufounos, and R. G. Baraniuk, "Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors," IEEE Trans. Inf. Theory, 2013.

Distributed Compressed Sensing

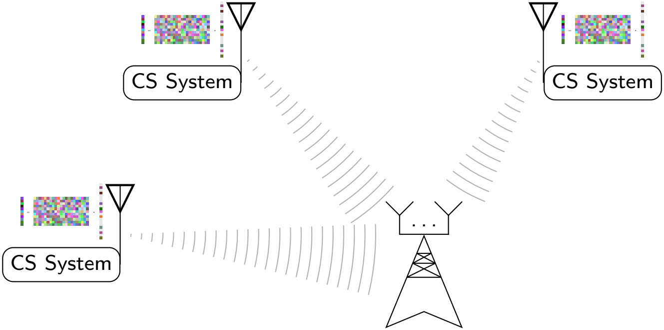
Motivation



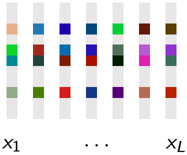
- Distributed sensing, joint reconstruction
- Signals share common structure
- Applications such as sensor networks, MRI or MIMO systems



Motivation



- Distributed sensing, joint reconstruction
- Signals share common structure
- Applications such as sensor networks, MRI or MIMO systems



Idea: use joint signal structure to reduce measurement rates

Distributed CS with 1-Bit Measurements

(our work)

Model

- Signal $X = [x_1, \ldots, x_L] \in \mathcal{S}_{s,L}$ with

$$\mathcal{S}_{s,L} = \left\{ Z \in \mathbb{R}^{n \times L} \colon \underbrace{|\mathrm{supp}(Z)| \leq s}_{\text{row-sparsity}},\ \underbrace{\|z_\ell\|_2 = \frac{\|Z\|_F}{\sqrt{L}} \ \forall \ell \in [L]}_{\text{normalized columns}} \right\}.$$

Model

- Signal $X = [x_1, \ldots, x_L] \in \mathcal{S}_{s,L}$ with

$$\mathcal{S}_{s,L} = \left\{ Z \in \mathbb{R}^{n \times L} \colon \underbrace{|\mathrm{supp}(Z)| \leq s}_{\text{row-sparsity}}, \underbrace{\|z_\ell\|_2 = \frac{\|Z\|_F}{\sqrt{L}} \;\; \forall \ell \in [L]}_{\text{normalized columns}} \right\}.$$

- Distributed 1-bit sensing with properly scaled Gaussian A_ℓ:

$$\underbrace{y_\ell}_{m \times 1} = \mathrm{sign}\Big(\underbrace{A_\ell}_{m \times n} \underbrace{x_\ell}_{n \times 1} \Big), \quad \ell \in [L]$$

Model

- Signal $X = [x_1, \ldots, x_L] \in \mathcal{S}_{s,L}$ with

$$\mathcal{S}_{s,L} = \left\{ Z \in \mathbb{R}^{n \times L} \colon \underbrace{|\mathrm{supp}(Z)| \leq s}_{\text{row-sparsity}}, \underbrace{\|z_\ell\|_2 = \frac{\|Z\|_F}{\sqrt{L}} \ \forall \ell \in [L]}_{\text{normalized columns}} \right\}.$$

- Distributed 1-bit sensing with properly scaled Gaussian A_ℓ:

$$\underbrace{y_\ell}_{m \times 1} = \mathrm{sign}\Big(\underbrace{A_\ell}_{m \times n} \underbrace{x_\ell}_{n \times 1} \Big), \quad \ell \in [L]$$

- Reconstruction:

$$\hat{X} = [\hat{x}_1, \ldots, \hat{x}_L] = \underbrace{\mathbb{H}_s\Big(A_1^T y_1, \ldots, A_L^T y_L \Big)}_{\text{hard thresholding}}$$

  where $\mathbb{H}_s(\cdot)$ keeps the s largest (in $\|\cdot\|_2$) rows.

Model

- Signal $X = [x_1, \ldots, x_L] \in \mathcal{S}_{s,L}$ with

$$\mathcal{S}_{s,L} = \left\{ Z \in \mathbb{R}^{n \times L} \colon \underbrace{|\mathrm{supp}(Z)| \leq s}_{\text{row-sparsity}}, \underbrace{\|z_\ell\|_2 = \frac{\|Z\|_F}{\sqrt{L}} \;\; \forall \ell \in [L]}_{\text{normalized columns}} \right\}.$$

- Distributed 1-bit sensing with properly scaled Gaussian A_ℓ:

$$\underbrace{y_\ell}_{m \times 1} = \mathrm{sign}\Big(\underbrace{A_\ell}_{m \times n} \underbrace{x_\ell}_{n \times 1} \Big), \quad \ell \in [L]$$

- Reconstruction:

$$\hat{X} = [\hat{x}_1, \ldots, \hat{x}_L] = \underbrace{\mathbb{H}_s\Big(A_1^\mathsf{T} y_1, \ldots, A_L^\mathsf{T} y_L \Big)}_{\text{hard thresholding}}$$

  where $\mathbb{H}_s(\cdot)$ keeps the s largest (in $\|\cdot\|_2$) rows.

Question: how many measurements per signal are necessary?

Main Result

Theorem

Fix $X \in \mathcal{S}_{s,L}$ *with* $\|X\|_F = 1$. *Choose*

$$mL \gtrsim \delta^{-2} s \Big(\ln\big(en/s\big) + L \Big)$$

and let $y_\ell = \operatorname{sign}(A_\ell x_\ell)$ *for* $\ell \in [L]$. *Then, with probability (over the* A_ℓ*s) at least* $1 - \exp\big(-c\delta^2 mL\big)$,

$$\Big\| X - \hat{X} \Big\|_F \lesssim \sqrt{\delta}$$

where $\hat{X}$ *is the result of a single hard thresholding step.*

Main Result

Theorem

Fix $X \in \mathcal{S}_{s,L}$ *with* $\|X\|_F = 1$. *Choose*

$$mL \gtrsim \delta^{-2} s \Big(\ln(en/s) + L\Big)$$

and let $y_\ell = \mathrm{sign}(A_\ell x_\ell)$ *for* $\ell \in [L]$. *Then, with probability (over the* A_ℓ*s) at least* $1 - \exp\left(-c\delta^2 mL\right)$,

$$\left\|X - \hat{X}\right\|_F \lesssim \sqrt{\delta}$$

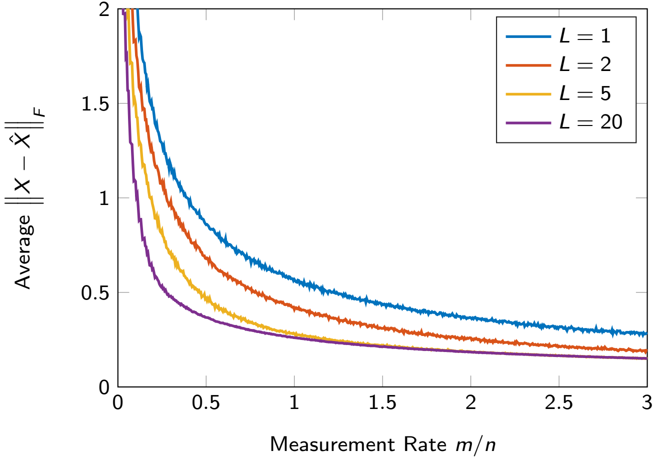
where $\hat{X}$ *is the result of a single hard thresholding step.*

Remarks:

- For $L \gtrsim \ln(en/s)$, we need effectively $m \gtrsim \delta^{-2} s$ measurements per signal
- Non-uniform recovery result
- Bounds total error of the L signals

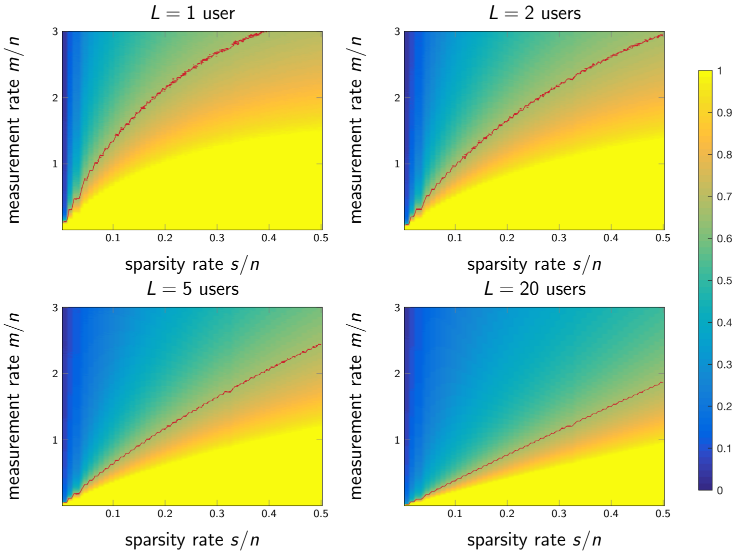
Numerical Experiments

Numerical Experiments (1)



$n = 100$, $s = 5$, 500 trials, $\|X\|_F = 1$,
$x_1, \ldots, x_L$ uniform on sphere in a random s-dim subspace of $\mathbb{R}^n$

Numerical Experiments (2)



Average error $\|X - \hat{X}\|_F$ over 500 trials, $n = 100$

Proof Strategy

Proof Ingredients

- Derive[1] appropriate RIP for $A_1, \ldots, A_L$ and $\mathcal{S}_{s,L}$

[1] inspired by Y. Plan and R. Vershynin, "Dimension reduction by random hyperplane tesselation," Discrete & Computational Geometry, 2014.

$\mathrm{RIP}_{2,1}$

Denote

$$A = \begin{pmatrix} \boxed{A_1} & & \\ & \ddots & \\ & & \boxed{A_L} \end{pmatrix}$$

and recall

$$\mathcal{S}_{s,L} = \Big\{ Z \in \mathbb{R}^{n \times L} \colon \underbrace{|\mathrm{supp}(Z)| \le s}_{\text{row-sparsity}},\ \underbrace{\|z_\ell\|_2 = \frac{\|Z\|_F}{\sqrt{L}} \ \ \forall \ell \in [L]}_{\text{normalized columns}} \Big\}.$$

Lemma ($\mathrm{RIP}_{2,1}$)

Let the entries of A_ℓ, $\ell \in [L]$, be iid $\mathcal{N}\big(0, \pi/(2Lm^2)\big)$. For

$$mL \gtrsim \delta^{-2} s \Big(\ln(en/s) + L \Big)$$

the A satisfies

$$(1-\delta)\,\|X\|_F \le \|A \cdot \mathrm{vec}(X)\|_1 \le (1+\delta)\,\|X\|_F$$

for all $X \in \mathcal{S}_{s,L}$ with probability at least $1 - \exp\left(-c\delta^2 mL\right)$.

Proof Ingredients

- Derive[2] appropriate RIP for $A_1, \dots, A_L$ and $\mathcal{S}_{s,L}$
- Show concentration of $\left\| \left[A_1^\mathsf{T} y_1, \dots, A_L^\mathsf{T} y_L \right]_S \right\|_F$

[2] inspired by Y. Plan and R. Vershynin, "Dimension reduction by random hyperplane tesselation," Discrete & Computational Geometry, 2014.

Concentration of $\left\|\left[A_1^{\mathsf{T}} y_1, \ldots, A_L^{\mathsf{T}} y_L\right]_S\right\|_F$

Lemma

Let $B = \underbrace{\left[A_1^{\mathsf{T}} y_1, \ldots, A_L^{\mathsf{T}} y_L\right]_S}_{\textit{restricted to } \mathrm{supp}(X)} \in \mathrm{R}^{n \times L}$. *Then,*

$$\mathbb{P}\left(\left|\|B\|_F - \sqrt{\mathrm{E}\left[\|B\|_F^2\right]}\right| \geq \delta\right) \leq 2 \exp\left(-c \min\left\{\frac{\delta^4 m^2}{s^2}, \frac{\delta^2 m}{s}\right\}\right).$$

- Concentration for a single X if $m \gtrsim \delta^{-2} s$
- Since $\left\|A_\ell^{\mathsf{T}} y_\ell\right\|_2$ is independent of ℓ, we can find $\tilde{B} \in \mathcal{S}_{s,L}$ close to B

Proof Ingredients

- Derive[3] appropriate RIP for $A_1, \ldots, A_L$ and $\mathcal{S}_{s,L}$
- Show concentration of $\left\| \left[A_1^\top y_1, \ldots, A_L^\top y_L \right]_S \right\|_F$
- Conclude that[4] $\left\| X - \hat{X} \right\|_F \leq 2\sqrt{10\delta}$

[3] inspired by Y. Plan and R. Vershynin, "Dimension reduction by random hyperplane tesselation," Discrete & Computational Geometry, 2014.

[4] inspired by S. Foucart, "Flavors of compressive sensing", Int. Conf. Approx. Th., 2016

Summary

Summary / Remarks

- Non-uniform approximation result for distributed 1-bit CS with hard thresholding
- For $L \gtrsim \ln(en/s)$, measurements per signal scale linearly in the row-sparsity
- Theoretical scaling clearly visible in numerical experiments

Thanks!

arXiv paper:

"Analysis of Hard-Thresholding for Distributed Compressed Sensing with One-Bit Measurements",

https://arxiv.org/abs/1805.03486