DEMOCRACY IN ACTION
Quantization, Saturation, and Compressive Sensing

Jason Laska
COLLABORATORS

Petros Boufounos

Mark Davenport

Richard Baraniuk
If we could first know where we are, and whither we are tending, we could then better judge what to do, and how to do it.

-Abraham Lincoln
Sparsity / Compressibility

Images

N pixels

Chirps

N wideband samples

K \ll N large wavelet coefficients

K \ll N large Gabor coefficients (time-freq)
Sparsity / Compressibility

Wireless spectrum

- N wideband samples
- sparse in (time-frequency)
- most of spectrum is not occupied

K << N

large Gabor coefficients
(time-freq)

Measured Spectrum Occupancy in Chicago and New York City

[http://www.sharedspectrum.com/measurements/recent.html]
Signal Models

Sparse
- \( K \) nonzero signal coefficients

- e.g. man-made signals such as MSK, QPSK, sparse in time-frequency

Compressible
- signal coefficients decay according to
  \[
  |x_n| \propto n^{-1/p} \quad p \leq 1
  \]
- approximate by \( K \) largest coefficients

- e.g. wavelet coefficients of natural images are in range: \(.3 < p < .7\) [DeVore, Jawerth, Lucier]

**coefficients**
- \( K=5 \)

**sorted coefficients**
- K-term approximation
- power law decay
Compressive Sensing

Hallmarks
- non-adaptive linear measurements
- integrates sensing, compression, processing
- measurements are *democratic*, contain a similar amount of information
- progressive reconstruction

Notation
- $\|u\|_2 = \left( \sum_{n=1}^{N} |u_n|^2 \right)^{1/2}$
- $\|u\|_1 = \sum_{n=1}^{N} |u_n|$
**Measurement Model**

\[ y = \Phi \times x \]

- \( y \) is \( M \times 1 \) measurements
- \( \Phi \) is \( M \times N \) nonzero entries
- \( x \) is \( N \times 1 \) sparse signal
- \( K \) is \( K \)-sparse or compressible
- \( \Phi \) is random

\( K < M \ll N \)
**Matrix Requirements**

**Restricted Isometry Property (RIP)**

\[ a \|x\|_2^2 \leq \|\Phi x\|_2^2 \leq b \|x\|_2^2 \]

for all K-sparse \( x \)

- RIP of order 2K implies: for all K-sparse \( u, v \)

\[ a \leq \frac{\|\Phi u - \Phi v\|_2^2}{\|u - v\|_2^2} \leq b \]

for \( 0 < a \leq b < \infty \)

and \( b/a < 1 + \sqrt{2} \)
If elements of $\Phi$ are drawn from subGaussian distribution, then $\Phi$ has RIP with high probability for

$$M = O(K \log(N/K))$$
Reconstruction

$M \times 1$ measurements
$\Phi$
$N \times 1$ sparse signal
$K$ nonzero entries

$y = \Phi x$

**Optimization**

- Bounded measurement noise: $(\|e\|_2 < \epsilon)$ [Candes, Romberg, Tao; Donoho]

$$\hat{x} = \arg\min_x \|x\|_1 \quad \text{s.t.} \quad \|\Phi x - y\|_2 \leq \epsilon$$

**Greedy Algorithms**

- (Orthogonal) Matching Pursuit (OMP) [Tropp, Gilbert]
- Iterative Hard Thresholding (IHT) [Blumensath, Davies]
- Compressive Sampling Matching Pursuit (CoSaMP) [Needell, Tropp]
Reconstruction Guarantees

signal (length-$N$) $\xrightarrow{\Phi}$ measurements (length-$M$) $\xrightarrow{\Phi \x}$ $y = \Phi \x$ $\xrightarrow{+}$ $y + e$ $\xrightarrow{\text{reconstruction}}$ signal estimate

(best $K$-term approximation)

reconstruction error (optimization)

error due to measurement noise

$\| \x - \hat{x} \|_2 \leq C_0 \epsilon + C_1 \frac{\|\x - \x_K\|_1}{\sqrt{K}}$

[Candes, Romberg]

similar guarantees for greedy reconstruction approaches
Random Demodulator
- for Fourier-sparse signals
- tradeoff complexity of ADC with complexity of “randomized hardware”
- maps analog signal to discrete measurements

\[ x(t) \times p_c(t) \rightarrow \text{integrator} \rightarrow \text{ADC} \rightarrow \text{quantizer} \rightarrow y[n] \]

- Pseudorandom Number Generator
- low rate ADC
- Nyquist rate “chipping sequence”

[Tropp, Laska, Romberg, Duarte, Baraniuk; Laska, Kirolos, Duarte, Ragheb, Baraniuk, Massoud]
Single Pixel Camera
- digital micromirror device (DMD) displays random 0/1 patterns
- photodiode computes “optical inner product”
- photodiode voltage is quantized
Quantization

- $\Delta$ quantization interval
- error per measurement bounded:

$$|x_n - R(x_n)| \leq \Delta/2$$

Finite Dynamic Range Quantization

- $G$ “saturation level” (dynamic range)
- $B$ bit-rate (bits per sample)
- quantization interval is $\Delta = 2^{-B+1}G$
- measurements above $G$ saturate
- saturation error is unbounded

all practical quantizers have finite dynamic range
Quantization Error

Infinite dynamic range error is dominated by saturation.

Saturation rate is effectively zero.

Finite dynamic range error can still be significant.

Message: avoid saturation
Avoiding Saturation

Scale down signal such that little or no saturation occurs

Typical rule 6 saturation events per million

Too conservative?
- quantization error increases as gain decreases
- saturation events may be rare

achieved with automatic gain control (AGC)
**CS WITH QUANTIZATION**

- **signal:** (length-N)
- **measurements:** (length-M) \( y = \Phi x \) \((M<<N)\)
- **quantizer:** \( R(\cdot) \)
- **reconstruction:** (non-linear) \( R(y) \)
- **noise:** \( e = y - R(y) \)
- **signal estimate:** \( \hat{x} \)

**Problem**
- ◆ error is unbounded due to saturation events
- ◆ CS results assume bounded errors

**Conventional Approach**
- ◆ avoid saturation  scale down measurements

*can we do better?*
All power tends to corrupt and absolute power corrupts absolutely.

-Lord Acton
**Saturated Measurements**

- \( \tilde{M} \): measurements below saturation level
- \( M - \tilde{M} \): saturated measurements

**Democratic Measurements** (before quantization and saturation)
- each measurement contains roughly the same amount of information

**Saturated Measurements**
- easy to detect
- have magnitude greater than \( G \)
**Saturation Rejection**

**Rejection Approach**
- simply discard saturated measurements

\[
M \times 1 \text{ measurements} \rightarrow y = \Phi \times x \rightarrow N \times 1 \text{ sparse signal}
\]

\[
M \times N \text{ nonzero entries}
\]
Saturation Rejection

Rejection Approach
- simply discard saturated measurements
- discard corresponding rows of $\Phi$

$M \times 1$ measurements

$N \times 1$ sparse signal

$K$ nonzero entries

$M \times N$
**Saturation Rejection**

**Rejection Approach**
- simply discard saturated measurements
- discard corresponding rows of $\Phi$

\[
\tilde{M} \times 1 \quad \text{measurements} \quad = \quad \tilde{y} \quad \tilde{\Phi} \quad \tilde{M} \times N \\
\text{apply standard reconstruction algorithms to} \quad \tilde{y}, \tilde{\Phi}
\]

ex: $\hat{x} = \arg\min_x \|x\|_1 \; \text{s.t.} \; \|\tilde{\Phi}x - \tilde{y}\|_2 < \epsilon$
Saturation Rejection

\[ \tilde{M} \times 1 \text{ measurements} \]

\[ \tilde{\Phi} \]

\[ \tilde{y} = \arg\min_x \|x\|_1 \text{ s.t. } \|\tilde{\Phi}x - \tilde{y}\|_2 < \epsilon \]

Hallmarks
- any out-of-the-box CS reconstruction algorithm can be used
- quantization error on remaining measurements is bounded
- exploits democracy of measurements
- can be used for other types of processing (such as compressive matched filtering for detection)

**but... we are throwing away information**
**Saturation Consistency**

Let's consider a sparse signal measurement scenario with saturation constraints.

**Consistent Approach**
- Modify recovery algorithms so that the solution is consistent with saturation level.
- Add saturation constraint:

\[
|y_s| > G
\]

Where \( y_s \) represents saturated measurements, \( G \) is a threshold, and \( |\cdot| \) denotes the absolute value.

The diagram illustrates the relationship between measurements, sparse signal, and saturated measurements.

- **Measurements**: \( M \times 1 \)
- **Sparse Signal**: \( N \times 1 \)
- **Saturated Measurements**: \( K \) nonzero entries

**Mathematical Equation**

\[
\begin{align*}
M \times 1 & \quad \text{measurements} \\
\Phi & \quad \text{matrix of signal} \\
\times & \quad \text{vector of ones} \\
N \times 1 & \quad \text{sparse signal} \\
K & \quad \text{nonzero entries}
\end{align*}
\]
Saturation Consistency

\[ \Phi^S_+ : \text{rows corresponding to positive saturated measurements} \]
\[ \Phi^S_- : \text{rows corresponding to negative saturated measurements} \]

New \((M - \widetilde{M}) \times N\) matrix:

\[ \hat{\Phi} \triangleq \begin{bmatrix} \Phi^S_+ \\ -\Phi^S_- \end{bmatrix} \]
Saturation Consistency

\[ \hat{x} = \arg\min_x \|x\|_1 \quad \text{s.t.} \quad \|\tilde{\Phi}x - \tilde{y}\|_2 < \epsilon \quad \text{and} \quad \tilde{\Phi}x > G \cdot 1 \]

\[ \hat{x} \text{ is the vector of ones} \]

**Consistent Approach** (optimization)

**Measurement error term** (quantization)

**Saturation consistency constraint**
**Consistent Approach** (optimization) cont...
- exploits *democracy* of measurements
- alternative “measurement error” terms can be used

\[ \| \tilde{\Phi} x - \tilde{y} \|_p < \epsilon \quad p > 2 \]

[Jacques, Hammond, Fadili]

**Consistent Approach** (greedy)
- often faster than optimization
- we introduce saturation consistent CoSaMP (SC-CoSaMP)

**Overview**
- find new supports
- update support set
- estimate coefficients
- prune result
**SC-CoSaMP** (greedy)

$n$ denotes iteration

while (not converged)

compute proxy: $p \leftarrow \tilde{\Phi}^T \left( \tilde{y} - \tilde{\Phi} \hat{x}^{[n]} \right) + \hat{\Phi}^T \tilde{h} \left( G \cdot 1 - \hat{\Phi} \hat{x}^{[n]} \right)$

(\(\tilde{h}(\cdot)\) selects positive elements of vector)

merge support: $\Omega \leftarrow \text{union of}$

- support of largest $2K$ coefficients of $p$
- support of $\hat{x}^{[n]}$

estimate coefficients: $\hat{x}^{[n+1]} \leftarrow \text{argmin}_x \left\| \tilde{\Phi}_\Omega x - \tilde{y} \right\|_2^2 \text{ s.t. } \tilde{\Phi}_\Omega x \geq G \cdot 1$

prune: $\hat{x}^{[n+1]} \leftarrow \text{keep largest } K \text{ coefficients of } \hat{x}^{[n+1]}$

$n \leftarrow n + 1$

end
THREE APPROACHES

Conventional Approach
- *avoid* saturation, scale down measurements

Rejection Approach
- simply *discard* saturated measurements

\[ \hat{x} = \arg\min_{x} \|x\|_1 \quad \text{s.t.} \quad \|\tilde{\Phi}x - \tilde{y}\|_2 < \epsilon \]

Consistent Approach
- use *all* measurements, put *constraint* on saturated ones
  (optimization)

\[ \hat{x} = \arg\min_{x} \|x\|_1 \quad \text{s.t.} \quad \|\tilde{\Phi}x - \tilde{y}\|_2 < \epsilon \quad \text{and} \quad \tilde{\Phi}x > G \cdot 1 \]
  (greedy)
- SC-CoSaMP, SC-IHT
Random Measurements

and

Democracy

Information is the currency of democracy.

-Thomas Jefferson (attributed)
**Setup**

\( \Phi : M \times N \) matrix with \( \mathcal{N}(0, 1/\tilde{M}) \) entries

\( \Phi^\Gamma \) denotes the rows of \( \Phi \) indexed by the set \( \Gamma \)

**Definition**

- We call \( \Phi \) \( \tilde{M} \)-democratic if we have,

\[
a \| x \|_2^2 \leq \| \Phi^\Gamma x \|_2^2 \leq b \| x \|_2^2
\]

for every \( \Gamma \) with \( |\Gamma| = \tilde{M} \) and for all \( K \)-sparse \( x \)

\( a > 0 \)

\( b > a \)

**Notes**

- stronger than RIP for \( \tilde{M} \times N \) matrix
- \( \tilde{\Phi} \) is an instance of \( \Phi^\Gamma \) for signal dependent \( \Gamma \)
Democracy

Theorem

- $\Phi$ is $\widetilde{M}$-democratic with probability exceeding $1 - P_F$

where

$$P_F \leq \left( \frac{3eN}{\epsilon K} \right)^K (P_\alpha + P_\beta)$$

we can pick $M > \mu(\varrho)$, such that for a fixed $\widetilde{M}/M$, $P_F < \varrho < 1$
Proof sketch (that \( \Phi \) is \( \widetilde{M} \)-democratic)

- concentration of measure for \( \Phi^\Gamma \) (in this result, elements of \( \Phi \) are \( \mathcal{N}(0, 1) \))
  - for any \( \mathbf{s} \) in \( \mathbb{R}^N \), draw a random \( \Phi \)
    \[
    \alpha \widetilde{M} \| \mathbf{s} \|_2^2 \leq \| \Phi^\Gamma \mathbf{s} \|_2^2 \leq \beta \widetilde{M} \| \mathbf{s} \|_2^2
    \]
    with high probability

- restricted isometry of \( \Phi^\Gamma \) ("standard" procedure)
  - fix K-dimensional subspace
  - pick set of points \( \mathbf{s} \in S \) such that for any \( \mathbf{x} \) we have \( \| \mathbf{x} - \mathbf{s} \|_2 < \epsilon \)
  - apply concentration of measure to \( \| \Phi^\Gamma \mathbf{s} \|_2 \) to bound \( \| \Phi^\Gamma \mathbf{x} \|_2 \)
  - apply union bound over all K-dimensional subspaces
Concentration of measure: detail
- only need to consider two bounding cases
  - keep rows corresponding to smallest magnitude measurements (for $\alpha$)
  - keep rows corresponding to largest magnitude measurements (for $\beta$)

Example: lower bound

- $y$ are Gaussian
- let $U(y)$ select $\widetilde{M}$ smallest elements of $y$
- if given $\widetilde{M}$-th order statistic of $|y|$, $u$, then
  solve
  
  \[ P_\alpha = \mathbb{P}(\|U(y)\|_2^2 \leq \alpha \widetilde{M}) \]
  \[ = \int_0^\infty \mathbb{P}(\|y\|_2^2 \leq \alpha \widetilde{M} | u) f_{\widetilde{M}:M}(u) du \]

- $y | u$ distributed as truncated Gaussian, use Markov inequality to obtain bound

Definition
- draw $M$ i.i.d. variables, the $\widetilde{M}$-th order statistic is the $\widetilde{M}$-th largest element
Simulations

No man’s knowledge here can go beyond his experience.

-John Locke
Simulation Setup

Signal models
- K-sparse: coefficients are i.i.d. Gaussian, random locations
- p-Compressible: coefficient magnitudes chosen as \( |x_n| \propto n^{-1/p} \)
  assign random signs and positions
  signals are scaled to have unit norm

Measurement matrix
- elements are i.i.d. Gaussian with variance \( 1/M \)

Reconstruction metric
- Signal-to-Noise ratio (SNR)
  \[
  \text{SNR} \triangleq 10 \log_{10} \left( \frac{\|x\|^2_2}{\|x - \hat{x}\|^2_2} \right)
  \]
Three approaches

- **conventional approach**
  - reconstruct with saturated measurements

- **rejection approach**
  - discard saturated measurements

- **consistent approach**
  - use saturated measurements as constraint
**K-sparse Signals**

Fixed parameters: \( N = 1024 \), \( K = 20 \), \( B = 4 \)  
[algorithm: cvx]

M/N = 2/16  
(few measurements)

- **Consistent approach** performs only slightly better.
- **Rejection approach** performs worse.

```
SNR

Saturation Level (G)
```

```
Saturation Rate
```

```
Avg Saturation Rate
```
K-sparse Signals

Fixed parameters: $N = 1024$, $K = 20$, $B = 4$ [algorithm: cvx]

$M/N = 6/16$

- Both approaches perform better for a fixed $G$
- Too much saturation for rejection approach

SNR vs Saturation Rate (G)

15dB too much saturation for rejection approach
Fixed parameters $N = 1024, K = 20, B = 4$ [algorithm: cvx]

M/N = 6/16

optimal performance occurs at significantly nonzero saturation rate
**K-sparse Signals**

Fixed parameters: \(N = 1024\), \(K = 20\), \(B = 4\)  
[algorithm: cvx]

M/N = 15/16  
(many measurements)

34dB
**K-sparse Signals**

Fixed parameters: $N = 1024$, $K = 20$, $B = 4$

M/N = 15/16

(many measurements)

![Graph showing SNR and Saturation Rate](image)

- **Avg Saturation Rate**
- **Conventional**
- **Reject**
- **Consistent**

[Algorithm: cvx]
**K-sparse Signals**

Fixed parameters: $N = 1024$, $K = 20$, $B = 4$  
[algorithm: cvx]

- M/N = 2/16
- M/N = 6/16
- M/N = 15/16

**similar performance for all approaches**

saturation rate should be close to zero

rejection, consistent approaches achieve higher optimal SNR

best performance at nonzero saturation rate
**K-sparse Signals**

Fixed parameters: \( N = 1024 \), \( K = 20 \), \( B = 4 \)

- **M/N = 2/16**
- **M/N = 6/16**
- **M/N = 15/16**

Performance gain increases as a function of M

How do optimal performances compare?
**SC-Cosamp**

**Fixed parameters**
- $N = 1024$
- $K = 20$
- $B = 4$

**rejection approach**
- best SNR is 20dB higher than conventional approach

**consistent approach**
- best SNR is 23dB higher than conventional approach

rejection approach SNR is only 3dB lower than consistent approach SNR

**what does consistency buy us?**

![Graph showing SNR vs. M/N for Conventional, Reject, and Consistent approaches.](image)

- Conventional
- Reject
- Consistent

- Maximum SNR (dB)
- M/N

- Best SNR is 20dB higher than conventional approach
- Consistent approach best SNR is 23dB higher than conventional approach
- Rejection approach SNR is only 3dB lower than consistent approach SNR
- Too few measurements
**Saturation Robustness**

Fixed parameters: \( N = 1024 \), \( M/N = 6/16 \), \( K = 20 \), \( B = 4 \)

*Saturation robustness*

- **range of saturation rates** such that the SNR of each approach is as good as the best conventional approach. SNR vs. saturation level graph is shown.

- SNR of each approach is compared to the conventional approach SNR.
Saturation robustness

range of saturation rates such that SNR of each approach is as good as best conventional approach SNR

Saturation Robustness

Fixed parameters $N = 1024$, $M/N = 6/16$, $K = 20$, $B = 4$ [algorithm: sc-cosamp]

Saturation rate

SNR of each approach is as good as best conventional approach SNR

Saturation Level (G)

SNR

Saturation Rate

Conventional

Reject

Consistent

rejection approach "robustness"
**Saturation Robustness**

Fixed parameters: \( N = 1024 \), \( M/N = 6/16 \), \( K = 20 \), \( B = 4 \) [algorithm: sc-cosamp]

**Saturation robustness**

*range of saturation rates* such that SNR of each approach is *as good* as best conventional approach SNR.

- **Consistent approach** "robustness"
- **Rejection approach** "robustness"

![Graph showing saturation robustness](image)
Saturation Robustness

Fixed parameters $N = 1024$ $M/N = 6/16$ $K = 20$ $B = 4$

Saturation robustness

- range of saturation rates such that SNR of each approach is as good as best conventional approach SNR

consistent approach is more robust to saturation events
Compressible Signals

Fixed parameters  \( N = 1024 \)  \( M/N=6/16 \)  \( B=4 \) [algorithm: cvx]

- **p=0.4**
- **p=0.8**
- **p=1**

p=0.4: Conventional rejection, consistent approaches achieve higher optimal SNR; best performance at non-zero saturation rate. Performance gain decreases as function of p.
Simulation results
- both rejection and consistent approaches can boost reconstruction performance (if enough measurements to spare)
- best performance occurs at significantly nonzero saturation rate
- best performance of both approaches is similar
- consistent approach is more robust to saturation events

Similar results for other measurement systems
- Rademacher (similar is used in CS camera)
- random demodulator
- random sampling
- Nyquist sampling of Fourier-sparse signals, CS reconstruction
Bureaucracy is not an obstacle to democracy but an inevitable compliment to it.

-Joseph Schumpeter
**Automatic Gain Control**

**Conventional AGC**
- scale down for no saturation
- requires complicated hardware and signal heuristics
- often not sensitive to drop in signal strength

**CS AGC**
- significantly nonzero saturation rate
- uses saturation rate only to determine gain

**Setup**
- process in time-blocks (indexed by $w$)
- each block determines the gain for the next
Saturation-based AGC

The system is based on saturation rate only.

\[ \theta[w] = \theta[w-1] + \nu(s - \hat{s}[w-1]) \]

- \( \theta \): gain
- \( s \): desired saturation rate
- \( \hat{s} \): measured saturation rate
- \( w \): block iteration
- \( \nu \): feedback parameter (controls responsiveness)
AGC in Action

Signal strength drops by 90%

Range of quantizer $[-1, 1]$

Set $s = 0.2$

Block-size: 32 measurements

Uses saturation rate only: sensitive to decrease in signal strength
What is not yet done is only what we have not yet attempted to do.

-Alexis de Tocqueville
Fountain Codes [Luby]
- sends $N$ packets, as long as at least $M$ are received, the original data can be reconstructed
- robust to erasures

Multiple Description Coding (MDC)
- partition data into sets (ex: odd, even samples)
- compress each set independently
- reconstruct via interpolation
- progressive and robust to erasures and errors

Democracy ➔ CS measurements are robust to erasures
CS reconstruction ➔ progressive
However strong the general case for democracy, it is not an ultimate or absolute value, and must be judged by what it will achieve. It is probably the best method of achieving certain ends, but not an end in itself.

-F. A. von Hayek
Conclusions

Democracy

- can use any subset of CS measurements (if subset is large enough)

Two New Approaches

- discard saturated measurements
- include saturated measurements as constraint

Hallmarks

- improved reconstruction performance
- improved robustness to saturation
- nonzero saturation rates are encouraged

Extensions

- simple AGC based only on saturation rate

rethink the approach to saturation mitigation

DSP.RICE.EDU/CS