Sparsity in imaging: Compressed sensing

Clarice Poon
University of Bath

January, 2019

Compressed sensing

Candés, Romberg & Tao (2006); Donoho (2006)

Task: Given $y_0 = Ax_0$ where $A : \mathbb{R}^N \to \mathbb{R}^m$ with $N \gg m$, recover x_0.

In general, this is impossible, since we have more unknowns than knowns.

Compressed sensing

Candés, Romberg & Tao (2006); Donoho (2006)

Task: Given $y_0 = Ax_0$ where $A : \mathbb{R}^N \to \mathbb{R}^m$ with $N \gg m$, recover x_0.

In general, this is impossible, since we have more unknowns than knowns.

Suppose that $f \in \mathbb{C}^N$ is sparse in some orthonormal basis Ψ. That is, $f = \Psi x_0$ for some sparse vector x_0. Then,

$$y = Af = A \circ \Psi x_0 = \Phi x_0.$$

Solve instead

$$\min_x \|x\|_0 \text{ subject to } \Phi x = y$$

Compressed sensing

Candés, Romberg & Tao (2006); Donoho (2006)

Task: Given $y_0 = Ax_0$ where $A : \mathbb{R}^N \to \mathbb{R}^m$ with $N \gg m$, recover x_0.

In general, this is impossible, since we have more unknowns than knowns.

Suppose that $f \in \mathbb{C}^N$ is sparse in some orthonormal basis Ψ. That is, $f = \Psi x_0$ for some sparse vector x_0. Then,

$$y = Af = A \circ \Psi x_0 = \Phi x_0.$$

Solve instead

$$\min_x \|x\|_0 \text{ subject to } \Phi x = y$$

- Even though this problem might have a unique solution, it is nontrivial to find an algorithm to solve this.

Compressed sensing

Candés, Romberg & Tao (2006); Donoho (2006)

Task: Given $y_0 = Ax_0$ where $A : \mathbb{R}^N \to \mathbb{R}^m$ with $N \gg m$, recover x_0.

In general, this is impossible, since we have more unknowns than knowns.

Suppose that $f \in \mathbb{C}^N$ is sparse in some orthonormal basis Ψ. That is, $f = \Psi x_0$ for some sparse vector x_0. Then,

$$y = Af = A \circ \Psi x_0 = \Phi x_0.$$

Solve instead

$$\min_x \|x\|_0 \text{ subject to } \Phi x = y$$

- Even though this problem might have a unique solution, it is nontrivial to find an algorithm to solve this.
- Naively, we can attempt to solve $A_S u = y$ for all subsets S of size s. However, it is unpractical to check all $\binom{N}{s}$ such subsets! E.g. if $N = 1000$, $s = 10$, then there are $\binom{1000}{10} \geqslant (1000/10)^{10} = 10^{20}$ linear systems of size 10×10. Even if each system is solved in 10^{-10}s, this approach requires 10^{10} s > 300 years.

Compressed sensing

Candés, Romberg & Tao (2006); Donoho (2006)

Task: Given $y_0 = Ax_0$ where $A : \mathbb{R}^N \to \mathbb{R}^m$ with $N \gg m$, recover x_0.

In general, this is impossible, since we have more unknowns than knowns.

Suppose that $f \in \mathbb{C}^N$ is sparse in some orthonormal basis Ψ. That is, $f = \Psi x_0$ for some sparse vector x_0. Then,

$$y = Af = A \circ \Psi x_0 = \Phi x_0.$$

Solve instead

$$\min_x \|x\|_0 \text{ subject to } \Phi x = y$$

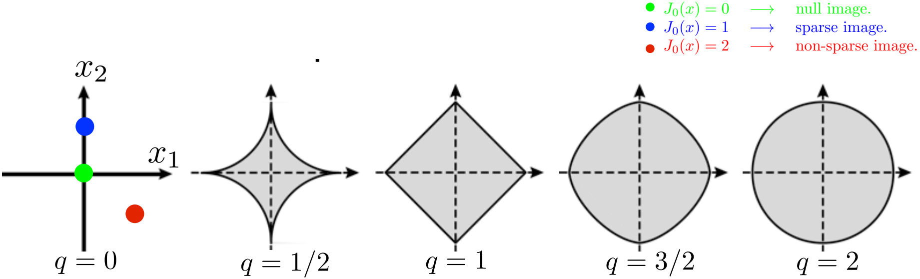
- Even though this problem might have a unique solution, it is nontrivial to find an algorithm to solve this.
- Naively, we can attempt to solve $A_S u = y$ for all subsets S of size s. However, it is unpractical to check all $\binom{N}{s}$ such subsets! E.g. if $N = 1000$, $s = 10$, then there are $\binom{1000}{10} \geqslant (1000/10)^{10} = 10^{20}$ linear systems of size 10×10. Even if each system is solved in 10^{-10}s, this approach requires 10^{10} s > 300 years.
- In general, the ℓ^0 problem can be transposed into an NP-hard problem.

Compressed sensing

Candés, Romberg & Tao (2006); Donoho (2006)

Let $\|x\|_q^q = \sum_j |x_j|^q$. Convex when $q \geqslant 1$ and "close to" ℓ_0 for small q.

$$\min_x \|x\|_p \text{ subject to } \Phi x = y$$



Compressed sensing

Candés, Romberg & Tao (2006); Donoho (2006)

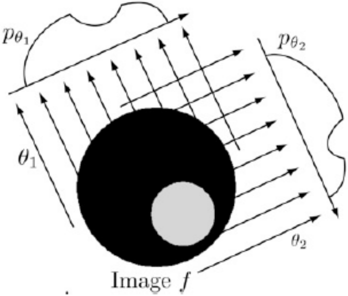
Key outcome of compressed sensing:
We can recover sparse vectors of length N from $m \ll N$ randomised linear measurements by solving the following convex optimisation problem:

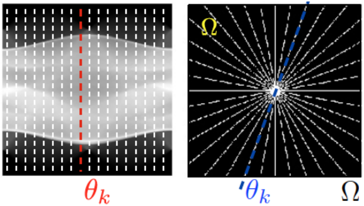
$$\min_x \|x\|_1 \text{ subject to } \Phi x = y. \tag{BP}$$

Applications of compressed sensing – Fourier measurements

Many imaging devices can be seen as providing pointwise samples of the Fourier transform.

- Magnetic resonance imaging
- Radio interferometry
- Electron microscopy
- Tomography.





For tomography, if p_θ is the Radon projection of f at angle θ, then the Fourier splice theorem says:

$$\hat{p}_\theta(t) = \hat{f}(t\cos(\theta), t\sin(\theta)).$$

We therefore are interested in $y = P_\Omega \mathcal{F} W x$.

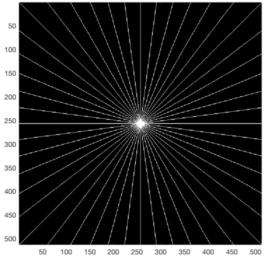
The matlab phantom experiment [Candès, Romberg and Tao '06]

Let $P_\Omega \mathcal{F} x = (\hat{x}_j)_{j\in\Omega}$. Given observations $y_0 = P_\Omega \mathcal{F} x_0$, take the reconstruction z as

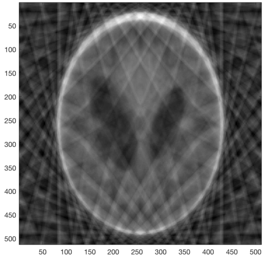
$$\operatorname*{argmin}_x \|Wx\|_1 \text{ subject to } P_\Omega \mathcal{F} x = y_0$$

If W is invertible, this is equivalent to

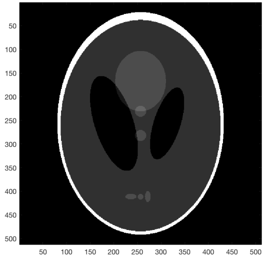
$$\operatorname*{argmin}_x \|x\|_1 \text{ subject to } P_\Omega \mathcal{F} W^{-1} x = y_0$$



Sampling map Ω



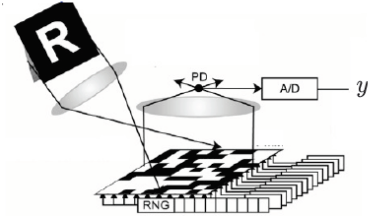
$\mathcal{F}^{-1} P_\Omega y_0$



Sparse reconstruction

The single pixel camera [Duarte, Davenporte, Takhar, Laska, Sun, Kelly, Baraniuk '08]

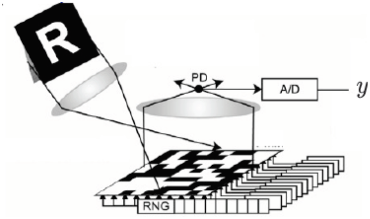
- Let $z = \mathbb{R}^N$.
- The single pixel camera is a microarry consisting of N mirrors, each of which can be switched on or off individually.
- The light from the image is reflected on the micro array, and a lens then combines all reflected beams in one sensor.



Each measurement is $\langle z, b \rangle$ where b is a vector consisting of 1's at locations where the mirrors are 'on' and 0 where the mirrors are 'off'.

The single pixel camera [Duarte, Davenporte, Takhar, Laska, Sun, Kelly, Baraniuk '08]

- Let $z = \mathbb{R}^N$.
- The single pixel camera is a microarry consisting of N mirrors, each of which can be switched on or off individually.
- The light from the image is reflected on the micro array, and a lens then combines all reflected beams in one sensor.



Each measurement is $\langle z, b \rangle$ where b is a vector consisting of 1's at locations where the mirrors are 'on' and 0 where the mirrors are 'off'.

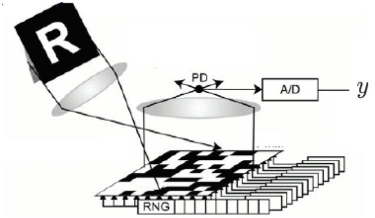
$m/N = 1$

$m/N = 0.16$

$m/N = 0.02$

The single pixel camera [Duarte, Davenporte, Takhar, Laska, Sun, Kelly, Baraniuk '08]

- Let $z = \mathbb{R}^N$.
- The single pixel camera is a microarry consisting of N mirrors, each of which can be switched on or off individually.
- The light from the image is reflected on the micro array, and a lens then combines all reflected beams in one sensor.



Each measurement is $\langle z,\, b\rangle$ where b is a vector consisting of 1's at locations where the mirrors are 'on' and 0 where the mirrors are 'off'.

Link to Bernoulli measurements

We can think of this as recovering sparse x from $y = Az = AW^* x$. where $A \in \{-1, 1\}^{m\times N}$ a Bernoulli random matrix (entries take values ± 1 with equal probability:
if $a \in \{-1, 1\}^N$ is a Bernoulli sequence, then

$$b_j^1 = \begin{cases} 1 & a_j = 1 \\ 0 & a_j = -1 \end{cases} \quad \text{and} \quad b_j^2 = \begin{cases} 1 & a_j = -1 \\ 0 & a_j = 1 \end{cases}$$

we have $\langle z,\, a\rangle = \langle z,\, b^1\rangle - \langle z,\, b^2\rangle$. So, $2m$ measurements is equivalent to taking m Bernoulli measurements.

Outline

1. Minimal number of measurements
2. Conditions for uniform recovery of sparse vectors via ℓ^1 minimisation
3. Recovery with incoherent bases
   - Theoretical results - Non-uniform recovery

Outline

1. Minimal number of measurements
2. Conditions for uniform recovery of sparse vectors via ℓ^1 minimisation
3. Recovery with incoherent bases
   - Theoretical results - Non-uniform recovery

Lower bound on sampling complexity

Task 1: Find $A \in \mathbb{C}^{m\times N}$ and recovery maps $\Delta : \mathbb{C}^m \to \mathbb{C}^N$ such that $\Delta(Ax) = x$ for all $x \in \mathbb{C}^N$ s-sparse.

In general, we need $m \geqslant 2s$*.*

Lower bound on sampling complexity

Task 1: Find $A \in \mathbb{C}^{m \times N}$ and recovery maps $\Delta : \mathbb{C}^m \to \mathbb{C}^N$ such that $\Delta(Ax) = x$ for all $x \in \mathbb{C}^N$ s-sparse.

In general, we need $m \geqslant 2s$.

Task 2: Find $A \in \mathbb{C}^{m \times N}$ and recovery maps $\Delta : \mathbb{C}^m \to \mathbb{C}^N$ such that

$$\|x - \Delta(Ax)\|_2 \leqslant \frac{C}{\sqrt{s}} \sigma_s(x)_1, \qquad \forall x \in \mathbb{C}^N.$$

In order for (A, Δ) to be stable of order s, we need $m \geqslant Cs \ln(eN/s)$.

Gelfand widths

Given $K \subset X$ where X is a normed space, the Gelfand m-width are:

$$d^m(K, X) \overset{\text{def.}}{=} \inf \left\{ \sup_{x \in K \cap L^m} \|x\| \ ; \ L^m \subset X, \quad \operatorname{codim}(L^m) \leqslant m \right\}$$

Gelfand widths

Given $K \subset X$ where X is a normed space, the Gelfand m-width are:

$$d^m(K,X) \overset{\text{def.}}{=} \inf \left\{ \sup_{x \in K \cap \mathcal{N}(A)} \|x\| \ ; \ A : X \to \mathbb{R}^m \text{ linear} \right\}$$

Measures the extent to which one can determine elements of K from m linear measurements.

[Kashin '77, Garnaev & Gluskin '84] proved

$$d^m(B_1^N, \ell_2^N) \asymp \min\left(1, \sqrt{\frac{\ln(eN/m)}{m}}\right).$$

where B_1^N is the ℓ^1 ball and ℓ_2^N is the N-dimensional vector space with norm $\|\cdot\|_2$.

Gelfand widths

Given $K \subset X$ where X is a normed space, the Gelfand m-width are:

$$d^m(K,X) \overset{\text{def.}}{=} \inf \left\{ \sup_{x \in K \cap \mathcal{N}(A)} \|x\| \ ; \ A : X \to \mathbb{R}^m \text{ linear} \right\}$$

Measures the extent to which one can determine elements of K from m linear measurements.

[Kashin '77, Garnaev & Gluskin '84] proved

$$d^m(B_1^N, \ell_2^N) \asymp \min\left(1, \sqrt{\frac{\ln(eN/m)}{m}}\right).$$

where B_1^N is the ℓ^1 ball and ℓ_2^N is the N-dimensional vector space with norm $\|\cdot\|_2$.

Consequence of $d^m(B_1^N, \ell_2^N) \gtrsim \sqrt{\frac{\ln(eN/m)}{m}}$ is $m \gtrsim s \ln(eN/s)$.

If (A, Δ) is stable of order s, then for $v \in \mathcal{N}(A) \cap B_1^N$, stable recovery of v_S and v_{S^c} respectively means:

$$\|-v_S - \Delta(A(-v_S))\| \leqslant 0 \implies -v_S = \Delta(A(-v_S)) = \Delta(Av_{S^c})$$

$$\|v_{S^c} - \Delta(Av_{S^c})\| \leqslant \frac{C}{\sqrt{s}} \sigma_s(v_{S^c})_1 \leqslant \frac{C}{\sqrt{s}} \|v\|_1 \implies \|v\|_2 \leqslant \frac{C}{\sqrt{s}}$$

So, we have $d^m(B_1^N, \ell_2^n) \leqslant C/\sqrt{s}$ which implies that $m \gtrsim s \ln(eN/m)$.

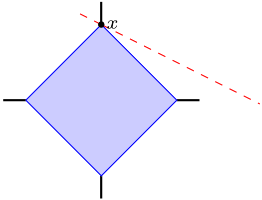
Outline

Null space property

Note that x uniquely minimises

$$\min_z \|z\|_1 \text{ subject to } Az = Ax$$

if and only if $\mathcal{F}_x \cap \mathcal{B}_x = \{x\}$ where $\mathcal{F}_x \stackrel{\text{def.}}{=} \{z \,;\, Az = Ax\}$ and $\mathcal{B}_x \stackrel{\text{def.}}{=} \{z \,;\, \|z\|_1 \leqslant \|x\|_1\}$



Null space property

$A \in \mathbb{C}^{m\times N}$ is said to satisfy the NSP relative to a set $S \subset [N]$ if

$$\|v_S\|_1 < \|v_{S^c}\|_1, \qquad \forall v \in \mathcal{N}(A) \setminus \{0\}$$

It is said to satisfy the NSP of order s if this holds for all $S \subset [N]$ with $|S| \leqslant s$.

Null space property

Theorem

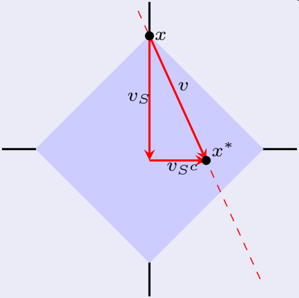
Given $A \in \mathbb{C}^{m \times N}$, *every* $x \in \mathbb{C}^N$ *supported on* $S \subset [N]$ *is the unique solution to (BP) if and only if* A *satisfies the NSP relative to set* S.

Null space property

Theorem

Given $A \in \mathbb{C}^{m \times N}$, every $x \in \mathbb{C}^N$ supported on $S \subset [N]$ is the unique solution to (BP) if and only if A satisfies the NSP relative to set S.

Spe. $Ax^* = Ax$, $x^* \neq x$, $\|x^*\|_1 \leqslant \|x\|_1$



$v \overset{\text{def.}}{=} x^* - x \in \mathcal{N}(A) \setminus \{0\}$ satisfies:

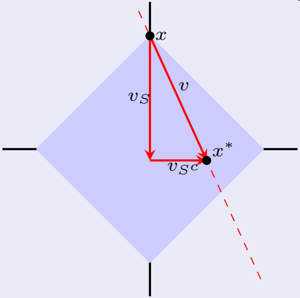
$$\begin{aligned}
&\|v_{S^c}\|_1 = \left\|x^*_{S^c}\right\|_1 \\
&= \left\|x^*_{S^c}\right\|_1 - \|x\|_1 + \left\|x - x^*_S + x^*_S\right\|_1 \\
&\leqslant \left\|x^*_{S^c}\right\|_1 - \|x\|_1 + \|v_S\|_1 + \left\|x^*_S\right\|_1 \\
&\leqslant \|v_S\|_1
\end{aligned}$$

Null space property

Theorem

Given $A \in \mathbb{C}^{m\times N}$, every $x \in \mathbb{C}^N$ supported on $S \subset [N]$ is the unique solution to (BP) if and only if A satisfies the NSP relative to set S.

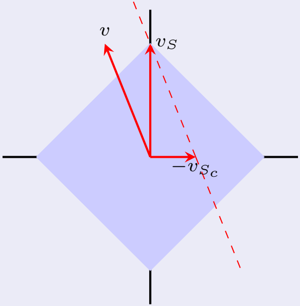
Spe. $Ax^* = Ax$, $x^* \neq x$, $\|x^*\|_1 \leqslant \|x\|_1$



$v \stackrel{\text{def.}}{=} x^* - x \in \mathcal{N}(A) \setminus \{0\}$ satisfies:

$$\begin{aligned}
&\|v_{S^c}\|_1 = \left\|x^*_{S^c}\right\|_1 \\
&= \left\|x^*_{S^c}\right\|_1 - \|x\|_1 + \left\|x - x^*_S + x^*_S\right\|_1 \\
&\leqslant \left\|x^*_{S^c}\right\|_1 - \|x\|_1 + \|v_S\|_1 + \left\|x^*_S\right\|_1 \\
&\leqslant \|v_S\|_1
\end{aligned}$$

Spe. $\exists v \in \mathcal{N}(A) \setminus \{0\}$ s.t. $\|v_S\|_1 \geqslant \|v_{S^c}\|_1$



Let $x \stackrel{\text{def.}}{=} v_S$. Then, $Av_S = -Av_{S^c}$ but x is not the unique solution to (BP).

$\mathcal{F}_x = \{z\ ;\ Az = Ax\}$ is the dotted red line.

Robust and stable recovery

Let $y = Ax + e$ with $\|e\| \leqslant \eta$. What conditions should we impose on A such that

$$\Delta_{BP}^{\eta}(y) \stackrel{\text{def.}}{=} \operatorname{argmin} \|z\|_1 \;\text{ subject to }\; \|Az - y\|_2 \leqslant \eta.$$

satisfies $\left\|x - \Delta_{BP}^{\eta}(y)\right\|_2 \leqslant \frac{C}{\sqrt{s}}\sigma_s(x)_1 + D\eta$ for some $C, D > 0$?

Robust and stable recovery

Let $y = Ax + e$ with $\|e\| \leqslant \eta$. What conditions should we impose on A such that

$$\Delta_{BP}^{\eta}(y) \stackrel{\text{def.}}{=} \operatorname{argmin} \|z\|_1 \ \text{ subject to } \ \|Az - y\|_2 \leqslant \eta.$$

satisfies $\left\|x - \Delta_{BP}^{\eta}(y)\right\|_2 \leqslant \frac{C}{\sqrt{s}} \sigma_s(x)_1 + D\eta$ for some $C, D > 0$?

Robust null space property

We say that A satisfies the robust NSP with constant $\rho, \tau > 0$ if

$$\|v_S\|_2 \leqslant \frac{\rho}{\sqrt{s}} \|v_{S^c}\|_1 + \tau \|Av\|_2 \,, \qquad \forall v \in \mathbb{C}^N.$$

Robust and stable recovery

Let $y = Ax + e$ with $\|e\| \leqslant \eta$. What conditions should we impose on A such that

$$\Delta_{BP}^{\eta}(y) \stackrel{\text{def.}}{=} \operatorname{argmin} \|z\|_1 \ \text{ subject to } \ \|Az - y\|_2 \leqslant \eta.$$

satisfies $\left\|x - \Delta_{BP}^{\eta}(y)\right\|_2 \leqslant \frac{C}{\sqrt{s}}\sigma_s(x)_1 + D\eta$ for some $C, D > 0$?

Robust null space property

We say that A satisfies the robust NSP with constant $\rho, \tau > 0$ if

$$\|v_S\|_2 \leqslant \frac{\rho}{\sqrt{s}} \|v_{S^c}\|_1 + \tau \|Av\|_2 \,, \qquad \forall v \in \mathbb{C}^N.$$

- robust NSP with $\rho < 1$ implies the NSP.
- If A satisfies the robust NSP with $\rho < 1$, then this is sufficient for robust and stable recovery.
- If we have stable and robust recovery, then setting $x \stackrel{\text{def.}}{=} v \in \mathbb{C}^N$, $e = -Av$ and $\eta = \|Av\|_2$, we have $\Delta_{BP}^{\eta}(Ax + e) = 0$ and $\|v\|_2 \leqslant \frac{C}{\sqrt{s}}\sigma_s(v)_1 + D\|Av\|_2$. So, this condition is necessary.

The restricted isometry property

This is one way to assess the quality of the matrix A for recovering s-sparse vectors.

The RIP

The sth restricted isometry constant δ_s of a matrix A is the smallest $\delta > 0$ such that

$$(1-\delta)\left\|x\right\|_2^2 \leqslant \left\|Ax\right\|_2^2 \leqslant (1+\delta)\left\|x\right\|_2^2,$$

for all s-sparse vectors $x \in \mathbb{C}^N$.

- $\delta_s = \max_{|S| \leqslant s} \left\|A_S^* A_S - \mathrm{Id}\right\|$.
- All singular values of A_S are restricted to $[1-\delta_s, 1+\delta_s]$.

The restricted isometry property

This is one way to assess the quality of the matrix A for recovering s-sparse vectors.

The RIP

The sth restricted isometry constant δ_s of a matrix A is the smallest $\delta > 0$ such that

$$(1-\delta)\|x\|_2^2 \leqslant \|Ax\|_2^2 \leqslant (1+\delta)\|x\|_2^2,$$

for all s-sparse vectors $x \in \mathbb{C}^N$.

- $\delta_s = \max_{|S| \leqslant s} \left\| A_S^* A_S - \mathrm{Id} \right\|$.
- All singular values of A_S are restricted to $[1-\delta_s, 1+\delta_s]$.

Theorem (RIP $\implies$ robust NSP $\implies$ robust and stable recovery)

If $\delta_{2s} < \frac{1}{\sqrt{2}}$, then A satisfies the robust NSP of order s with $\rho \in (0,1)$ and $\tau > 0$ dependent only on δ_{2s}. So, the RIP implies that $\left\| x - \Delta_{BP}^{\eta}(Ax+e) \right\|_2 \leqslant \frac{C}{\sqrt{s}} \sigma_s(x)_1 + D\eta$ for some $C, D > 0$ which depend only on δ_{2s}.

Random Gaussian matrices [Mendelson et al '09, Baraniuk et al '08]

Theorem

Let $A \in \mathbb{R}^{m \times N}$ with entries as iid $\mathcal{N}(0,1)$. Let $\tilde{A} = \frac{1}{\sqrt{m}} A$. Then, provided that $m \geqslant C\delta^{-2} s \ln(eN/s)$, wp $\geqslant 1 - 2\exp\left(-\frac{m\delta^2}{128}\right)$, $\tilde{A}$ has RIP constant $\delta_s \leqslant \delta$.

Random Gaussian matrices [Mendelson et al '09, Baraniuk et al '08]

Theorem

Let $A \in \mathbb{R}^{m\times N}$ *with entries as iid* $\mathcal{N}(0,1)$*. Let* $\tilde{A} = \frac{1}{\sqrt{m}} A$*. Then, provided that* $m \geqslant C\delta^{-2} s \ln(eN/s)$*, wp* $\geqslant 1 - 2\exp\left(-\frac{m\delta^2}{128}\right)$*,* $\tilde{A}$ *has RIP constant* $\delta_s \leqslant \delta$*.*

Step 1, Concentration inequality: For fixed $x \in \mathbb{R}^N$ and $t > 0$,

$$\mathbb{P}\left(\left|\left\|\tilde{A}x\right\|_2^2 - \|x\|_2^2\right| > t\,\|x\|_2^2\right) \leqslant 2\exp\left(-\frac{mt^2}{16}\right).$$

- Note that $(\tilde{A}x)_i = \frac{1}{\sqrt{m}} \sum_{j=1}^N A_{ij} x_j = \frac{\|x\|_2}{\sqrt{m}} g_i$ where $g_i = \mathcal{N}(0,1)$.
- $\mathbb{P}\left(\left\|\tilde{A}x\right\|^2 > (1+t)\right) = \mathbb{P}\left(\frac{1}{m}\sum_i g_i^2 > (1+t)\right) = \mathbb{P}\left(\exp\left(u\sum_i g_i^2\right) > \exp\left(um(1+t)\right)\right)$
- By Markov's inequality *, this is upper bounded by

$$\frac{\mathbb{E}\left(\exp\left(u\sum_i g_i^2\right)\right)}{\exp\left(um(1+t)\right)} \underbrace{=}_{\text{indep.}} \prod_{i=1}^m \frac{\mathbb{E}\left(\exp\left(ug_i^2\right)\right)}{\exp\left(u(1+t)\right)} \underbrace{=}_{\text{moment gen. fn.}^\dagger} \left(\frac{1/\sqrt{1-2u}}{\exp(u(1+t))}\right)^m$$

- Choosing $u = t/8 < 1/4$, this is exponentially decaying in t, in particular, upper bounded by $\exp(-mt^2/16)$.

$^*\mathbb{P}(|X| \geqslant t) \leqslant \mathbb{E}\,|X|\,/t$

†For $a < 1/2$, $\mathbb{E}[\exp(ag^2)] = \frac{1}{\sqrt{1-2a}}$

Random Gaussian matrices [Mendelson et al '09, Baraniuk et al '08]

Theorem

Let $A \in \mathbb{R}^{m\times N}$ with entries as iid $\mathcal{N}(0,1)$. Let $\tilde{A} = \frac{1}{\sqrt{m}}A$. Then, provided that $m \geqslant C\delta^{-2} s \ln(eN/s)$, wp $\geqslant 1 - 2\exp\left(-\frac{m\delta^2}{128}\right)$, $\tilde{A}$ has RIP constant $\delta_s \leqslant \delta$.

Step 2: Fix $S \subset [N]$ with $|S| = s$. Then $\left\| \tilde{A}_S^* \tilde{A}_S - \mathrm{Id} \right\| \leqslant \delta$ whp.

- The unit sphere of $\mathbb{R}^s$ can be covered by $n \leqslant (1+2/\rho)^s$ balls of radius ρ.
- Let $\Sigma_S \stackrel{\text{def.}}{=} \left\{ z \in \mathbb{R}^N \;;\; \mathrm{Supp}(z) \subseteq S \right\}$. There exists ℓ_2 normalised $u_1, \ldots, u_n \in \Sigma_S$, $n \leqslant (1+2/\rho)^s$ s.t. for all $x \in \Sigma_S$ with $\|x\| = 1$, there exists k s.t. $\|x - u_k\| \leqslant \rho$.
- Let $B \stackrel{\text{def.}}{=} \tilde{A}_S^* \tilde{A}_S - \mathrm{Id}$.

$$\mathbb{P}(\exists k \in [n],\ |\langle B u_k, u_k\rangle| > t) = \mathbb{P}\left(\exists k \in [n],\ \left| \left\| \tilde{A} u_k \right\|_2^2 - \|u_k\|_2^2 \right| > t\right)$$

$$\leqslant 2n \exp\left(-\frac{mt^2}{16}\right) \leqslant 2(1+2/\rho)^s \exp\left(-\frac{mt^2}{16}\right) = 2\exp\left(\ln(9)s - \frac{m\delta^2}{64}\right) \stackrel{\text{def.}}{=} \varepsilon.$$

if $\rho = 1/4$ and $t = \delta/2$.

- This means that w.p. $1-\varepsilon$, $\|B\| \leqslant \delta$:

$$|\langle Bx, x\rangle| = |\langle B u_k, u_k\rangle + \langle B(x+u_k), (x-u_k)\rangle| \leqslant \frac{\delta}{2} + \|B\| \, \|x+u_k\| \, \|x-u_k\|$$

$$\leqslant \frac{\delta}{2} + 2\rho \|B\| = \frac{\delta}{2} + \frac{1}{2}\|B\|.$$

Random Gaussian matrices [Mendelson et al '09, Baraniuk et al '08]

Theorem

Let $A \in \mathbb{R}^{m\times N}$ *with entries as iid* $\mathcal{N}(0,1)$*. Let* $\tilde{A} = \frac{1}{\sqrt{m}} A$*. Then, provided that* $m \geqslant C\delta^{-2} s \ln(eN/s)$*, wp* $\geqslant 1 - 2\exp\left(-\frac{m\delta^2}{128}\right)$*,* $\tilde{A}$ *has RIP constant* $\delta_s \leqslant \delta$*.*

Step 3, Union bound: There are $\binom{N}{s} \leqslant (eN/s)^s$ subsets of size s in $[N]$. Therefore,

$$\mathbb{P}(\delta_s > \delta) = \mathbb{P}\left(\left\| \tilde{A}_S^* \tilde{A}_S - \mathrm{Id} \right\| > \delta \text{ for some } S \subset [N], |S| = s\right)$$

$$\leqslant 2(eN/s)^s \exp\left(\ln(9)s - \frac{m\delta^2}{64}\right) \leqslant 2\exp\left(-\frac{m\delta^2}{128}\right)$$

provided that $\ln(9e)s\ln(eN/s) \leqslant m\delta^2/128$, i.e. $m \geqslant C\delta^{-2} s\ln(eN/s)$.

Random Gaussian matrices [Mendelson et al '09, Baraniuk et al '08]

Theorem

Let $A \in \mathbb{R}^{m\times N}$ *with entries as iid* $\mathcal{N}(0,1)$. *Let* $\tilde{A} = \frac{1}{\sqrt{m}}A$. *Then, provided that* $m \geqslant C\delta^{-2} s \ln(eN/s)$, *wp* $\geqslant 1 - 2\exp\left(-\frac{m\delta^2}{128}\right)$, $\tilde{A}$ *has RIP constant* $\delta_s \leqslant \delta$.

Step 3, Union bound: There are $\binom{N}{s} \leqslant (eN/s)^s$ subsets of size s in $[N]$. Therefore,

$$\begin{aligned}\mathbb{P}(\delta_s > \delta) &= \mathbb{P}\left(\left\|\tilde{A}_S^*\tilde{A}_S - \mathrm{Id}\right\| > \delta \text{ for some } S \subset [N], |S| = s\right)\\ &\leqslant 2(eN/s)^s \exp\left(\ln(9)s - \frac{m\delta^2}{64}\right) \leqslant 2\exp\left(-\frac{m\delta^2}{128}\right)\end{aligned}$$

provided that $\ln(9e)s\ln(eN/s) \leqslant m\delta^2/128$, i.e. $m \geqslant C\delta^{-2}s\ln(eN/s)$.

Remarks:

- Similar result of random Bernoulli matrices.
- Let $U \in \mathbb{R}^{N\times N}$ be unitary. Then, $\delta_s(\tilde{A}U^*) \leqslant \delta$ with the same probability, since given any $x \in \mathbb{C}^N$, let $x' \stackrel{\text{def.}}{=} U^* x$:

$$\mathbb{P}\left(\left|\left\|\tilde{A}U^*x\right\|_2^2 - \|x\|^2\right| > t\,\|x\|_2^2\right) = \mathbb{P}\left(\left|\left\|\tilde{A}x'\right\|_2^2 - \left\|x'\right\|^2\right| > t\,\left\|x'\right\|_2^2\right) \leqslant 2\exp\left(\frac{-mt^2}{16}\right).$$

Summary

Compressed sensing allows for the recovery of s-sparse vectors $x \in \mathbb{C}^N$ from randomised linear measurements $Ax \in \mathbb{C}^m$ with $m \ll N$ via ℓ^1-minimisation.

- To guarantee the stable recovery of s-sparse signals, we need at least $m = \mathcal{O}(s \log(N/s))$ measurements (for **any** method).
- the NSP is a necessary and sufficient condition for the recovery of s-sparse signals.
- the robust NSP is a sufficient (and almost necessary) condition for the stable and robust recovery of s-sparse signals.
- if a matrix has sufficiently small RIP constant δ_s, then it satisfies the robust-NSP.
- random Gaussian/random Bernoulli matrices satisfy the RIP with $m = \mathcal{O}(s \log(N/s))$.

Outline

1. Minimal number of measurements
2. Conditions for uniform recovery of sparse vectors via ℓ^1 minimisation
3. Recovery with incoherent bases
   - Theoretical results - Non-uniform recovery

Setup

Suppose that $V = [v_1|\cdots|v_N] \in \mathbb{C}^{N\times N}$ and $W = [w_1|\cdots|w_N] \in \mathbb{C}^{N\times N}$ are unitary matrices. Let $z \in \mathbb{C}^N$ be the signal of interest.

- Observe $\langle z, w_j\rangle$ for $j \in \Omega$ where $\Omega \subseteq [N]$ is a randomly chosen set of indices.
- z is s-sparse in V, that is, $z = Vx$ where $x \in \Sigma_s$.

Therefore, we want to recover x from

$$y = P_\Omega U x, \quad \text{where} \quad U = W^* V.$$

Setup

Suppose that $V = [v_1|\cdots|v_N] \in \mathbb{C}^{N\times N}$ and $W = [w_1|\cdots|w_N] \in \mathbb{C}^{N\times N}$ are unitary matrices. Let $z \in \mathbb{C}^N$ be the signal of interest.

- Observe $\langle z, w_j\rangle$ for $j \in \Omega$ where $\Omega \subseteq [N]$ is a randomly chosen set of indices.
- z is s-sparse in V, that is, $z = Vx$ where $x \in \Sigma_s$.

Therefore, we want to recover x from

$$y = P_\Omega U x, \quad \text{where} \quad U = W^* V.$$

Definition

*The **coherence** of V and W is $\mu \stackrel{\text{def.}}{=} \max_{k,\ell} |\langle v_\ell, w_k\rangle|$. In the following, let $K \stackrel{\text{def.}}{=} \sqrt{N}\mu$.*

Setup

Suppose that $V = [v_1|\cdots|v_N] \in \mathbb{C}^{N\times N}$ and $W = [w_1|\cdots|w_N] \in \mathbb{C}^{N\times N}$ are unitary matrices. Let $z \in \mathbb{C}^N$ be the signal of interest.

- Observe $\langle z, w_j\rangle$ for $j \in \Omega$ where $\Omega \subseteq [N]$ is a randomly chosen set of indices.
- z is s-sparse in V, that is, $z = Vx$ where $x \in \Sigma_s$.

Therefore, we want to recover x from

$$y = P_\Omega U x, \quad \text{where} \quad U = W^* V.$$

Definition

*The **coherence** of V and W is $\mu \stackrel{\text{def.}}{=} \max_{k,\ell} |\langle v_\ell, w_k\rangle|$. In the following, let $K \stackrel{\text{def.}}{=} \sqrt{N}\mu$.*

Clearly, $\mu \leqslant 1$, and since W and V are unitary, we have

$$1 = \|w_k\|^2 = \sum_{\ell=1}^{N} |\langle w_k, v_\ell\rangle|^2 \leqslant N\mu^2$$

so $\mu \geqslant \frac{1}{\sqrt{N}}$. When $\mu = \frac{1}{\sqrt{N}}$, we say that V and W are maximally incoherent. So,

$$\mu \in [\frac{1}{\sqrt{N}}, 1] \quad \text{and} \quad K \in [1, \sqrt{N}]$$

Examples

- The Fourier transform $W = \frac{1}{\sqrt{N}} \left(e^{i2\pi(\ell-1)(k-1)/N}\right)_{k,\ell=1}^{N}$ is maximally incoherent with the canonical basis $V = \mathrm{Id}_N$, with $\mu = \frac{1}{\sqrt{N}}$.

Examples

- The Fourier transform $W = \frac{1}{\sqrt{N}} \left(e^{i2\pi(\ell-1)(k-1)/N}\right)_{k,\ell=1}^{N}$ is maximally incoherent with the canonical basis $V = \mathrm{Id}_N$, with $\mu = \frac{1}{\sqrt{N}}$.
- The Hadamard transform is maximally incoherence with the canonical basis, where the Hadamard transform is $H \stackrel{\text{def.}}{=} H_n \in \mathbb{R}^{2^n \times 2^n}$ is defined recursively by

$$H_n = \frac{1}{\sqrt{2}} \begin{pmatrix} H_{n-1} & H_{n-1} \\ H_{n-1} & -H_{n-1} \end{pmatrix}, \qquad H_0 = 1.$$

  It can be computed in $\mathcal{O}(N \log(N))$ time and is useful in modelling systems where there are 'on/off' measurements, such as the single-pixel camera, or Fluorescence microscopy.

Examples

- The Fourier transform $W = \frac{1}{\sqrt{N}} \left(e^{i2\pi(\ell-1)(k-1)/N}\right)_{k,\ell=1}^{N}$ is maximally incoherent with the canonical basis $V = \mathrm{Id}_N$, with $\mu = \frac{1}{\sqrt{N}}$.
- The Hadamard transform is maximally incoherence with the canonical basis, where the Hadamard transform is $H \stackrel{\text{def.}}{=} H_n \in \mathbb{R}^{2^n \times 2^n}$ is defined recursively by

$$H_n = \frac{1}{\sqrt{2}} \begin{pmatrix} H_{n-1} & H_{n-1} \\ H_{n-1} & -H_{n-1} \end{pmatrix}, \qquad H_0 = 1.$$

  It can be computed in $\mathcal{O}(N \log(N))$ time and is useful in modelling systems where there are 'on/off' measurements, such as the single-pixel camera, or Fluorescence microscopy.
- The Haar wavelet basis is maximally incoherent with the noiselet basis.

Examples

- The Fourier transform $W = \frac{1}{\sqrt{N}} \left(e^{i2\pi(\ell-1)(k-1)/N} \right)_{k,\ell=1}^{N}$ is maximally incoherent with the canonical basis $V = \mathrm{Id}_N$, with $\mu = \frac{1}{\sqrt{N}}$.
- The Hadamard transform is maximally incoherence with the canonical basis, where the Hadamard transform is $H \stackrel{\text{def.}}{=} H_n \in \mathbb{R}^{2^n \times 2^n}$ is defined recursively by

$$H_n = \frac{1}{\sqrt{2}} \begin{pmatrix} H_{n-1} & H_{n-1} \\ H_{n-1} & -H_{n-1} \end{pmatrix}, \qquad H_0 = 1.$$

  It can be computed in $\mathcal{O}(N \log(N))$ time and is useful in modelling systems where there are 'on/off' measurements, such as the single-pixel camera, or Fluorescence microscopy.
- The Haar wavelet basis is maximally incoherent with the noiselet basis.
- Any basis is maximally coherent with itself, as $\mu = 1$.

Examples

- The Fourier transform $W = \frac{1}{\sqrt{N}} \left(e^{i2\pi(\ell-1)(k-1)/N}\right)_{k,\ell=1}^{N}$ is maximally incoherent with the canonical basis $V = \mathrm{Id}_N$, with $\mu = \frac{1}{\sqrt{N}}$.
- The Hadamard transform is maximally incoherence with the canonical basis, where the Hadamard transform is $H \stackrel{\text{def.}}{=} H_n \in \mathbb{R}^{2^n \times 2^n}$ is defined recursively by

$$H_n = \frac{1}{\sqrt{2}} \begin{pmatrix} H_{n-1} & H_{n-1} \\ H_{n-1} & -H_{n-1} \end{pmatrix}, \qquad H_0 = 1.$$

  It can be computed in $\mathcal{O}(N \log(N))$ time and is useful in modelling systems where there are 'on/off' measurements, such as the single-pixel camera, or Fluorescence microscopy.
- The Haar wavelet basis is maximally incoherent with the noiselet basis.
- Any basis is maximally coherent with itself, as $\mu = 1$.

Uniform recovery guarantee

If

$$m \stackrel{\text{def.}}{=} |\Omega| \gtrsim K^2 \delta^{-2} s \ln^4(N),$$

then $\sqrt{\frac{N}{m}} U$ satisfies $\delta_s \leqslant \delta$ with probability at least $1 - N^{-\ln^3(N)}$. This guarantees uniform recovery of all s-sparse vectors.

Examples

- The Fourier transform $W = \frac{1}{\sqrt{N}} \left(e^{i2\pi(\ell-1)(k-1)/N}\right)_{k,\ell=1}^{N}$ is maximally incoherent with the canonical basis $V = \mathrm{Id}_N$, with $\mu = \frac{1}{\sqrt{N}}$.
- The Hadamard transform is maximally incoherence with the canonical basis, where the Hadamard transform is $H \overset{\text{def.}}{=} H_n \in \mathbb{R}^{2^n \times 2^n}$ is defined recursively by

$$H_n = \frac{1}{\sqrt{2}} \begin{pmatrix} H_{n-1} & H_{n-1} \\ H_{n-1} & -H_{n-1} \end{pmatrix}, \qquad H_0 = 1.$$

  It can be computed in $\mathcal{O}(N \log(N))$ time and is useful in modelling systems where there are 'on/off' measurements, such as the single-pixel camera, or Fluorescence microscopy.
- The Haar wavelet basis is maximally incoherent with the noiselet basis.
- Any basis is maximally coherent with itself, as $\mu = 1$.

Uniform recovery guarantee

If

$$m \overset{\text{def.}}{=} |\Omega| \gtrsim K^2 \delta^{-2} s \ln^4(N),$$

then $\sqrt{\frac{N}{m}} U$ satisfies $\delta_s \leqslant \delta$ with probability at least $1 - N^{-\ln^3(N)}$. This guarantees uniform recovery of all s-sparse vectors.

Note that $\log(N)^4$ is not so small... for $N = 1000$, $\log(N) \approx 6.9$ but $\log^4(N) > 2N$!.

Uniform vs nonuniform guarantees

So far, we have seen that NSP, robust NSP, RIP guarantee recovery of all s-sparse vectors. In particular, we have seen the following uniform recovery guarantee:

$$\mathbb{P}\left(\forall x \in \Sigma_s,\ \Delta_{BP}(Ax+e) \text{ recovers } x\right) \geqslant 1-\varepsilon$$

Uniform vs nonuniform guarantees

So far, we have seen that NSP, robust NSP, RIP guarantee recovery of all s-sparse vectors. In particular, we have seen the following uniform recovery guarantee:

$$\mathbb{P}\left(\forall x \in \Sigma_s,\ \Delta_{BP}(Ax+e) \text{ recovers } x\right) \geqslant 1-\varepsilon$$

However, we could ask for a weaker statement: given one vector $x \in \Sigma_s$, show that

$$\mathbb{P}\left(\Delta_{BP}(Ax+e) \text{ recovers } x\right) \geqslant 1-\varepsilon$$

Uniform vs nonuniform guarantees

So far, we have seen that NSP, robust NSP, RIP guarantee recovery of all s-sparse vectors. In particular, we have seen the following uniform recovery guarantee:

$$\mathbb{P}\left(\forall x \in \Sigma_s,\ \Delta_{BP}(Ax+e) \text{ recovers } x\right) \geqslant 1-\varepsilon$$

However, we could ask for a weaker statement: given one vector $x \in \Sigma_s$, show that

$$\mathbb{P}\left(\Delta_{BP}(Ax+e) \text{ recovers } x\right) \geqslant 1-\varepsilon$$

In this section, we shall derive nonuniform recovery statements when $m \gtrsim K^2 s \ln(N)$.

Uniform vs nonuniform guarantees

So far, we have seen that NSP, robust NSP, RIP guarantee recovery of all s-sparse vectors. In particular, we have seen the following uniform recovery guarantee:

$$\mathbb{P}\left(\forall x \in \Sigma_s,\ \Delta_{BP}(Ax+e) \text{ recovers } x\right) \geqslant 1-\varepsilon$$

However, we could ask for a weaker statement: given one vector $x \in \Sigma_s$, show that

$$\mathbb{P}\left(\Delta_{BP}(Ax+e) \text{ recovers } x\right) \geqslant 1-\varepsilon$$

In this section, we shall derive nonuniform recovery statements when $m \gtrsim K^2 s \ln(N)$.

Remark

Recall that stable recovery requires $m \gtrsim s\ln(N/s)$, and random Gaussian matrices achieve this optimal rate.
However, one can show that for subsampled orthonormal systems, if we want

$$\|\Delta_{BP}(x) - x\|_1 \lesssim \sigma_s(x)_1$$

to hold for all vectors x, then necessarily, $m \gtrsim s\ln(N)$.

Non-universal recovery and dual certificates

RIP and NSP are concerned with the recovery of all s-sparse vectors or all vectors supported on some $S \subset [N]$. What if we are only interested in the recovery of one vector x?

Theorem

Given $A \in \mathbb{C}^{m \times N}$, $x \in \mathbb{C}^N$ *with support* S *is the unique minimiser of BP with* $y = Ax$ *if either*

(a) $|\langle \mathrm{sign}(x)_S, v \rangle| < \|v_{S^c}\|_1$ *for all* $v \in \mathcal{N}(A) \setminus \{0\}$,

(b) A_S *is injective and* $\exists h \in \mathbb{C}^m$ *s.t.*

$$(A^* h)_S = \mathrm{sign}(x_S) \quad \text{and} \quad \|(A^* h)_{S^c}\|_\infty < 1$$

Non-universal recovery and dual certificates

RIP and NSP are concerned with the recovery of all s-sparse vectors or all vectors supported on some $S \subset [N]$. What if we are only interested in the recovery of one vector x?

Theorem

Given $A \in \mathbb{C}^{m \times N}$, $x \in \mathbb{C}^N$ with support S is the unique minimiser of BP with $y = Ax$ if either

(a) $|\langle \mathrm{sign}(x)_S, v \rangle| < \|v_{S^c}\|_1$ *for all* $v \in \mathcal{N}(A) \setminus \{0\}$,

(b) A_S *is injective and* $\exists h \in \mathbb{C}^m$ *s.t.*

$$(A^*h)_S = \mathrm{sign}(x_S) \quad \textit{and} \quad \|(A^*h)_{S^c}\|_\infty < 1$$

- (a) and (b) are equivalent.
- The null space property relative to S implies (a).
- A^*h is called a **dual certificate**.
- The converse is also true in the real setting, but false in general.

Dual certificates guarantee robust and stable recovery

Theorem (Dual certificate)

Suppose that [a]

$$\|A_S^* A_S - \mathrm{Id}\| \leqslant \frac{1}{2} \quad \textit{and} \quad \max_{\ell \in S^c} \left\|A_S^* A_{\{\ell\}}\right\|_2 \leqslant 1,$$

and there exists $u = A^* h$ *such that*

$$u_S = \mathrm{sign}(x_S) \quad \textit{and} \quad \|u_{S^c}\|_\infty \leqslant \frac{1}{2} \quad \textit{and} \quad \|h\|_2 \leqslant 2\sqrt{s}.$$

Then any minimizer x^* *to* $\min_z \|z\|_1$ *subject to* $\|Az - y\|_2 \leqslant \eta$ *where* $y = Ax + e$ *with* $\|e\| \leqslant \eta$ *satisfies*

$$\|x - x^*\|_2 \lesssim \sigma_s(x)_1 + \sqrt{s}\eta.$$

[a] For simplicity, I have made constants in the upper bounds explicit here.

Application to our problem

Our aim: recover x from $y = P_\Omega U x + e$, where U is a unitary matrix and $\Omega \overset{\text{def.}}{=} \{k_\ell\}_{\ell=1}^m$ are chosen iid unif. rand.

Application to our problem

Our aim: recover x from $y = P_\Omega U x + e$, where U is a unitary matrix and $\Omega \stackrel{\text{def.}}{=} \{k_\ell\}_{\ell=1}^m$ are chosen iid unif. rand.

Let $A \stackrel{\text{def.}}{=} \sqrt{\frac{N}{m}} P_\Omega U$. Then $\mathbb{E}[A^* A] = \mathrm{Id}$ and $\mathbb{E}[A_S^* A_{S^c}] = 0$:

$$\mathbb{E}[(A^* A)_{i,j}] = \frac{N}{m} \sum_{\ell=1}^{m} \mathbb{E}[U_{i,k_\ell}^* U_{k_\ell,j}] = \frac{N}{m} \sum_{\ell=1}^{m} \sum_{k=1}^{N} \frac{1}{N} U_{i,k}^* U_{k,j} = (U^* U)_{i,j} = \delta_{ij}.$$

Application to our problem

Our aim: recover x from $y = P_\Omega U x + e$, where U is a unitary matrix and $\Omega \stackrel{\text{def.}}{=} \{k_\ell\}_{\ell=1}^m$ are chosen iid unif. rand.

Let $A \stackrel{\text{def.}}{=} \sqrt{\frac{N}{m}} P_\Omega U$. Then $\mathbb{E}[A^* A] = \mathrm{Id}$ and $\mathbb{E}[A_S^* A_{S^c}] = 0$:

$$\mathbb{E}[(A^*A)_{i,j}] = \frac{N}{m}\sum_{\ell=1}^m \mathbb{E}[U^*_{i,k_\ell} U_{k_\ell,j}] = \frac{N}{m}\sum_{\ell=1}^m \sum_{k=1}^N \frac{1}{N} U^*_{i,k} U_{k,j} = (U^*U)_{i,j} = \delta_{ij}.$$

Key question: How large does m need to be such that with probability at least $1-\rho$,

- $\left\| A_S^* A_S - \mathrm{Id} \right\| \leqslant \frac{1}{2}$ and $\max_{\ell \in S^c} \left\| A_S^* A_{\{\ell\}} \right\|_2 \leqslant 1$,
- there exists $u = A^* h$ such that

$$u_S = \mathrm{sign}(x_S) \quad \text{and} \quad \|u_{S^c}\|_\infty \leqslant \frac{1}{2} \quad \text{and} \quad \|h\|_2 \leqslant 2\sqrt{s}.$$

Application to our problem

Our aim: recover x from $y = P_\Omega U x + e$, where U is a unitary matrix and $\Omega \stackrel{\text{def.}}{=} \{k_\ell\}_{\ell=1}^m$ are chosen iid unif. rand.

Let $A \stackrel{\text{def.}}{=} \sqrt{\frac{N}{m}} P_\Omega U$. Then $\mathbb{E}[A^* A] = \mathrm{Id}$ and $\mathbb{E}[A_S^* A_{S^c}] = 0$:

$$\mathbb{E}[(A^*A)_{i,j}] = \frac{N}{m}\sum_{\ell=1}^m \mathbb{E}[U^*_{i,k_\ell} U_{k_\ell,j}] = \frac{N}{m}\sum_{\ell=1}^m \sum_{k=1}^N \frac{1}{N} U^*_{i,k} U_{k,j} = (U^*U)_{i,j} = \delta_{ij}.$$

Key question: How large does m need to be such that with probability at least $1-\rho$,

- $\left\| A_S^* A_S - \mathrm{Id} \right\| \leqslant \frac{1}{2}$ and $\max_{\ell \in S^c} \left\| A_S^* A_{\{\ell\}} \right\|_2 \leqslant 1$,
- there exists $u = A^* h$ such that

$$u_S = \mathrm{sign}(x_S) \quad \text{and} \quad \|u_{S^c}\|_\infty \leqslant \frac{1}{2} \quad \text{and} \quad \|h\|_2 \leqslant 2\sqrt{s}.$$

Let $\eta \geqslant \|e\|$. This would guarantee that any solution $\tilde{x}$ to

$$\min \|z\|_1 \text{ subject to } \|P_\Omega U z - y\| \leqslant \eta$$

satisfies

$$\|\tilde{x} - x\|_2 \lesssim \sigma_s(x)_1 + \sqrt{s}\eta.$$

Existence of dual certificates

A natural candidate of a certificate is the Fuchs certificate:

$$u = A^* A_S (A_S^* A_S)^{-1} \operatorname{sign}(x_S).$$

Note that $u_S = \operatorname{sign}(x_S)$ and we simply need to check that $|u_{S^c}| < 1$. Therefore,

- we simply need to control $A_{S^c}^* A_S$ and $(A_S^* A_S)^{-1}$.
- NB: $\mathbb{E}[A_{S^c}^* A_S] = 0$ and $\mathbb{E}[A_S^* A_S] = \mathrm{Id}$.

Existence of dual certificates

A natural candidate of a certificate is the Fuchs certificate:

$$u = A^* A_S (A_S^* A_S)^{-1} \operatorname{sign}(x_S).$$

Note that $u_S = \operatorname{sign}(x_S)$ and we simply need to check that $|u_{S^c}| < 1$. Therefore,

- we simply need to control $A_{S^c}^* A_S$ and $(A_S^* A_S)^{-1}$.
- NB: $\mathbb{E}[A_{S^c}^* A_S] = 0$ and $\mathbb{E}[A_S^* A_S] = \mathrm{Id}$.

Probabilistic bounds (proved using Bernstein concentration inequalities).
With probability at least $1 - \varepsilon$,

(I) $\left\| A_S^* A_S - \mathrm{Id} \right\| \leqslant \delta$ if $m \gtrsim K^2 \delta^{-2} s \ln(2s\varepsilon^{-1})$.

(II) $\max_{j \in S^c} \left\| A_S^* a_j \right\| \leqslant t$ if $m \gtrsim K^2 \max\left(\ln^2(N\varepsilon^{-1}), st^{-2}\right)$.

(III) $\max_{j \in S^c} \left| \langle \operatorname{sign}(x_S),\, A_S^* a_j \rangle \right| \leqslant r$ if $m \gtrsim K^2 s r^{-2} \ln(N\varepsilon^{-1})$.

Existence of dual certificates

A natural candidate of a certificate is the Fuchs certificate:

$$u = A^* A_S (A_S^* A_S)^{-1} \operatorname{sign}(x_S).$$

Note that $u_S = \operatorname{sign}(x_S)$ and we simply need to check that $|u_{S^c}| < 1$. Therefore,

- we simply need to control $A_{S^c}^* A_S$ and $(A_S^* A_S)^{-1}$.
- NB: $\mathbb{E}[A_{S^c}^* A_S] = 0$ and $\mathbb{E}[A_S^* A_S] = \mathrm{Id}$.

Probabilistic bounds (proved using Bernstein concentration inequalities).
With probability at least $1 - \varepsilon$,

(I) $\left\| A_S^* A_S - \mathrm{Id} \right\| \leqslant \delta$ if $m \gtrsim K^2 \delta^{-2} s \ln(2s\varepsilon^{-1})$.

(II) $\max_{j \in S^c} \left\| A_S^* a_j \right\| \leqslant t$ if $m \gtrsim K^2 \max\left(\ln^2(N\varepsilon^{-1}), st^{-2}\right)$.

(III) $\max_{j \in S^c} \left| \langle \operatorname{sign}(x_S),\, A_S^* a_j \rangle \right| \leqslant r$ if $m \gtrsim K^2 s r^{-2} \ln(N\varepsilon^{-1})$. *

* (III) comes from the stronger result "For a fixed vector v, with probability at least $1 - \delta$, $\max_{j \in S^c} |\langle v,\, A_S^* a_j \rangle| \leqslant \frac{r\|v\|}{\sqrt{s}}$ if $m \gtrsim K^2 s r^{-2} \ln(N\varepsilon^{-1})$"

Naive approaches

To control $u_j = (A^* A_S (A_S^* A_S)^{-1} \operatorname{sign}(x_S))_j$ for $j \notin S$...

Naive approaches

To control $u_j = (A^* A_S (A_S^* A_S)^{-1} \operatorname{sign}(x_S))_j$ for $j \notin S$...

Naive approach:

$$|u_j| = \left|\langle (A_S^* A_S)^{-1} A_S^* a_j, \operatorname{sign}(x_S)\rangle\right| \leqslant \|A_S^* a_j\| \left\|(A_S^* A_S)^{-1}\right\| \sqrt{s} < 1$$

if $\left\|(A_S^* A_S)^{-1}\right\| < 2$ and $\left\|A_S^* a_j\right\| < \frac{1}{2\sqrt{s}}$ for all $j \in S^c$.

This holds with probability at least ε if

$$m \gtrsim K^2 \max\left(s \ln(2s/\varepsilon), \ln^2(N/\varepsilon), s^2\right).$$

Naive approaches

To control $u_j = (A^* A_S(A_S^* A_S)^{-1}\,\mathrm{sign}(x_S))_j$ for $j \notin S$...

Naive approach:

$$|u_j| = \left|\langle (A_S^* A_S)^{-1} A_S^* a_j, \, \mathrm{sign}(x_S)\rangle\right| \leqslant \|A_S^* a_j\| \, \left\|(A_S^* A_S)^{-1}\right\| \sqrt{s} < 1$$

if $\left\|(A_S^* A_S)^{-1}\right\| < 2$ and $\left\|A_S^* a_j\right\| < \frac{1}{2\sqrt{s}}$ for all $j \in S^c$.

This holds with probability at least ε if

$$m \gtrsim K^2 \max\left(s\ln(2s/\varepsilon), \ln^2(N/\varepsilon), s^2\right).$$

Slightly less naive approach:

$$\begin{aligned} |u_j| &= \left|\langle (A_S^* A_S)^{-1} A_S^* a_j, \, \mathrm{sign}(x_S)\rangle\right| \\ &\leqslant \left|\langle \left((A_S^* A_S)^{-1} - \mathrm{Id}\right) A_S^* a_j, \, \mathrm{sign}(x_S)\rangle\right| + |\langle A_S^* a_j, \, \mathrm{sign}(x_S)\rangle| \\ &\leqslant \underbrace{\left\|(A_S^* A_S)^{-1} - \mathrm{Id}\right\|}_{< \frac{1}{\sqrt{2\sqrt{s}}}} \underbrace{\left\|A_S^* a_j\right\|}_{< \frac{1}{\sqrt{2\sqrt{s}}}} \sqrt{s} + \underbrace{|\langle A_S^* a_j, \, \mathrm{sign}(x_S)\rangle|}_{< \frac{1}{2}} < 1 \end{aligned}$$

holds provided that

$$m \gtrsim K^2 \max\left(s^{3/2}\ln(s/\varepsilon), s\ln(N/\varepsilon), \ln^2(N/\varepsilon)\right).$$

Optimal number of samples via random signs [Candès & Romberg '07, Tropp '08]

Lemma (Hoeffding's inequality)

Given $v \in \mathbb{C}^s$, *if* $\alpha \geqslant \|v\|$ *and* σ *is a Rademacher sequence,*

$$\mathbb{P}(\langle v, \sigma \rangle \geqslant w) \leqslant 2 \exp\left(-w^2/(2\alpha^2)\right).$$

Optimal number of samples via random signs [Candès & Romberg '07, Tropp '08]

Lemma (Hoeffding's inequality)

Given $v \in \mathbb{C}^s$, if $\alpha \geqslant \|v\|$ and σ is a Rademacher sequence,

$$\mathbb{P}(\langle v, \sigma \rangle \geqslant w) \leqslant 2 \exp\left(-w^2/(2\alpha^2)\right).$$

We already know that w.p. at least $1 - \varepsilon'$, if $m \gtrsim K^2 \max\left(s \ln(2s/\varepsilon'), \ln^2(N/\varepsilon'), \frac{s}{t^2}\right)$, then $v_j \overset{\text{def.}}{=} (A_S^* A_S)^{-1} A_S^* a_j$ satisfies

$$\|v_j\|^2 \leqslant \left\|(A_S^* A_S)^{-1}\right\|^2 \|A^* a_j\|^2 \leqslant \frac{t^2}{(1-\delta)^2} = 2t^2 \overset{\text{def.}}{=} \alpha^2.$$

Optimal number of samples via random signs [Candès & Romberg '07, Tropp '08]

Lemma (Hoeffding's inequality)

Given $v \in \mathbb{C}^s$, if $\alpha \geqslant \|v\|$ and σ is a Rademacher sequence,

$$\mathbb{P}(\langle v, \sigma \rangle \geqslant w) \leqslant 2 \exp\left(-w^2/(2\alpha^2)\right).$$

We already know that w.p. at least $1 - \varepsilon'$, if $m \gtrsim K^2 \max\left(s \ln(2s/\varepsilon'), \ln^2(N/\varepsilon'), \frac{s}{t^2}\right)$, then $v_j \stackrel{\text{def.}}{=} (A_S^* A_S)^{-1} A_S^* a_j$ satisfies

$$\|v_j\|^2 \leqslant \left\|(A_S^* A_S)^{-1}\right\|^2 \|A^* a_j\|^2 \leqslant \frac{t^2}{(1-\delta)^2} = 2t^2 \stackrel{\text{def.}}{=} \alpha^2.$$

Assume that sign(x_S) is a Rademacher sequence and recall that $|u_j| = |\langle v_j, \operatorname{sign}(x_S)\rangle|$.

$$\begin{aligned}\mathbb{P}(\exists j \in S^c,\ |u_j| > \frac{1}{2}) &\leqslant N\mathbb{P}(|u_j| > \frac{1}{2} \mid \|v_j\| \leqslant \alpha) + \mathbb{P}\left(\exists j \in S^c\ \|v_j\| > \alpha\right) \\ &\leqslant N \exp\left(-1/(16t^2)\right) + \varepsilon' \leqslant \varepsilon,\end{aligned}$$

if $\varepsilon' = \varepsilon/2$ and $t^2 = (16 \ln(2N/\varepsilon))^{-1}$.

Optimal number of samples via random signs [Candès & Romberg '07, Tropp '08]

Lemma (Hoeffding's inequality)

Given $v \in \mathbb{C}^s$, if $\alpha \geqslant \|v\|$ and σ is a Rademacher sequence,

$$\mathbb{P}(\langle v, \sigma \rangle \geqslant w) \leqslant 2 \exp\left(-w^2/(2\alpha^2)\right).$$

We already know that w.p. at least $1 - \varepsilon'$, if $m \gtrsim K^2 \max\left(s \ln(2s/\varepsilon'), \ln^2(N/\varepsilon'), \frac{s}{t^2}\right)$, then $v_j \overset{\text{def.}}{=} (A_S^* A_S)^{-1} A_S^* a_j$ satisfies

$$\|v_j\|^2 \leqslant \left\|(A_S^* A_S)^{-1}\right\|^2 \|A^* a_j\|^2 \leqslant \frac{t^2}{(1-\delta)^2} = 2t^2 \overset{\text{def.}}{=} \alpha^2.$$

Assume that sign(x_S) is a Rademacher sequence and recall that $|u_j| = |\langle v_j, \text{sign}(x_S)\rangle|$.

$$\mathbb{P}(\exists j \in S^c,\ |u_j| > \frac{1}{2}) \leqslant N\mathbb{P}(|u_j| > \frac{1}{2} |\, \|v_j\| \leqslant \alpha) + \mathbb{P}\left(\exists j \in S^c\ \|v_j\| > \alpha\right)$$
$$\leqslant N \exp\left(-1/(16t^2)\right) + \varepsilon' \leqslant \varepsilon,$$

if $\varepsilon' = \varepsilon/2$ and $t^2 = (16 \ln(2N/\varepsilon))^{-1}$.

i.e. $\mathbb{P}(\exists j \in S^c,\ |u_j| > 1) \leqslant \varepsilon$ provided that

$$m \gtrsim K^2 \max\left(s \ln(N/\varepsilon), \ln^2(N/\varepsilon)\right).$$

Recovery statement

Let U be an unitary matrix, $\mu \stackrel{\text{def.}}{=} \max_{k,j} \left|U_{k,j}\right|$ and $K \stackrel{\text{def.}}{=} \sqrt{N}\mu$. We want to recover $x \in \mathbb{C}^N$ from $y = P_\Omega U x + e$ where Ω consists of m indices chosen uniformly at random.

We have so far shown:

Theorem

Suppose that $\operatorname{sign}(x)$ *is a Rademacher sequence and* $m \gtrsim K^2 \max\left(s \ln(N/\varepsilon), \ln^2(N/\varepsilon)\right)$. *Let* $\eta \geqslant \|e\|$. *Then, with probability at least* $1-\varepsilon$, *any solution* $\tilde{x}$ *to*

$$\min \|z\|_1 \;\; \textit{subject to} \;\; \|P_\Omega U z - y\| \leqslant \eta$$

satisfies

$$\|\tilde{x} - x\|_2 \lesssim \sigma_s(x)_1 + \sqrt{s}\eta.$$

Recovery statement

Let U be an unitary matrix, $\mu \stackrel{\text{def.}}{=} \max_{k,j} \left|U_{k,j}\right|$ and $K \stackrel{\text{def.}}{=} \sqrt{N}\mu$. We want to recover $x \in \mathbb{C}^N$ from $y = P_\Omega U x + e$ where Ω consists of m indices chosen uniformly at random.

We have so far shown:

Theorem

Suppose that $\operatorname{sign}(x)$ *is a Rademacher sequence and* $m \gtrsim K^2 \max\left(s\ln(N/\varepsilon), \ln^2(N/\varepsilon)\right)$*.* *Let* $\eta \geqslant \|e\|$*. Then, with probability at least* $1-\varepsilon$*, any solution* $\tilde{x}$ *to*

$$\min \|z\|_1 \ \textit{subject to} \ \|P_\Omega U z - y\| \leqslant \eta$$

satisfies

$$\|\tilde{x} - x\|_2 \lesssim \sigma_s(x)_1 + \sqrt{s}\eta.$$

The assumptions in red can be replaced by

$$m \gtrsim K^2 s \ln(N) \ln(\varepsilon^{-1}).$$

using the idea of **inexact dual certificates** and a **golfing scheme** dual certificate construction (which constructs a different certificate to the Fuchs certificate).

Optimal sampling complexity without the random signs assumption

Recall that

Theorem (Dual certificate)

Suppose that [a]

$$\|A_S^* A_S - \mathrm{Id}\| \leqslant \frac{1}{2} \quad \text{and} \quad \max_{\ell \in S^c} \left\|A_S^* A_{\{\ell\}}\right\|_2 \leqslant 1,$$

and there exists $u = A^* h$ *such that*

$$u_S = \mathrm{sign}(x_S) \quad \text{and} \quad \|u_{S^c}\|_\infty \leqslant \frac{1}{2} \quad \text{and} \quad \|h\|_2 \leqslant 2\sqrt{s}.$$

Then any minimizer x^* *to* $\|z\|_1$ *subject to* $\|Az - y\|_2 \leqslant \eta$ *where* $y = Ax + e$ *with* $\|e\| \leqslant \eta$ *satisfies*

$$\|x - x^*\|_2 \lesssim \sigma_s(x)_1 + \sqrt{s}\eta.$$

[a] For simplicity, I have made constants in the upper bounds explicit here.

Optimal sampling complexity without the random signs assumption

Theorem (Inexact Dual certificate)

Suppose that [a]

$$\|A_S^* A_S - \mathrm{Id}\| \leqslant \frac{1}{2} \quad \textit{and} \quad \max_{\ell \in S^c} \left\|A_S^* A_{\{\ell\}}\right\|_2 \leqslant 1,$$

and there exists $u = A^* h$ *such that*

$$\|u_S - \mathrm{sign}(x_S)\| \leqslant \frac{1}{8} \quad \textit{and} \quad \|u_{S^c}\|_\infty \leqslant \frac{1}{4} \quad \textit{and} \quad \|h\|_2 \leqslant 2\sqrt{s}.$$

Then any minimizer x^* *to* $\|z\|_1$ *subject to* $\|Az - y\|_2 \leqslant \eta$ *where* $y = Ax + e$ *with* $\|e\| \leqslant \eta$ *satisfies*

$$\|x - x^*\|_2 \lesssim \sigma_s(x)_1 + \sqrt{s}\eta.$$

[a] For simplicity, I have made constants in the upper bounds explicit here.

Proof: Inexact dual certificate implies dual certificate.

Let $v \stackrel{\text{def.}}{=} u + \tilde{u}$ where $\tilde{u} \stackrel{\text{def.}}{=} A^* A_S (A_S^* A_S)^{-1} w$ and $w = \mathrm{sign}(x_S) - u_S$.
Note that

$$\|\tilde{u}_{S^c}\|_\infty \leqslant \|A_{S^c}^* A_S\|_{2\to\infty} \left\|(A_S^* A_S)^{-1}\right\|_{2\to 2} \|w\|_2 \leqslant \frac{1}{4}.$$

Therefore, $v_S = u_S + w_S = \mathrm{sign}(x_S)$ and $\|v_{S^c}\|_\infty \leqslant \|u_{S^c}\|_\infty + \|\tilde{u}_{S^c}\|_\infty \leqslant \frac{1}{2}$. □

Golfing Scheme [Gross '11, Candès & Plan '11]

The golfing scheme shows that with probability at least $1-\varepsilon$, there exists an inexact dual certificate when $m \gtrsim K^2 s \log(N/\varepsilon)$.

First observe that the Fuchs precertificate is

$$u = A^* A_S (A_S^* A_S)^{-1} \operatorname{sign}(x_S) = \sum_{n=1}^{\infty} A^* A_S \left(\mathrm{Id} - A_S^* A_S\right)^{n-1} \operatorname{sign}(x_S)$$

$$= \sum_{n=1}^{\infty} A^* A_S w_{n-1}, \quad \text{where} \quad w_n \stackrel{\text{def.}}{=} (\mathrm{Id} - A_S^* A_S) w_{n-1}, \; w_0 = \operatorname{sign}(x_S).$$

Golfing Scheme [Gross '11, Candès & Plan '11]

The golfing scheme shows that with probability at least $1-\varepsilon$, there exists an inexact dual certificate when $m \gtrsim K^2 s \log(N/\varepsilon)$.

First observe that the Fuchs precertificate is

$$u = A^* A_S (A_S^* A_S)^{-1} \operatorname{sign}(x_S) = \sum_{n=1}^{\infty} A^* A_S \left(\mathrm{Id} - A_S^* A_S\right)^{n-1} \operatorname{sign}(x_S)$$

$$= \sum_{n=1}^{\infty} A^* A_S w_{n-1}, \quad \text{where} \quad w_n \stackrel{\text{def.}}{=} (\mathrm{Id} - A_S^* A_S) w_{n-1}, \; w_0 = \operatorname{sign}(x_S).$$

Recall that $A = \sqrt{\frac{N}{m}} P_\Omega U$, where $\Omega = \{k_\ell\}_{\ell=1}^m$. Partition into L subsets $\Omega = \bigcup_{\ell=1}^L \Omega_\ell$, where Ω_ℓ consists of m_ℓ indices. Define $A^{(\ell)} \stackrel{\text{def.}}{=} \sqrt{\frac{N}{m_\ell}} P_{\Omega_\ell} U$.

Golfing Scheme [Gross '11, Candès & Plan '11]

The golfing scheme shows that with probability at least $1 - \varepsilon$, there exists an inexact dual certificate when $m \gtrsim K^2 s \log(N/\varepsilon)$.

First observe that the Fuchs precertificate is

$$u = A^* A_S (A_S^* A_S)^{-1} \operatorname{sign}(x_S) = \sum_{n=1}^{\infty} A^* A_S \left(\mathrm{Id} - A_S^* A_S\right)^{n-1} \operatorname{sign}(x_S)$$

$$= \sum_{n=1}^{\infty} A^* A_S w_{n-1}, \quad \text{where} \quad w_n \stackrel{\text{def.}}{=} (\mathrm{Id} - A_S^* A_S) w_{n-1},\ w_0 = \operatorname{sign}(x_S).$$

Recall that $A = \sqrt{\frac{N}{m}} P_\Omega U$, where $\Omega = \{k_\ell\}_{\ell=1}^m$. Partition into L subsets $\Omega = \bigcup_{\ell=1}^L \Omega_\ell$, where Ω_ℓ consists of m_ℓ indices. Define $A^{(\ell)} \stackrel{\text{def.}}{=} \sqrt{\frac{N}{m_\ell}} P_{\Omega_\ell} U$.

Consider the function

$$\tilde{u}^{(L)} \stackrel{\text{def.}}{=} \sum_{\ell=1}^{L} (A^{(\ell)})^* A_S^{(\ell)} \tilde{w}_{\ell-1}$$

where $\tilde{w}_\ell \stackrel{\text{def.}}{=} (\mathrm{Id} - (A_S^{(\ell)})^* A_S^{(\ell)}) \tilde{w}_{\ell-1},\ \tilde{w}_0 = \operatorname{sign}(x_S)$.

We still have $\tilde{u}^{(L)} \in \operatorname{Im}(A^*)$. The idea is that we have now decoupled the randomness.

Golfing Scheme

We have

$$\tilde{w}^{(\ell)} = \operatorname{sign}(x_S) - \tilde{u}_S^{(\ell)}.$$

If

(I) $\left\| \left(\mathrm{Id} - (A_S^{(\ell)})^* A_S^{(\ell)} \right) \tilde{w}_{\ell-1} \right\|_2 \leqslant r_\ell \, \|\tilde{w}_{\ell-1}\|_2$

(II) $\left\| (A_{S^c}^{(\ell)})^* A_S^{(\ell)} \tilde{w}_{\ell-1} \right\|_\infty \leqslant \frac{t_\ell}{\sqrt{s}} \, \|\tilde{w}_{\ell-1}\|_2,$

then

$$\left\| \operatorname{sign}(x_S) - \tilde{u}_S^{(L)} \right\| \leqslant \left\| \tilde{w}^{(L)} \right\| \leqslant \sqrt{s} \prod_{n=1}^{L} r_n$$

$$\left\| \tilde{u}_{S^c}^{(L)} \right\|_\infty \leqslant \sum_{\ell=1}^{L} \left\| (A_{S^c}^{(\ell)})^* A_S^{(\ell)} \tilde{w}_{\ell-1} \right\|_\infty \leqslant \sum_{\ell=1}^{L} \frac{t_\ell}{\sqrt{s}} \left\| \tilde{w}^{(\ell-1)} \right\|_2 \leqslant \sum_{\ell=1}^{L} t_\ell \prod_{j=1}^{\ell-1} r_j.$$

Golfing Scheme

We have

$$\tilde{w}^{(\ell)} = \mathrm{sign}(x_S) - \tilde{u}_S^{(\ell)}.$$

If

(I) $\left\| \left(\mathrm{Id} - (A_S^{(\ell)})^* A_S^{(\ell)} \right) \tilde{w}_{\ell-1} \right\|_2 \leqslant r_\ell \, \|\tilde{w}_{\ell-1}\|_2$

(II) $\left\| (A_{S^c}^{(\ell)})^* A_S^{(\ell)} \tilde{w}_{\ell-1} \right\|_\infty \leqslant \frac{t_\ell}{\sqrt{s}} \, \|\tilde{w}_{\ell-1}\|_2,$

then

$$\left\| \mathrm{sign}(x_S) - \tilde{u}_S^{(L)} \right\| \leqslant \left\| \tilde{w}^{(L)} \right\| \leqslant \sqrt{s} \prod_{n=1}^{L} r_n$$

$$\left\| \tilde{u}_{S^c}^{(L)} \right\|_\infty \leqslant \sum_{\ell=1}^{L} \left\| (A_{S^c}^{(\ell)})^* A_S^{(\ell)} \tilde{w}_{\ell-1} \right\|_\infty \leqslant \sum_{\ell=1}^{L} \frac{t_\ell}{\sqrt{s}} \left\| \tilde{w}^{(\ell-1)} \right\|_2 \leqslant \sum_{\ell=1}^{L} t_\ell \prod_{j=1}^{\ell-1} r_j.$$

- The idea is that by choosing r_ℓ, t_ℓ and L appropriately, one is guaranteed an inexact dual certificate with probability at least $1 - \varepsilon$ when

$$m = \sum_\ell m_\ell \gtrsim K^2 s \left(\ln(N) \ln(\varepsilon^{-1}) + \ln(s) \ln(s\varepsilon^{-1}) \right).$$

Golfing Scheme

We have

$$\tilde{w}^{(\ell)} = \mathrm{sign}(x_S) - \tilde{u}_S^{(\ell)}.$$

If

(I) $\left\|\left(\mathrm{Id} - (A_S^{(\ell)})^* A_S^{(\ell)}\right) \tilde{w}_{\ell-1}\right\|_2 \leqslant r_\ell \left\|\tilde{w}_{\ell-1}\right\|_2$

(II) $\left\|(A_{S^c}^{(\ell)})^* A_S^{(\ell)} \tilde{w}_{\ell-1}\right\|_\infty \leqslant \frac{t_\ell}{\sqrt{s}} \left\|\tilde{w}_{\ell-1}\right\|_2,$

then

$$\left\|\mathrm{sign}(x_S) - \tilde{u}_S^{(L)}\right\| \leqslant \left\|\tilde{w}^{(L)}\right\| \leqslant \sqrt{s} \prod_{n=1}^{L} r_n$$

$$\left\|\tilde{u}_{S^c}^{(L)}\right\|_\infty \leqslant \sum_{\ell=1}^{L} \left\|(A_{S^c}^{(\ell)})^* A_S^{(\ell)} \tilde{w}_{\ell-1}\right\|_\infty \leqslant \sum_{\ell=1}^{L} \frac{t_\ell}{\sqrt{s}} \left\|\tilde{w}^{(\ell-1)}\right\|_2 \leqslant \sum_{\ell=1}^{L} t_\ell \prod_{j=1}^{\ell-1} r_j.$$

- The idea is that by choosing r_ℓ, t_ℓ and L appropriately, one is guaranteed an inexact dual certificate with probability at least $1 - \varepsilon$ when

$$m = \sum_\ell m_\ell \gtrsim K^2 s \left(\ln(N) \ln(\varepsilon^{-1}) + \ln(s) \ln(s\varepsilon^{-1})\right).$$

- A slightly more refined argument where one is allowed to 'make mistakes' by choosing L slightly larger and throwing away the draws which violate (I) and (II) gives the optimal sampling complexity $m \gtrsim K^2 s \ln(N) \ln(\varepsilon^{-1})$.

Summary

Compressed sensing allows for the recovery of s-sparse vectors $x \in \mathbb{C}^N$ from randomised linear measurements $Ax \in \mathbb{C}^m$ with $m \ll N$ via ℓ^1-minimisation.

- To guarantee the stable recovery of s-sparse signals, we need at least $m = \mathcal{O}(s \log(N/s))$ measurements (for **any** method).
- the NSP is a necessary and sufficient condition for the recovery of s-sparse signals.
- the robust NSP is an almost necessary and sufficient condition for the stable and robust recovery of s-sparse signals.
- if a matrix has sufficiently small RIP constant δ_s, then it satisfies the robust-NSP.
- random Gaussian/random Bernoulli matrices satisfy the RIP with $m = \mathcal{O}(s \log(N/s))$.

We considered the recovery of x from $P_\Omega W^* V x$, with $K = \sqrt{N} \cdot \max_{i,j} |\langle v_j, w_i \rangle|$.

- NSP, robust NSP and RIP are conditions for uniform recovery. They can be hard to establish. For non-uniform recovery results, we look to the construction of dual certificates.
- A dual certificate is an element of $\mathrm{Im}(A^*)$ which interpolates $\mathrm{sign}(x_0)$ exactly.
- The Fuchs certificate $A^* A_S (A_S^* A_S)^{-1} \mathrm{sign}(x_0)$ is a natural candidate for a dual certificate. We can prove that this is indeed a dual certificate provided that $\mathrm{sign}(x_0)$ is a Rademacher sequence when $m = \mathcal{O}(sK^2 \log(N))$.
- The golfing scheme provides another construction of a dual certificate, and allows us to remove the random signs assumption while retaining the optimal sampling complexity.

Sources

- “A Mathematical Introduction to Compressive Sensing” by Simon Foucart & Holger Rauhut.
- “Flavors of Compressive Sensing” by Simon Foucart.