# Compressed Sensing

Roman Vershynin

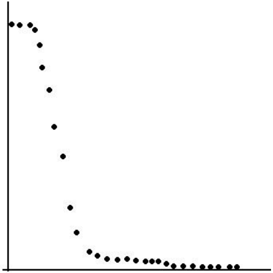
University of California, Davis

September 2007

# Sampling

- Compressed Sensing is a new theory of sampling.
- What's wrong with the old theory?
- Huge information is acquired by sampling, but most of it is waste.

Example: digital photography.



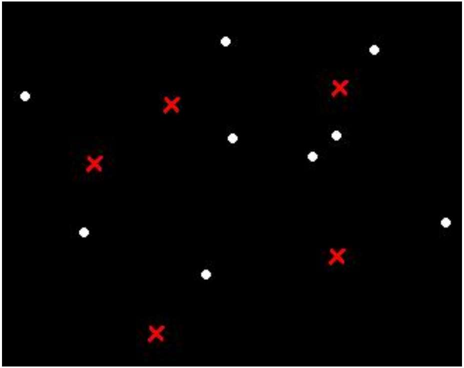
1. Sample at pixels ($1000 \times 1000 = 10^6$); get raw image.
2. Compress: compute wavelet coefficients. They decay fast. Retain only few largest (100); the rest is waste.

# Sampling

- Traditional sampling is wasteful:
  first acquiring the whole signal (huge information),
  then compressing (extracting small useful information).
- Compressed Sensing: combine acquisition and compression into one step. Compress at the time of sampling. No raw images.
- Proposed by [Donoho] and [Candes-Tao] in 2004.

# How Compressed Sensing is possible?

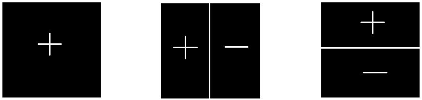
- Undersample. Take fewer samples than the full dimension $1000 \times 1000$.
- Difficulty:



- Sampling points will probably miss all stars. Get black sky.

# How to correctly undersample?

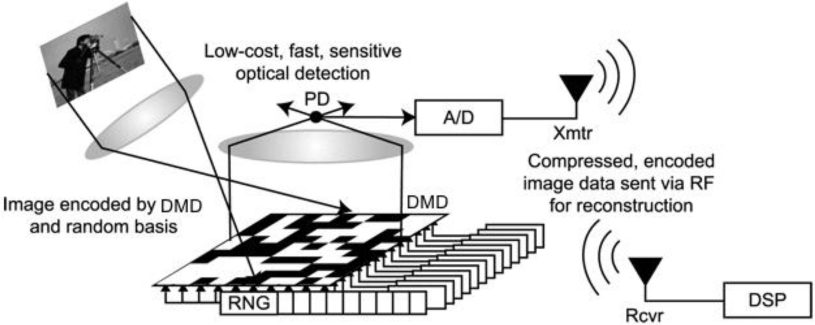
- Instead of sampling at few points,
  sample few linear combinations of all points.
- Example. Sum the values at all points with these signs:




  This gets us 3 samples of the signal.
- Will not miss the stars: all 3 sampled values are positive.
  Such sampling gets us a “collective” information about the image.
- For 3 stars, one expects that 3 samples suffice (instead of $10^6$).
- Yet a better way: for each pixel, assign a random sign ($+$ or $-$).
  Sum the values with these signs.

# Can one really sample this way?

Compressed Sensing Camera being built at Rice:




- http://www.dsp.ece.rice.edu/cs/
- See also Terry Tao's blog, *Single pixel cameras.*
- How do they reconstruct the image? - later.

# General setup of Compressed Sensing

- Model: finite-dimensional sparse signal

$$\text{Signal:} \quad f \in \mathbb{R}^d, \quad |\operatorname{supp}(f)| = n \ll d.$$

  In the starry night example, $n = 3$ and $d = 10^6$.

- Take $N \ll d$ samples of $f$ using some linear operator $A : \mathbb{R}^d \to \mathbb{R}^N$

$$\text{Measurement:} \quad Af \in \mathbb{R}^N,$$

  Thus $Af$ records $N$ different linear combinations of all values of $f$.

- Traditional sampling, at points $x_1, \ldots, x_N$:

$$Af = (f(x_1), \ldots, f(x_N)).$$

- Compressed Sensing:

$$Af \text{ with } A \text{ arbitrary.}$$

  Ideal: $N \sim n$ rather than $d$. Thus 3 measurements rather $10^6$.

# Alternative view: sampling in a different basis

- Compressed Sensing $=$ undersampling in a different basis.
- $U$: some change of basis in $\mathbb{R}^N$.

$$\text{Measurement:} Af = (Uf(\omega_1), \dots, Uf(\omega_N)),$$

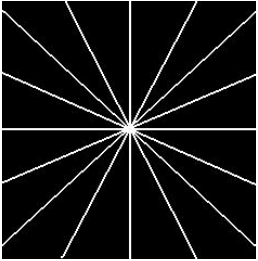
  i.e. $N$ samples of $Uf$ at randomly chosen points.
- What bases are good?
  The more uncorrelated with the original basis, the better.
- Why? To remove any sparsity from $Uf$.
  Change of basis spreads the information from stars to the whole sky. Then can sample effectively.
- What bases are uncorrelated?
  Time - Frequency. Change of basis: Fourier transform.

# Fourier measurements

- $f$: signal sparse in time domain.
  We shall sample it in frequency domain. Fourier measurements:

$$\hat{f}(\omega_1), \ldots, \hat{f}(\omega_N).$$

- Example: medical imaging (MRI).
  Equivalent to sampling $\hat{f}$ along radial lines:

Problem. How to reconstruct $f$? - later.

# Fourier measurements

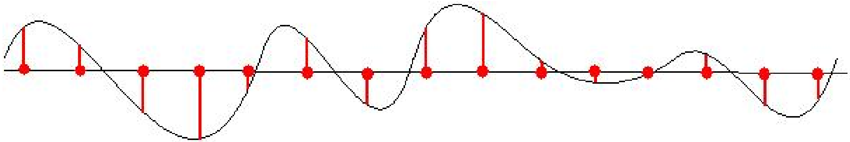
- Or vice versa, $f$: signal sparse in *frequency* domain.
  We shall sample it in *time*: $f(t_1), \dots, f(t_N)$.
- Example: A/D conversion of sound (CD recordings).
- Sparsity assumption: $f$ is band-limited; max frequency is $\Omega/2\pi$:

$$\operatorname{supp}(\hat{f}) \subseteq [-\Omega, \Omega].$$

**Nyquist-Shannon Sampling Theorem**

$f$ can be reconstructed from its time samples at frequency $\Omega/\pi$:

$$f(t) = \sum_{k \in \mathbb{Z}} f(n\pi/\Omega)\,\mathrm{sinc}(\Omega t - n\pi), \qquad \mathrm{sinc}(t) = \sin(t)/t.$$



# Uncertainty Principle

- What makes Compressed Sensing work is the assumption:

  No signal $f$ can be sparse in both bases.

  (If $f$ is sparse in the original basis, then it is *not* sparse in the other basis; so we can sample there).
- This is an abstract form of the Uncertainty Principle: no signal can be localized in both time and frequency.
- Uncertainty Principle thus becomes a *positive* statement.
  This was set forth in [Donoho-Stark, 1989].

# Uncertainty Principle

- $f \in L^2(\mathbb{R})$ and its Fourier transform can not be both supported on compact sets. More generally,

**Proposition (Uncertainty Principle: Nazarov, 1993)**

*If $f \neq 0$ is concentrated on a set $T$ ($1 - \varepsilon$ fraction of its norm) and $\hat{f}$ is concentrated on a set $\Omega$ ($1 - \delta$ fraction of its norm) then*

$$|T| \cdot |\Omega| > c \log \frac{c}{\varepsilon + \delta},$$

*where $c > 0$ is an absolute constant.*

- For Compressed Sensing, we need a discrete version:

# Uncertainty Principle

**Theorem (Discrete Uncertainty Principle: Tao, 2005)**

*Assume $d$ is prime. If $f \in \mathbb{R}^d$ is supported on a set $T$ and $\hat{f}$ is supported on a set $\Omega$ then*

$$|T| + |\Omega| > d.$$

- Not true for composite $d$. (Easy).
- Uncertainty Principle implies Compressed Sensing:

# Uncertainty Principle $\rightarrow$ Compressed Sensing

## Corollary

*Assume $d$ is prime. Then every $n$-sparse signal $f \in \mathbb{R}^d$ is uniquely determined by the values of $\hat{f}$ at any $2n$ points.*

## Proof.

Assume there are two different $n$-sparse signals $f$ and $g$ whose Fourier transforms agree at $2n$ points. Thus

$$|\operatorname{supp}(\widehat{f-g})| \leq d - 2n.$$

Since both $f$ and $g$ are $n$-sparse,

$$|\operatorname{supp}(f-g)| \leq 2n.$$

This contradicts Tao's Uncertainty Principle. $\square$

# Algorithmic reconstruction

- How to reconstruct $f$ from its measurement $Af$? (Formula? Algorithm?)
- $L_2$-minimization (least-squares problem) is a traditional approach:

$$\text{Minimize } \|g\|_2 \text{ subject to } Ag = Af.$$

- Never gives exact reconstruction, even if $f$ is sparse.
- $L_1$-minimization, advocated by Donoho and collaborators:

$$\text{Minimize } \|g\|_1 \text{ subject to } Ag = Af.$$

- Can lead to exact reconstruction if $f$ is sparse.
- Heuristically, $L_1$-norm promotes sparsity: for all vectors of the same $L_2$ norm, the $L_1$ norm is smaller for sparse vectors.

# Reconstruction: $L_1$ minimization

- Reconstruction possible $\Leftrightarrow$ the measurement operator is *one-to-one* for sparse vectors.
- Algorithm exists if $A$ is an *almost isometry* for sparse vectors:

**Theorem (Candes-Tao, 2004)**

*Suppose the measurement operator $A$ is a restricted isometry:*

$$0.8\|f\|_2 \le \|Af\|_2 \le 1.2\|f\|_2 \quad \text{for all } 3n\text{-sparse signals } f.$$

*Then $f$ can be reconstructed from the measurements $Af$ as the solution to the convex optimization problem*

$$\min \|g\|_1 \quad \text{subject to} \quad Ag = Af.$$

- Linear program. Various solvers available.
- Restricted isometry $=$ Uniform Uncertainty Principle for $A$.

# Examples of restricted isometries

- What measurement maps $A : \mathbb{R}^d \to \mathbb{R}^N$ are restricted isometries?

$$0.8\|f\| \le C\|Af\| \le 1.2\|f\| \quad \text{for all } 3n\text{-sparse signals } f.$$

- Need $N \ll d$, so $A$ is not even one-to-one on the whole space.
- Random operators are with high probability:
- Random Gaussian matrix, for $N \sim n \log d$.
  See [Candes-Tao-Rudelson-V., 2005]
- Random Bernoulli matrix ($\pm 1$ entries), for $N \sim n \log d$.
  See [Mendelson-Pajor-Tomczak, 2006].
  Camera at Rice uses these measurements.
- Random Fourier measurements, for $N \sim n \log^4 d$.
  [Rudelson-V., 2005]; see also [Candes-Tao, 2005] with exponent 6.
  Medical imaging uses these measurements.
  Lowering the exponent to 1 may be very hard.

Problem. Deterministic constructions, for $N \sim n \log^{O(1)} d$.

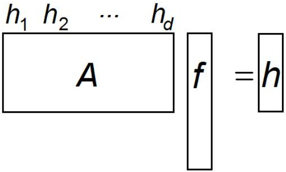
- Currently best known: $N \sim n d^{o(1)}$. [Indyk, 2007].

# Alternatives to $L_1$ minimization

- $L_1$ minimization is based on Linear Programming. Often too slow:
- Theory: no strongly polynomial guarantees on the time.
- Practice: current software too slow for images.
- Alternatives?
- Greedy methods. Compute the support of $f$ iteratively, perhaps one point at a time. (Then easy to compute $f$).
- Greedy methods are well developed in Approximation Theory:

# Compressed Sensing in Approximation Theory

- View the measurement operator (matrix) as $A = (h_1, \ldots, h_d)$.
  View the unknown signal as $f = (c_1, \ldots, c_d)$. (Only $n$ nonzeros).




- View the measurement $h = Af$ as a linear combination

$$h = \sum_{k=1}^{d} c_k h_k .$$

  Then the reconstruction problem becomes:

**Approximation Theory Problem.**

Given a dictionary of functions $(h_1, \ldots, h_d)$ and a function $h$ which is a linear combination of $n \ll d$ dictionary elements, *find these elements*. More generally, given arbitrary $h$, find its best $n$-term approximation.

# Orthogonal Matching Pursuit

- Approximation Theory Problem. Given a dictionary of functions $(h_1, \dots, h_d)$ and a function $h$ which is a linear combination of $n \ll d$ dictionary elements, *find these elements*.
- Idea of greedy methods. Elements $h_k$ that are more correlated with $h$ (with bigger $|\langle h_k, h\rangle|$) are more likely to participate.

**Orthogonal Matching Pursuit (OMP).**

Choose the element $h_k$ most correlated with $h$.
Subtract its contribution: $h := h - c_k h_k$, with $c_k$ minimizing the $L_2$ norm.
Iterate $n$ times.

- Problem. When is OMP correct?
- For random Gaussian dictionary, and for *one* sparse signal $f$, OMP is correct with high probability. [Gilbert, Tropp, 2005]
- But not for all $f$ [Rauhut, 2007]. No deterministic guarantees.
- Nevertheless, practitioners prefer greedy methods. (Camera at Rice uses OMP).

# Regularized Orthogonal Matching Pursuit

- Approximation Theory Problem. Given a dictionary of functions $A = (h_1, \dots, h_d)$ and a function $h$ which is a linear combination of $n \ll d$ dictionary elements, *find these elements*.
- Idea: interpret the Restricted Isometry Condition

$$(1-\varepsilon)\|f\|_2 \le \|Af\|_2 \le (1+\varepsilon)\|f\|_2 \quad \text{for all } 3n\text{-sparse signals } f:$$

"Every $n$ dictionary elements form an almost orthonormal system".

**Regularized Orthogonal Matching Pursuit (ROMP).**

Choose $n$ elements of the dictionary most correlated with $h$...
Subtract their contribution from from $h$. Iterate $2n$ times.

**Theorem (Needell, V., 2007)**

*Under the Restricted Isometry Condition with $\varepsilon \sim 1/\sqrt{\log n}$, the ROMP is correct.*

- Essentially closes the gap between two major methods in Compressed Sensing: $L_1$ minimization and greedy algorithms.

# Summary

- Compressed Sensing: combines acquisition and compression of signals into a single step.
- Uncertainty Principle is a guarantee for Compressed Sensing.
- Under U.P. (Restricted Isometry Condition), one can reconstruct exactly any sparse signal:
  1. using $L_1$-minimization methods (based on Linear Programming);
  2. using greedy methods (based on Approximation Theory).

# Further directions:

- Streaming algorithms: superfast reconstruction, in time proportional to sparsity not the dimension. See survey [Muthukrishnan], works [Gilbert-Tropp-Strauss-V.]
- Compressible signals (not sparse): see [Candes-Tao]
- Connections to high-dimensional convex geometry (polytopes, sections of convex bodies): [Rudelson-V.], [Donoho-Tanner], [Lyubarskii-V.]
- Duality with Error Correction [Candes-Rudelson-Tao-V.]
- Duality with Vector Quantization [Lyubarskii-V.]
- Lots of practical applications: medical imaging, seismology, ...

Compressed Sensing Webpage: http://www.dsp.ece.rice.edu/cs/