Compressed Sensing

Theory and Applications

Muhammad Ahmad Kaleem

Engineering Science, University of Toronto

Canadian Undergraduate Mathematics Conference
June 22 2023

Table of Contents

Introduction

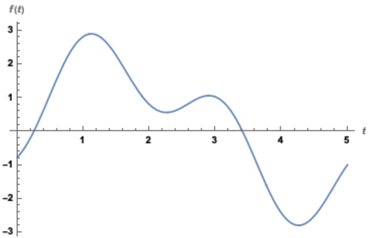
- Data and signals everywhere and growing rapidly
- Amount of data generated larger than total storage capacity and communication bandwidth: efficient storage and transmission important
- Traditional methods of signal acquisition and reconstruction often expensive, don't take into account specifics of signal

Compressed Sensing

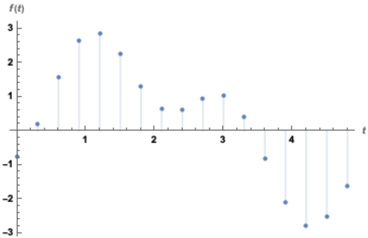
- Signal processing paradigm which improves sampling and recovery of signals
- Focused on finding solutions to underdetermined linear systems and leveraging sparsity, incoherence
- At the intersection of signal processing, statistics, approximation theory etc.
- Initially developed by mathematicians and engineers David Donoho, Emmanuel Candes, Justin Romberg, and Terence Tao in 2004
- Also known as compressive sampling

Signals

- A signal is a representation of a physical phenomenon e.g. audio, video, images
- Can be represented mathematically as an information bearing function (of several variables e.g. time)
- May be continuous or discrete



(a) Continuous



(b) Discrete

Signal Processing Basics

Fourier Transform

Fourier series (for periodic functions)

$$g(t) = \sum_{n=-\infty}^{\infty} c_n e^{i2\pi f_0 nt}$$

$$c_n = \frac{1}{T_0} \int_{-\frac{T_0}{2}}^{\frac{T_0}{2}} g(t) e^{-i2\pi f_0 nt}\, dt$$

Fourier Transform

$$G(f) = \int_{-\infty}^{\infty} g(t) e^{-2\pi ift}\, dt, \quad g(t) = \int_{-\infty}^{\infty} G(f) e^{2\pi ift}\, df$$

- Decompose function into frequency components
- "Change of basis"

Fourier Transform (Example)

- $g(t) = \begin{cases} 1 & |t| \leq \frac{T}{2} \\ 0 & |t| > \frac{T}{2} \end{cases}$ leads to $G(f) = \frac{T\sin(\pi fT)}{\pi fT} = T\ sinc(fT)$
- "Time" and "Frequency" domain interpretation

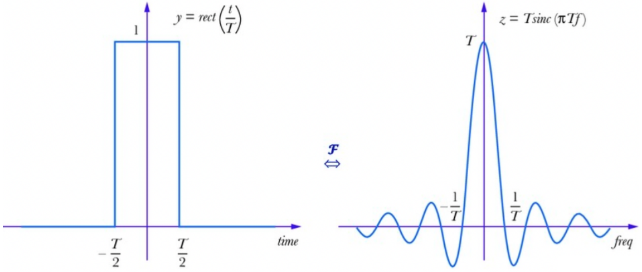


Figure: (Credit: Luo, 2017)

Discrete Fourier Transform

- Fourier Transform can be extended to discrete signals

Discrete Fourier Transform

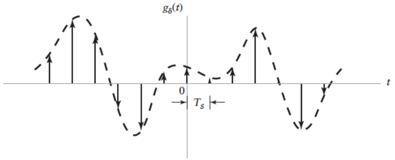
$$X[k] = \frac{1}{\sqrt{n}} \sum_{s=0}^{n-1} x[s] e^{-2\pi i k s / n}, \quad x[s] = \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} X[k] \cdot e^{2\pi i k s / n}$$

- Converts $x \in \mathbb{C}^n$ to $X \in \mathbb{C}^n$
- $\psi_j = \frac{1}{\sqrt{n}} e^{2\pi i k s / n}$ can be viewed as the Fourier basis

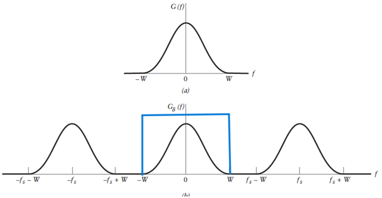
Sampling

Nyquist-Shannon Sampling Theorem

A continuous signal $g(t)$ with maximum frequency (bandwidth) W can be reconstructed perfectly with discrete samples $g(iT_s)$ if $T_s < \frac{1}{2W}$ i.e. using a sampling frequency $f_s > 2W$ (Nyquist rate).



(a) Sampling Process in time domain



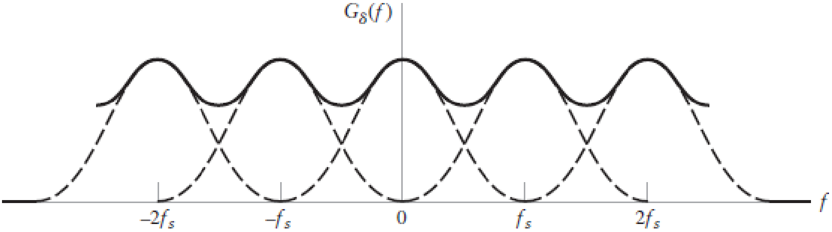
(b) Sampling Process in frequency domain

Figure: (Credit: Haykin and Moher, 2006)

- Conversion from continuous time to discrete time

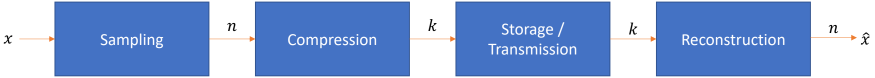
Sampling (cont'd)

- If signal sampled with frequency below the Nyquist rate, **aliasing** occurs



- Worst case bound on samples required, does not consider specifics of signal

Conventional Approach to Sampling



- Acquire n samples of continuous signal x, generate discrete signal $x \in \mathbb{R}^n$
- n samples compressed to k dimensions for storage ($k << n$)
- Signal reconstructed back to n dimensions

Problem Setting

Motivation

- Acquiring n samples and then compressing is wasteful
- n may be very high depending on Nyquist rate
- Obtaining samples may be expensive
- Instead, directly acquire compressed data
- Replace samples by m general measurements:

Sparsity

- Under a certain basis, many signals have a sparse representation
- Consider orthonormal basis $\Psi = [\psi_1 \dots \psi_n] \in \mathbb{R}^{n \times n}$
  - Call Ψ the **representation basis**
- Let $x = \sum_{i=1}^{n} s_i \psi_i$ or equivalently $x = \Psi s$, $s = \begin{bmatrix} s_1 \\ \vdots \\ s_n \end{bmatrix}$
  - By orthogonality, $s = \Psi^T x \implies s_i = \psi_i^T x = \langle x, \psi_i \rangle$
- x k-**sparse** if s has $\leq k$ nonzero elements

Sparsity (Example)

- Consider a camera which takes an image x with n pixels
- Consider representation of image in wavelet basis: n coefficients s_i
- Keep only a fraction of the largest s_i's, zero all other coefficients

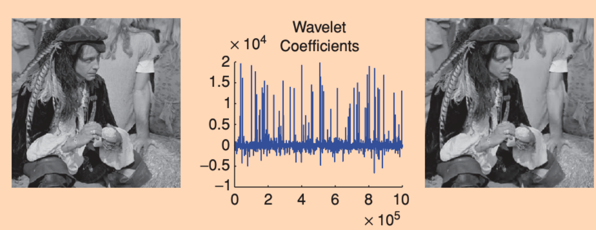


Figure: Image before and after zeroing out lowest 97.5 % coefficients (Credit: Candès and Wakin, 2008)

- JPEG 2000 lossy compression

Sensing Problem

- Consider a discrete signal of interest $x \in \mathbb{R}^n$
- Obtain information about the signal with linear functionals: $y_k = \langle x, \phi_k \rangle$, $k = 1, \ldots, m$: $y = \Phi x$, $y \in \mathbb{R}^m$
- $\Phi = \begin{bmatrix} \phi_1^T \\ \ldots \\ \phi_m^T \end{bmatrix} \in \mathbb{R}^{m \times n}$ is the **measurement matrix**, Undersampled situation $(m << n)$
- Want to recover x given y
- Specifically interested in setting where x is k-sparse $k < m << n$
- Rewrite $y = \Phi x = \Phi \Psi s = As$. $A = \Phi \Psi$ is the **sensing matrix**.
- Assume WLOG that $\Psi = I$ i.e. x is sparse in the space domain and so $y = Ax$
    - For general Ψ we can replace x with s.

Sensing Problem (cont'd)

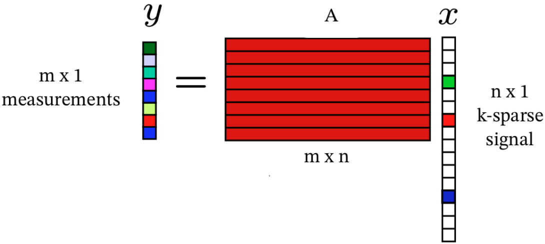


Figure: Compressed Sensing Measurement (Credit: Baraniuk, 2007)

- Want to solve $Ax = y$ for x
- Underdetermined linear system: Fewer measurements than variables $m << n$
  - In general, infinitely many candidates x such that $y = Ax$
    - Example: r in the nullspace of A, then $A(x + r) = Ax$
- **Main Idea:** If x k-sparse, $(k < m)$ the problem can be solved uniquely (assuming conditions on A)

Main Results

Incoherent Sampling

Coherence

The coherence μ between sensing basis Φ and representation basis Ψ is $\mu(\Phi, \Psi) = \sqrt{n} \cdot \max_{k,j} |\langle \phi_k, \psi_j \rangle|$

- $1 \leq \mu \leq \sqrt{n}$
- Intuition: Largest pairwise correlation between elements in Φ and Ψ
- Compressed sensing focused on low coherence pairs: entries of A uniform in magnitude

Reconstruction Problem

- Recall: Given y, the sensing matrix A and the representation basis Ψ, want to reconstruct the sparse signal $x \in \mathbb{R}^n$
- Formulate recovery as a ℓ_0 optimization problem

$$\hat{x} = \arg min_x ||x||_{\ell_0} \text{ subject to } y = Ax \tag{1}$$

  - $||x||_{\ell_0} = \sum_{i=1}^{n} |x_i|^0 = |\{i : x_i \neq 0\}|$ is the sparsity of x

Proposition (Tao, 2009)

If any $2k$ columns of A are linearly independent, any k sparse signal can be uniquely recovered from $y = Ax$

  - Proof: By contradiction. Note that $x - x'$ is $2k$ sparse.
- Therefore ℓ_0 optimization gives unique sparse solution
- Main issue: ℓ_0 minimisation is computationally difficult (NP Hard)

Reconstruction Problem (cont'd)

- Instead formulate recovery as a convex optimization problem (basis pursuit)

$$\hat{x} = \min ||x||_{\ell_1} \text{ subject to } y = Ax \tag{2}$$

    - $||x||_{\ell_1} = \sum_{i=1}^{n} |x_i|$
- ℓ_1 norm as sparsity promoting objective
- Can be solved efficiently with linear programming

Theorem 1 (Candes and Romberg, 2007)

Given $x \in \mathbb{R}^n$ k-sparse in the basis Ψ, if $m \geq C \cdot \mu^2(\Phi, \Psi) \cdot k \cdot \log n$, the solution to optimization problem is exact with overwhelming probability

    - Remark: Lower coherence implies lower value of m

Geometry

- Consider example in $\mathbb{R}^3$
- Nullspace of A translated by x: $r \in \mathcal{N}(A)$, $x' = x + r, y = A(x + r) = Ax$

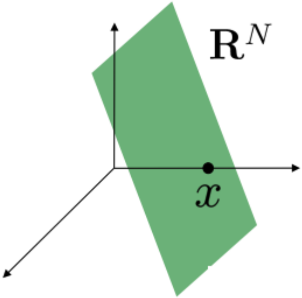


Figure: $\{x' : Ax' = y\}$ (Credit: Baraniuk, 2007)

- Wish to find reconstruction $\hat{x}$ under some criterion

Minimum Norm Reconstruction

- Minimizing ℓ_1 norm promotes sparsity in general
- $\min ||x||_{\ell_1}$ subject to $Ax = y$:

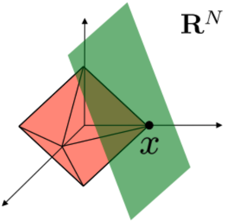


Figure: Solution obtained from ℓ_1 minimization (Credit: Baraniuk, 2007)

- Constant ℓ_1 norm corresponds to the octahedron
  - Recall $||x||_{\ell_1} = \sum_{i=1}^{n} |x_i|$
- Point of intersection with translated nullspace is on the coordinate axis
- Obtain sparse solution corresponding to true value of x

Restricted Isometry Property (RIP)

- RIP used to guarantee that solution to ℓ_1 reconstruction will be exact
- A satisfies RIP of order k if for any k-sparse vector x, we have

$$(1-\epsilon_k)||x||_2 \leq ||Ax||_2 \leq (1+\epsilon_k)||x||_2 \ , 0 < \epsilon_k << 1$$

  - Thus $||x_1 - x_2||_2 \approx ||Ax_1 - Ax_2||_2$ i.e. pairwise distances between sparse signals preserved in measurement space

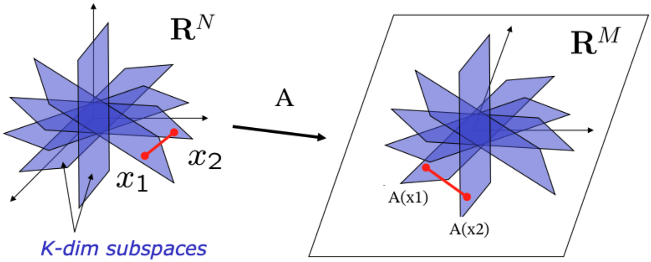


Figure: (Credit: Baraniuk, 2007)

Restricted Isometry Property (cont'd)

- RIP can be shown to be equivalent to condition that any $2k$ columns of A are linearly independent (relates to Proposition earlier)
  - Proof: By contradiction. Suppose $\exists\ x \neq 0$ $2k$ sparse s.t. $Ax = 0$. Then $(1 - \epsilon_{2k})||x||_2 \leq 0$.
- Provides a guarantee on reconstruction:

Theorem 2 (Candès and Wakin, 2008)

If $\epsilon_{2k} < \sqrt{2} - 1$, the solution to (2) satisfies $||\hat{x} - x||_1 \leq C_0 \cdot ||x - x_k||_1$ where x_k is the signal x with only the largest k values being nonzero.

- Thus for k sparse signals, the ℓ_1 reconstruction is exact.

Applications

Single Pixel Camera

- Compressed Sensing inspired design to reconstruct an image
- Using a single sensor, m measurements are acquired by using randomly generated patterns on an array (corresponds to ϕ_k)
- No need to collect n pixel values as a standard camera would do

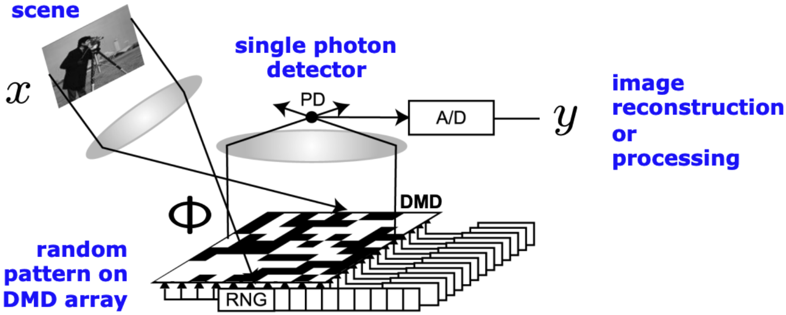


Figure: Design of Camera (Wakin et al., 2006)

Example: Samples from Single Pixel Camera

Figure: 16384 (n) pixel image, reconstruction with 1600 (m) measurements (Wakin et al., 2006)

Other Applications

- Medical Imaging, Inverse Problems e.g. MRI (Lustig et al., 2008)
    - Reduction in scan times while preserving quality
- Error Correcting Codes (Candes and Tao, 2005)
    - Coding matrix A, measurements $y = Ax + e$ where e is unknown sparse vector of errors, x is input vector
    - Recover x exactly even under significant proportion of errors in y
- Astronomy (Bobin et al., 2008)
    - Astronomical imaging and remote sensing
- Analog to Digital Conversion (Wakin et al., 2012)
    - Hardware design based on compressed sensing reduces sampling rate compared to conventional ADC hardware

Bibliography I

Baraniuk, Richard G. (2007). "Compressive Sensing [Lecture Notes]". In: *IEEE Signal Processing Magazine* 24.4, pp. 118–121. DOI: 10.1109/MSP.2007.4286571.

Bobin, Jérôme, Jean-Luc Starck, and Roland Ottensamer (2008). "Compressed sensing in astronomy". In: *IEEE Journal of Selected Topics in Signal Processing* 2.5, pp. 718–726.

Candes, E.J. and T. Tao (2005). "Decoding by linear programming". In: *IEEE Transactions on Information Theory* 51.12, pp. 4203–4215. DOI: 10.1109/TIT.2005.858979.

Candes, Emmanuel and Justin Romberg (2007). "Sparsity and incoherence in compressive sampling". In: *Inverse problems* 23.3, p. 969.

Candès, Emmanuel J and Michael B Wakin (2008). "An introduction to compressive sampling". In: *IEEE signal processing magazine* 25.2, pp. 21–30.

Haykin, S. and M. Moher (2006). *An Introduction to Analog and Digital Communications, 2nd Edition*. Wiley Textbooks. ISBN: 9781118313107. URL: https://books.google.ca/books?id=2RUcAAAAQBAJ.

Luo, Huageng (Dec. 2017). "Physics-based data analysis for wind turbine condition monitoring". In: *Clean Energy* 1. DOI: 10.1093/ce/zkx005.

Bibliography II

Lustig, Michael, David L. Donoho, Juan M. Santos, and John M. Pauly (2008). "Compressed Sensing MRI". In: *IEEE Signal Processing Magazine* 25.2, pp. 72–82. DOI: 10.1109/MSP.2007.914728.

Tao, Terence (2009). "Compressed sensing Or: the equation Ax= b, revisited". In: *Mahler Lecture Series.*

Wakin, Michael, Stephen Becker, Eric Nakamura, Michael Grant, Emilio Sovero, Daniel Ching, Juhwan Yoo, Justin Romberg, Azita Emami-Neyestanak, and Emmanuel Candes (2012). "A Nonuniform Sampler for Wideband Spectrally-Sparse Environments". In: *IEEE Journal on Emerging and Selected Topics in Circuits and Systems* 2.3, pp. 516–529. DOI: 10.1109/JETCAS.2012.2214635.

Wakin, Michael, Jason Laska, Marco Duarte, Dror Baron, Shriram Sarvotham, Dharmpal Takhar, Kevin Kelly, and Richard Baraniuk (Nov. 2006). "An Architecture for Compressive Imaging". In: pp. 1273–1276. DOI: 10.1109/ICIP.2006.312577.

Conventional Approach (Example)

- Consider an image, described as a continuous function $f(x, y)$ (light intensity at different positions)
- Sample this image into a 2D array of width W, height H: discrete image (with pixels)
- $f[r, s]$ where $r, s \in \mathbb{Z}$, $0 \leq r \leq H - 1$, $0 \leq s \leq W - 1$
- Image is compressed for storage
- Image reconstructed for viewing

Minimum Norm Reconstruction (cont'd)

- ℓ_2 norm reconstruction has a closed form solution (least squares) $\hat{x} = (A^T A)^{-1} A^T y$.
- Leads to solution which is incorrect and not sparse

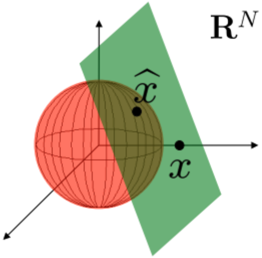


Figure: Solution obtained from ℓ_2 minimization (Credit: Baraniuk, 2007)

- Constant ℓ_2 norm corresponds to sphere
  - Recall $||x||_2 = \sqrt{\sum_{i=1}^{n} x_i^2}$

Random Sensing

- In practice, Φ can be generated randomly
  - Sample column vectors uniformly on the unit sphere of $\mathbb{R}^m$
  - Use iid Gaussian entries from $\mathcal{N}(0, \frac{1}{m})$
- Then for a fixed Ψ, with high probability Φ and Ψ are incoherent and $A = \Phi\Psi$ satisifes the RIP
  - Specifically, for satisfying RIP with high probability, require $m \geq C \cdot k \log(n/k)$
- Such measurement matrices Φ "universal": can construct Φ without knowledge of basis Ψ