Compressed Sensing (Day 1)

Sadashige Ishida & Raimundo Saona

June 2, 2020

What is compressed sensing?

Wikipedia says

> Compressed sensing (also known as compressive sensing, compressive sampling, or sparse sampling) is a signal processing technique for efficiently acquiring and reconstructing a signal, by finding solutions to underdetermined linear systems.

What is compressed sensing?

Let's break it down. Imagine a real-world situation.

We want to know the information of a signal $f \in \mathbb{C}^N$. Sometimes we do not get f itself, but something different via measurements e.g. coefficients of (discrete) Fourier transform of f. Besides, we often get only the partial information of $\hat{f}$ i.e. only a few Fourier coefficients.

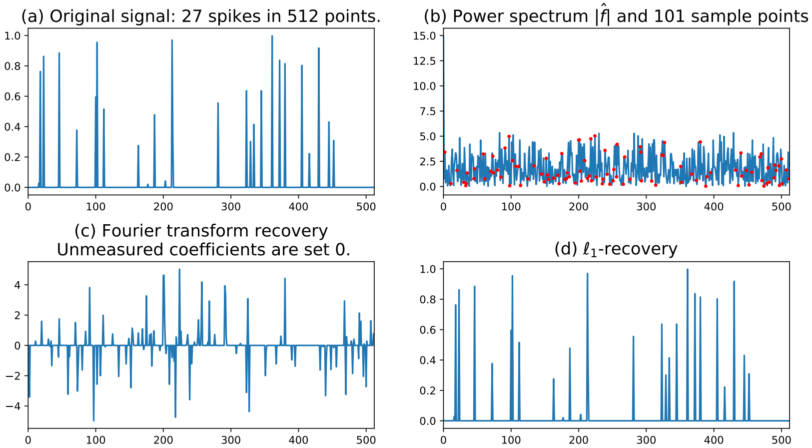
In order to get the original signal f, it seems we need all $\hat{f}$ because we know the relation

$$\mathcal{F} : f \longrightarrow \hat{f} \tag{1.1}$$

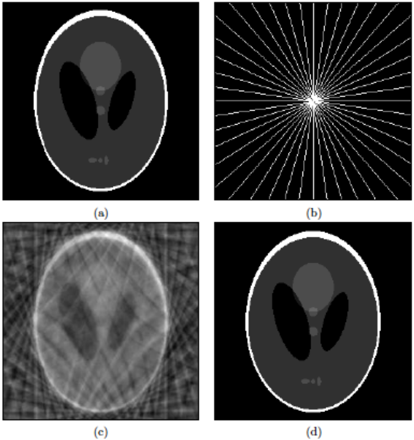
$$\mathcal{F}^{-1} : \hat{f} \longrightarrow f. \tag{1.2}$$

But in some situation, we can retrieve the exact information of f from the partial information of $\hat{f}$.

Example: 1D signal



Example: 2D image [1]



(a) Original image (b) Sample points in Fourier domain
(c) Fourier transform recovery (d) ℓ_1-recovery

Sparse and compressed

Compressed sensing

The exact recovery of a signal from partial measurements is possible if the signal is sparse, in other words, the information is compressed in a small domain in a certain sense.

Sparsity

A signal is sparse if the support of the signal is small.

Recovery condition

The main question of this lecture is

When can we get the exact information from partial measurements?

Discrete Fourier Transform

For a signal $f \in \mathbb{C}^N$, the discrete Fourier transform (DFT) $\mathcal{F} : f \to \hat{f}$ is given by,

$$\hat{f}(\omega) := \sum_{t=0}^{N-1} f(t) e^{-2\pi i t \omega / N}, \tag{1.3}$$

and the inverse $\mathcal{F}^{-1} : \hat{f} \to f$ is

$$f(t) := \frac{1}{N} \sum_{\omega=0}^{N-1} \hat{f}(\omega) e^{2\pi i t \omega / N}. \tag{1.4}$$

Discrete Fourier Transform

The matrix form of DFT is

$$\mathcal{F}_{\omega,t} := s^{t\omega}, \quad s := e^{-2\pi i/N} \tag{1.5}$$

where t, ω runs over $\{0, \ldots, N-1\}$.
For a pair $T, \Omega \subset \{0, \ldots, N-1\}$, we consider a partial DFT $\mathcal{F}_{T\to\Omega}$ by taking Ω rows and T columns.
E.g. let $N = 4$, $T = \{0, 2, 3\}$, $\Omega = \{1, 3\}$. Then

$$\mathcal{F} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & s & s^2 & s^3 \\ 1 & s^2 & s^4 & s^6 \\ 1 & s^3 & s^6 & s^9 \end{pmatrix}, \quad \mathcal{F}_{T\to\Omega} = \begin{pmatrix} 1 & s^2 & s^3 \\ 1 & s^6 & s^9 \end{pmatrix}. \tag{1.6}$$

In the sequel, we may abuse notations, Ω can be a subset of $\{0, \ldots, N-1\}$ or of $\{2\pi k/N : k \in \{0, \ldots, N-1\}\}$.

Recovery condition for the inverse Fourier transform

To recover f from

$$\hat{f}'(\omega) := \begin{cases} \hat{f}(\omega) & \omega \in \Omega \\ 0 & \text{otherwise}, \end{cases} \tag{1.7}$$

we need the measurements Ω to cover all non-zero $\hat{f}$.

Remark

Fourier transform recovery can be seen as a ℓ_2-optimization problem (also called *energy minimization*), as $\mathcal{F}^{-1}\hat{f}'$ is the solution of the problem

$$\min_{g \in \mathbb{C}^N} \|g\|_2 := \sum_t |g(t)|^2 = \frac{1}{N} \sum_\omega |\hat{g}(\omega)|^2, \quad \hat{g}|_\Omega = \hat{f}|_\Omega \tag{1.8}$$

The above equality is by Parseval's theorem (HW exercise).

Recovery condition for other norms

In this lecture, we will study signal recovery via ℓ_0 and ℓ_1 norms.

Remark

At this point, we expect:

$$\text{Easy to optimize } \ell_2 > \ell_1 > \ell_0,$$
$$\text{Recovery is likely correct } \ell_0 > \ell_1 > \ell_2.$$

Deterministic results

Here is a classical result.

Theorem 1.1 (Vetterli, Marziliano, and Blu [2])

Let $f \in \mathbb{C}^N$ be a signal and T be the support of f. Then f can be exactly recovered from $2|T|+1$ consecutive Fourier coefficients.

In fact, we can recover signals in a more relaxed condition.

Deterministic results

The previous theorem requires Ω to have $2|T|+1$ consecutive coefficients. The next theorem is stronger in the way it relaxes the condition of the measurement, but limited to prime N.

Theorem 1.2 (Candes, Romberg, and Tao [1])

Let N be a prime interger and $f \in \mathbb{C}^N$ be a signal, and T be the support of f. If we have measurements $\hat{f}$ on $\Omega \subset \{0,\ldots,N-1\}$ with

$$|T| \leq \frac{1}{2}|\Omega|, \tag{1.9}$$

then f is exactly recovered from $\hat{f}|_\Omega$.

Deterministic results

Theorem 1.3 (Candes, Romberg, and Tao [1])

Let N be a prime integer and $f \in \mathbb{C}^N$ be a signal, and T be the support of f. If we have measurements $\hat{f}$ on $\Omega \subset \{0, \ldots, N-1\}$ with

$$|T| \leq \frac{1}{2}|\Omega|, \tag{1.10}$$

then f is exactly recovered from $\hat{f}|_\Omega$.
The above inequality is actually sharp: For a fixed $\Omega \subsetneq \{0, \ldots, N-1\}$, there exist signals f and g with support T and S s.t.

$$|T|, |S| \leq \frac{1}{2}|\Omega| + 1 \tag{1.11}$$

and $\hat{f}|_\Omega = \hat{g}|_\Omega$.

Proof
We use the following lemma [1].

Lemma 1.4

Let N be a prime integer and $T, \Omega \subset \{0, \dots, N-1\}$. Then $\mathcal{F}_{T\to\Omega}$ is: bijective if $|T| = |\Omega|$, injective if $|T| \leq |\Omega|$, surjective if $|T| \geq |\Omega|$.

The first claim: Suppose that there exist f, g s.t. $\hat{f}|_\Omega = \hat{g}|_\Omega$ and $|\text{supp}(f)|, |\text{supp}(g)| \leq \frac{1}{2}|\Omega|$. Then, $\hat{f} - \hat{g}|_\Omega = 0$ and

$$|\text{supp}(f-g)| \leq |\text{supp}(f)| + |\text{supp}(g)| \leq |\Omega|.$$

Lemma 1.4 asserts that $\mathcal{F}_{\text{supp}(f-g)\to\Omega}$ is injective, namely $f - g = 0$.

Deterministic results

The second claim: Since $|\Omega| < N$, there are disjoint sets $T, S \subset \{0, \ldots, N-1\}$ s.t. $|T|, |S| \leq \frac{1}{2}|\Omega| + 1$ and $|T| + |S| = |\Omega| + 1$. Now pick up $\omega_0 \in \{0, \ldots, N-1\} \setminus \Omega$. Lemma 1.4 asserts that $\mathcal{F}_{T \cup S \to \Omega \cup \{\omega_0\}}$ is bijective. Pick up any $a \neq 0$, then

$$h := \mathcal{F}^{-1}_{T \cup S \to \Omega \cup \{\omega_0\}}(0, 0, \ldots, 0, a) \tag{1.12}$$

is supported on $T \cup S$ and non-zero. Finally, the signals defined by

$$f := \begin{cases} h & \text{on } T \\ 0 & \text{otherwise,} \end{cases} \qquad g := \begin{cases} -h & \text{on } S \\ 0 & \text{otherwise,} \end{cases}$$

satisfy $|\mathrm{supp}(f)|, |\mathrm{supp}(g)| \leq \frac{1}{2}|\Omega| + 1$ and $\hat{f} - \hat{g}|_\Omega = 0$, namely $\hat{f}|_\Omega = \hat{g}|_\Omega$.

Deterministic results

Remark

In Lemma 1.4 and namely Theorem 1.3, the assumption of N to be prime is essential. If N is not prime, the presence of non-trivial subgroups of $\{0, \ldots, N-1\}$ with modulo N would make the propositions fail.

(Homework exercise)
Find a counter example of Lemma 1.4 for non-prime N.

Deterministic results

Thanks to Theorem 1.3, we learned that we can reconstruct the signal from a given Ω and $\hat{f}|_\Omega$ if the signal is sparse i.e. if we find g s.t. $|\text{supp}(g)| \leq \frac{1}{2}|\Omega|$ and $\hat{g}|_\Omega = \hat{f}|_\Omega$, it's unique.
In principle, we can recover f by solving

$$\min_{g \in \mathbb{C}^N} \|g\|_0 := |\text{supp}(g)|, \quad \hat{g}|_\Omega = \hat{f}|_\Omega. \tag{1.13}$$

But solving a ℓ_0 problem directly is NP-hard. So, how can we practically recover the signal?

Deterministic results

To recover the signal, an algorithm would be:

Algorithm: A naive reconstruction

---

for $T \subset \{0, \ldots, N-1\}$ *s.t.* $|T| \leq \frac{1}{2}|\Omega|$ **do**

  if $\mathcal{F}_{T \to \Omega} f = \hat{f}|_{\Omega}$ is solvable **then**

    Solve for f and finish the iteration.

  end

end

---

[1]

This brute-force attack works, but it is computationally very expensive, especially if Ω is large.
For instance, if $|\Omega| \sim \frac{N}{2}$, the number of possible T is about $4^N 3^{-3N/4}$.

[1] It is a good idea to check the solvability of the system first by checking ranks of $(\mathcal{F}_{T \to \Omega})$ and $(\mathcal{F}_{T \to \Omega}, \hat{f}|_{\Omega})$ because it is cheaper than actually solving the system.

Probabilistic result

We learned a signal can be exactly recovered from partial measurements: $2|T|+1$ consecutive frequencies or arbitrary $2|T|$ frequencies for prime N. In practice, however, sparse signals can be very likely recovered from a more relaxed condition. Furthermore, they can be recovered via ℓ_1-optimization, which is more tractable norm than ℓ_0.

Probabilistic result

We introduce the notion of random sampling/measurement, which means we probabilistically get partial information of Fourier coefficients. For a fixed $0 < \tau < 1$, consider the sequence $\{I_k\}$ of N Bernoulli random variables,

$$I_k := \begin{cases} 1 & \text{with probability } \tau \\ 0 & \text{with probability } 1 - \tau. \end{cases} \tag{1.14}$$

From this, set

$$\Omega := \{k \subset \{0, \ldots, N-1\} \text{ s.t. } I_k = 1\}. \tag{1.15}$$

Note $\mathbb{E}(|\Omega|) = \tau N$.

Probabilistic result

Here is a rough statement of the main theorem.

Theorem 1.5 (Candes, Romberg, and Tao [1])

Let $f \in \mathbb{C}^N$ be a signal, and Ω be the random set. If f is supported on $T \subset \{0, \ldots, N-1\}$ and,

$$\mathbb{E}(|\Omega|) \geq Const.|T| \log N, \tag{1.16}$$

then, f can be exactly recovered from Ω and $\hat{f}|_\Omega$ with high probability as the unique minimizer of

$$\min_{g \in \mathbb{C}^N} \|g\|_1 := \sum_t |g(t)|, \quad \hat{g}|_\Omega = \hat{f}|_\Omega. \tag{1.17}$$

Remark the number of measurements Ω we need is approximately proportional to $|T|$.

Probabilistic result

Here is the exact statement of the main theorem:

Theorem 1.6 (Candes, Romberg, and Tao [1])

Let $f \in \mathbb{C}^N$ be a signal, and Ω be the random set, and M be an accuracy parameter. If f is supported on $T \subset \{0, \ldots, N-1\}$ and,

$$\mathbb{E}(|\Omega|) \geq |T| \log N / \alpha(M), \tag{1.18}$$

then, f can be exactly recovered from Ω and $\hat{f}|_\Omega$ with probability at least $1 - O(N^{-M})$ as the unique minimizer of the aformentioned ℓ_1-problem.

We will later see the explicit expression of $\alpha(M)$, but $1/\alpha(M)$ is roughly proportional to $M + 1$. As such, the higher M is, the more measurement Ω is needed to satisfy the inequality, but the closer to 1 the recovery probability $1 - O(N^{-M})$ is.

Probabilistic result

We will prove Theorem 1.6 from Day 2 to 4.

Remark

As Theorem 1.6 is a probabilistic result, there are Ω and signals that satisfy the inequality but slip out of the recovery even if $\text{supp}(f)$ is much smaller than $|\Omega|$. Typically such a signal has small support in the frequency domain, as such a large Ω but in $\{0, \ldots, N-1\} \setminus \text{supp}(\hat{f})$ gives no information.

(Homework exercise) Find an example.

Real-world application (1D signal)

A signal is often not sparse as is, but is sparse in some sense. In that case, we have a chance to use our technique.

Let's consider the variational signal $\delta(t) := g(t) - g(t-1)$ of a signal g [a]. Note $\sum_t \delta(t) = 0$, and for any ω,

$$\hat{\delta}(\omega) = \sum_t g(t)s^{t\omega} - \sum_t g(t-1)s^{t\omega} = (1 - s^{\omega})\hat{g}(\omega), \quad s = e^{-2\pi i/N}. \tag{1.19}$$

Therefore, we have another $\ell_1$problem,

$$\min_{\delta \in \mathbb{C}^N} \|\delta\|_1, \quad \hat{\delta}|_\Omega = (1 - s^{\omega})\hat{f}|_\Omega. \tag{1.20}$$

[a] $g(-1) := g(N-1)$

As such, if δ is sparse, we can recover δ using the measurement $\hat{f}|_\Omega$, then recover f from δ, up to offset.

2D version of the main theorem

As we expect, the main result can be extended to 2D.

Theorem 1.7 (Candes, Romberg, and Tao [1])

Let $f \in \mathbb{C}^{N \times N}$ be a 2D signal, and Ω be the random set, and M be an accuracy parameter. If f is supported on $T \subset \{0, \ldots, N-1\}^2$ and,

$$\mathbb{E}(|\Omega|) \geq |T| \log N / \alpha(M), \tag{1.21}$$

then, f can be exactly recovered from Ω and $\hat{f}|_{\Omega}$ with probability at least $1 - O(N^{-M})$ as the unique minimizer of the aformentioned ℓ_1-problem.

In the above theorem, 2D DFT $\mathcal{F} : f \to \hat{f}$ is defined by

$$\hat{f}(\omega_1, \omega_2) := \sum_{t_1, t_2} e^{\frac{-2\pi i (t_1 \omega_1 + t_2 \omega_2)}{N}} f(t_1, t_2). \tag{1.22}$$

Real-world application (2D image)

2D images with sparse variation can be reconstructed in a similar way.

For a real-valued 2D signal $g \in \mathbb{R}^{N\times N}$, let's consider the variational signal $\delta(t) := Dg_1(t) + iDg_2(t) \in \mathbb{C}^{N\times N}$ where

$$Dg_1(t_1, t_2) := g(t_1, t_2) - g(t_1 - 1, t_2) \quad (1.23)$$

$$Dg_2(t_1, t_2) := g(t_1, t_2) - g(t_1, t_2 - 1) \quad (1.24)$$

for $t = (t_1, t_2)$. Note $\sum_t \delta(t) = 0$, and for any $\omega = (\omega_1, \omega_2)$,

$$\hat{\delta}(\omega) = ((1 - s^{\omega_1}) + i(1 - s^{\omega_2}))\, \hat{g}(\omega), \quad s = e^{-2\pi i/N}. \quad (1.25)$$

Therefore, we have another $\ell_1$problem,

$$\min_{\delta\in\mathbb{C}^{N\times N}} \|\delta\|_1, \quad \hat{\delta}|_\Omega = ((1 - s^{\omega_1}) + i(1 - s^{\omega_2}))\, \hat{f}|_\Omega. \quad (1.26)$$

As such, if δ is sparse, we can recover δ using the measurement $\hat{f}|_\Omega$, then recover f from the real and imaginary part of δ, up to offset.

How to solve ℓ_1-optimization?

We learned that we can recover signals via ℓ_1-optimization. Since $|x|$ is a convex function, we can expect it is not very hard (if you are interested, google convex optimization).
In fact, ℓ_1-optimization can be seen as a type of problem called linear programming, which is known to be solved in polynomial time. In the last lecture (or HW exercise), we will see how to practically solve it.

Today's summary

We learned

- The concept of compressed sensing:
  We can exactly recover a signal from partial measurements if the signal is sparse.
- ℓ_0, ℓ_1, and ℓ_2-optimization problems
- Deterministic results
- A probabilistic result (main theorem) and a real-world application

What will happen in the following lectures

- Extensions of the main theorem for non-Fourier transform measurements
- Proof of the main theorem
- Related topics

E. Candes, J. Romberg, and T. Tao.
Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information, 2004.

M. Vetterli, P. Marziliano, and T. Blu.
Sampling signals with finite rate of innovation.
IEEE Transactions on Signal Processing, 50(6):1417–1428, 2002.