Richard G. Baraniuk

Compressive Sensing

The Shannon/Nyquist sampling theorem specifies that to avoid losing information when capturing a signal, one must sample at least two times faster than the signal bandwidth. In many applications, including digital image and video cameras, the Nyquist rate is so high that too many samples result, making compression a necessity prior to storage or transmission. In other applications, including imaging systems (medical scanners and radars) and high-speed analog-to-digital converters, increasing the sampling rate is very expensive.

This lecture note presents a new method to capture and represent compressible signals at a rate significantly below the Nyquist rate. This method, called compressive sensing, employs nonadaptive linear projections that preserve the structure of the signal; the signal is then reconstructed from these projections using an optimization process [1], [2].

RELEVANCE

The ideas presented here can be used to illustrate the links between data acquisition, compression, dimensionality reduction, and optimization in undergraduate and graduate digital signal processing, statistics, and applied mathematics courses.

PREREQUISITES

The prerequisites for understanding this lecture note material are linear algebra, basic optimization, and basic probability.

PROBLEM STATEMENT

COMPRESSIBLE SIGNALS

Consider a real-valued, finite-length, one-dimensional, discrete-time signal $\mathbf{x}$, which can be viewed as an $N \times 1$ column vector in $\mathbb{R}^N$ with elements $x[n]$, $n = 1, 2, \ldots, N$. (We treat an image or higher-dimensional data by vectorizing it into a long one-dimensional vector.) Any signal in $\mathbb{R}^N$ can be represented in terms of a basis of $N \times 1$ vectors $\{\psi_i\}_{i=1}^N$. For simplicity, assume that the basis is orthonormal. Using the $N \times N$ basis matrix $\Psi = [\psi_1|\psi_2|\ldots|\psi_N]$ with the vectors $\{\psi_i\}$ as columns, a signal $\mathbf{x}$ can be expressed as

$$\mathbf{x} = \sum_{i=1}^{N} s_i\,\psi_i \qquad \text{or} \qquad \mathbf{x} = \psi \mathbf{s} \qquad (1)$$

where $\mathbf{s}$ is the $N \times 1$ column vector of weighting coefficients $s_i = \langle \mathbf{x}, \psi_i \rangle = \psi_i^T \mathbf{x}$ and $\cdot^T$ denotes transposition. Clearly, $\mathbf{x}$ and $\mathbf{s}$ are equivalent representations of the signal, with $\mathbf{x}$ in the time or space domain and $\mathbf{s}$ in the Ψ domain.

The signal $\mathbf{x}$ is K-*sparse* if it is a linear combination of only K basis vectors; that is, only K of the s_i coefficients in (1) are nonzero and $(N - K)$ are zero. The case of interest is when $K \ll N$. The signal $\mathbf{x}$ is *compressible* if the representation (1) has just a few large coefficients and many small coefficients.

TRANSFORM CODING AND ITS INEFFICIENCIES

The fact that compressible signals are well approximated by K-sparse representations forms the foundation of transform coding [3]. In data acquisition systems (for example, digital cameras) transform coding plays a central role: the full N-sample signal $\mathbf{x}$ is acquired; the complete set of transform coefficients $\{s_i\}$ is computed via $\mathbf{s} = \Psi^T \mathbf{x}$; the K largest coefficients are located and the $(N - K)$ smallest coefficients are discarded; and the K values and locations of the largest coefficients are encoded. Unfortunately, this sample-then-compress framework suffers from three inherent inefficiencies. First, the initial number of samples N may be large even if the desired K is small. Second, the set of all N transform coefficients $\{s_i\}$ must be computed even though all but K of them will be discarded. Third, the locations of the large coefficients must be encoded, thus introducing an overhead.

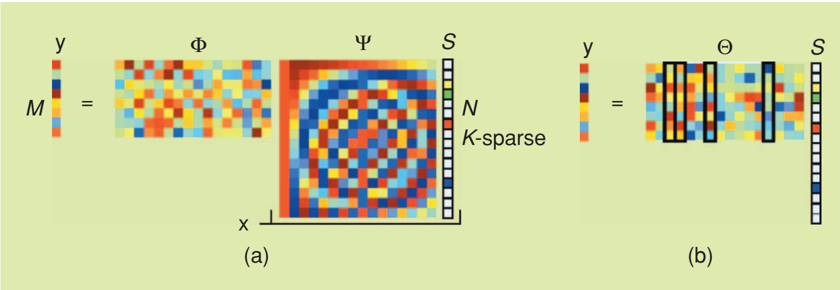
THE COMPRESSIVE SENSING PROBLEM

Compressive sensing address these inefficiencies by directly acquiring a compressed signal representation without going through the intermediate stage of acquiring N samples [1], [2]. Consider a general linear measurement process that computes $M < N$ inner products between $\mathbf{x}$ and a collection of vectors $\{\phi_j\}_{j=1}^M$ as in $y_j = \langle \mathbf{x}, \phi_j \rangle$. Arrange the measurements y_j in an $M \times 1$ vector $\mathbf{y}$ and the measurement vectors ϕ_j^T as rows in an $M \times N$ matrix Φ. Then, by substituting Ψ from (1), $\mathbf{y}$ can be written as

$$\mathbf{y} = \Phi \mathbf{x} = \Phi \Psi \mathbf{s} = \Theta \mathbf{s} \qquad (2)$$

where $\Theta = \Phi\Psi$ is an $M \times N$ matrix. The measurement process is not adaptive, meaning that Φ is fixed and does not depend on the signal $\mathbf{x}$. The problem consists of designing *a*) a *stable measurement matrix* Φ such that the salient information in any K-sparse or compressible signal is not damaged by the dimensionality reduction from $\mathbf{x} \in \mathbb{R}^N$ to $\mathbf{y} \in \mathbb{R}^M$ and *b*) a *reconstruction algorithm* to recover $\mathbf{x}$ from only $M \approx K$ measurements $\mathbf{y}$ (or about as many measurements as the number of coefficients recorded by a traditional transform coder).

1053-5888/07/$25.00©2007IEEE



[FIG1] (a) Compressive sensing measurement process with a random Gaussian measurement matrix Φ and discrete cosine transform (DCT) matrix Ψ. The vector of coefficients s is sparse with $K = 4$. (b) Measurement process with $\Theta = \Phi\Psi$. There are four columns that correspond to nonzero s_i coefficients; the measurement vector y is a linear combination of these columns.

SOLUTION

DESIGNING A STABLE MEASUREMENT MATRIX

The measurement matrix Φ must allow the reconstruction of the length-N signal x from $M < N$ measurements (the vector y). Since $M < N$, this problem appears ill-conditioned. If, however, x is K-sparse and the K locations of the nonzero coefficients in s are known, then the problem can be solved provided $M \geq K$. A necessary and sufficient condition for this simplified problem to be well conditioned is that, for any vector v sharing the same K nonzero entries as s and for some $\epsilon > 0$

$$1 - \epsilon \leq \frac{\|\Theta \mathrm{v}\|_2}{\|\mathrm{v}\|_2} \leq 1 + \epsilon \,. \qquad (3)$$

That is, the matrix Θ must preserve the lengths of these particular K-sparse vectors. Of course, in general the locations of the K nonzero entries in s are not known. However, a sufficient condition for a stable solution for both K-sparse and compressible signals is that Θ satisfies (3) for an arbitrary $3K$-sparse vector v. This condition is referred to as the *restricted isometry property* (RIP) [1]. A related condition, referred to as *incoherence*, requires that the rows $\{\phi_j\}$ of Φ cannot sparsely represent the columns $\{\psi_i\}$ of Ψ (and vice versa).

Direct construction of a measurement matrix Φ such that $\Theta = \Phi\Psi$ has the RIP requires verifying (3) for each of the $\binom{N}{K}$ possible combinations of K nonzero entries in the vector v of length N. However, both the RIP and incoherence can be achieved with high probability simply by selecting Φ as a random matrix. For instance, let the matrix elements $\phi_{j,i}$ be independent and identically distributed (iid) random variables from a Gaussian probability density function with mean zero and variance $1/N$ [1], [2], [4]. Then the measurements y are merely M different randomly weighted linear combinations of the elements of x, as illustrated in Figure 1(a). The Gaussian measurement matrix Φ has two interesting and useful properties:

- The matrix Φ is incoherent with the basis $\Psi = \mathbf{I}$ of delta spikes with high probability. More specifically, an $M \times N$ iid Gaussian matrix $\Theta = \Phi\mathbf{I} = \Phi$ can be shown to have the RIP with high probability if $M \geq cK \log(N/K)$, with c a small constant [1], [2], [4]. Therefore, K-sparse and compressible signals of length N can be recovered from only $M \geq cK \log(N/K) \ll N$ random Gaussian measurements.
- The matrix Φ is universal in the sense that $\Theta = \Phi\Psi$ will be iid Gaussian and thus have the RIP with high probability regardless of the choice of orthonormal basis Ψ.

DESIGNING A SIGNAL RECONSTRUCTION ALGORITHM

The signal reconstruction algorithm must take the M measurements in the vector y, the random measurement matrix Φ (or the random seed that generated it), and the basis Ψ and reconstruct the length-N signal x or, equivalently, its sparse coefficient vector s. For K-sparse signals, since $M < N$ in (2) there are infinitely many s′ that satisfy $\Theta\mathrm{s}' = \mathrm{y}$. This is because if $\Theta\mathrm{s} = \mathrm{y}$ then $\Theta(\mathrm{s} + \mathrm{r}) = \mathrm{y}$ for any vector r in the null space $\mathcal{N}(\Theta)$ of Θ. Therefore, the signal reconstruction algorithm aims to find the signal's sparse coefficient vector in the $(N - M)$-dimensional translated null space $\mathcal{H} = \mathcal{N}(\Theta) + \mathrm{s}$.

- **Minimum ℓ_2 norm reconstruction**: Define the ℓ_p norm of the vector s as $(\|\mathrm{s}\|_p)^p = \sum_{i=1}^{N} |s_i|^p$. The classical approach to inverse problems of this type is to find the vector in the translated null space with the smallest ℓ_2 norm (energy) by solving

$$\widehat{\mathrm{s}} = \operatorname{argmin} \|\mathrm{s}'\|_2 \;\text{ such that }\; \Theta\mathrm{s}' = \mathrm{y}. \qquad (4)$$

This optimization has the convenient closed-form solution $\widehat{\mathrm{s}} = \Theta^T(\Theta\Theta^T)^{-1}y$. Unfortunately, ℓ_2 minimization will almost never find a K-sparse solution, returning instead a nonsparse $\widehat{\mathrm{s}}$ with many nonzero elements.

- **Minimum ℓ_0 norm reconstruction**: Since the ℓ_2 norm measures signal energy and not signal sparsity, consider the ℓ_0 norm that counts the number of non-zero entries in s. (Hence a K-sparse vector has ℓ_0 norm equal to K.) The modified optimization

$$\widehat{\mathrm{s}} = \operatorname{argmin} \|\mathrm{s}'\|_0 \;\text{ such that }\; \Theta\mathrm{s}' = \mathrm{y} \qquad (5)$$

can recover a K-sparse signal exactly with high probability using only $M = K + 1$ iid Gaussian measurements [5]. Unfortunately, solving (5) is both numerically unstable and NP-complete, requiring an exhaustive enumeration of all $\binom{N}{K}$ possible locations of the nonzero entries in s.

- **Minimum ℓ_1 norm reconstruction**: Surprisingly, optimization based on the ℓ_1 norm

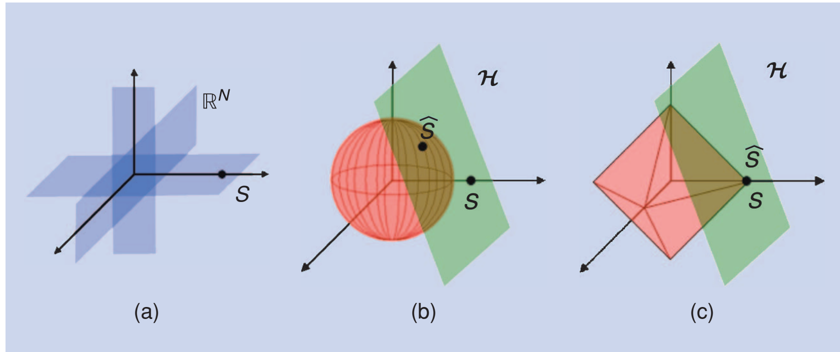
$$\widehat{\mathrm{s}} = \operatorname{argmin} \|\mathrm{s}'\|_1 \;\text{ such that }\; \Theta\mathrm{s}' = \mathrm{y} \qquad (6)$$

can exactly recover K-sparse signals and closely approximate compressible signals with high probability using only $M \geq cK \log(N/K)$ iid Gaussian measurements [1], [2]. This is a convex optimization problem that conveniently reduces to a linear program known as basis pursuit [1], [2] whose computational complexity is about $O(N^3)$. Other, related reconstruction algorithms are proposed in [6] and [7].

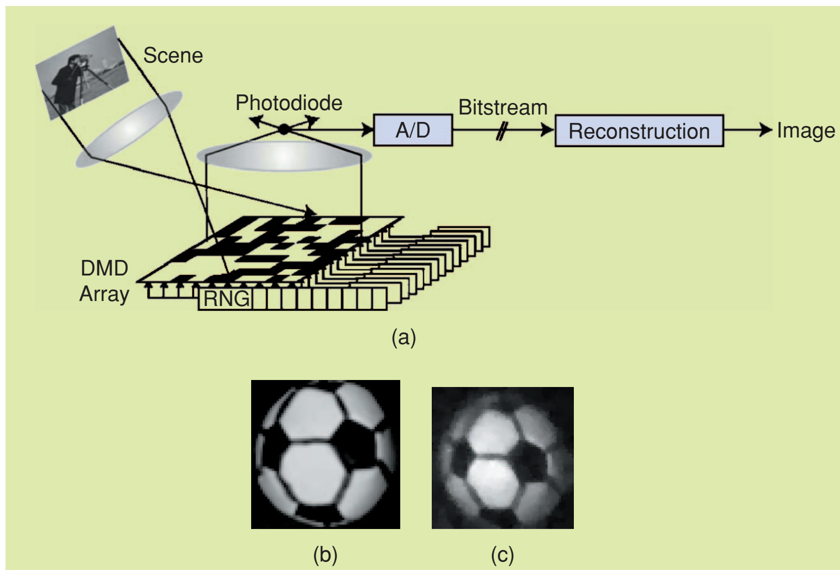
DISCUSSION

The geometry of the compressive sensing problem in $\mathbb{R}^N$ helps visualize why ℓ_2 reconstruction fails to find the sparse solution that can be identified by ℓ_1 reconstruction. The set of all K-sparse vectors $\mathbf{s}$ in $\mathbb{R}^N$ is a highly nonlinear space consisting of all K-dimensional hyperplanes that are aligned with the coordinate axes as shown in Figure 2(a). The translated null space $\mathcal{H} = \mathcal{N}(\Theta) + \mathbf{s}$ is oriented at a random angle due to the randomness in the matrix Θ as shown in Figure 2(b). (In practice $N, M, K \gg 3$, so any intuition based on three dimensions may be misleading.) The ℓ_2 minimizer $\widehat{\mathbf{s}}$ from (4) is the point on $\mathcal{H}$ closest to the origin. This point can be found by blowing up a hypersphere (the ℓ_2 ball) until it contacts $\mathcal{H}$. Due to the random orientation of $\mathcal{H}$, the closest point $\widehat{\mathbf{s}}$ will live away from the coordinate axes with high probability and hence will be neither sparse nor close to the correct answer $\mathbf{s}$. In contrast, the ℓ_1 ball in Figure 2(c) has points aligned with the coordinate axes. Therefore, when the ℓ_1 ball is blown up, it will first contact the translated null space $\mathcal{H}$ at a point near the coordinate axes, which is precisely where the sparse vector $\mathbf{s}$ is located.

While the focus here has been on discrete-time signals $\mathbf{x}$, compressive sensing also applies to sparse or compressible analog signals $x(t)$ that can be represented or approximated using only K out of N possible elements from a continuous basis or dictionary $\{\psi_i(t)\}_{i=1}^N$. While each $\psi_i(t)$ may have large bandwidth (and thus a high Nyquist rate), the signal $x(t)$ has only K degrees of freedom and thus can be measured at a much lower rate [8], [9].



[FIG2] (a) The subspaces containing two sparse vectors in $\mathbb{R}^3$ lie close to the coordinate axes. (b) Visualization of the ℓ_2 minimization (5) that finds the non-sparse point-of-contact $\widehat{s}$ between the ℓ_2 ball (hypersphere, in red) and the translated measurement matrix null space (in green). (c) Visualization of the ℓ_1 minimization solution that finds the sparse point-of-contact $\widehat{s}$ with high probability thanks to the pointiness of the ℓ_1 ball.



[FIG3] (a) Single-pixel, compressive sensing camera. (b) Conventional digital camera image of a soccer ball. (c) 64 × 64 black-and-white image $\widehat{x}$ of the same ball ($N = 4{,}096$ pixels) recovered from $M = 1{,}600$ random measurements taken by the camera in (a). The images in (b) and (c) are not meant to be aligned.

PRACTICAL EXAMPLE

As a practical example, consider a single-pixel, compressive digital camera that directly acquires M random linear measurements without first collecting the N pixel values [10]. As illustrated in Figure 3(a), the incident light-field corresponding to the desired image $\mathbf{x}$ is reflected off a digital micromirror device (DMD) consisting of an array of N tiny mirrors. (DMDs are present in many computer projectors and projection televisions.) The reflected light is then collected by a second lens and focused onto a single photodiode (the single pixel). Each mirror can be independently oriented either towards the photodiode (corresponding to a 1) or away from the photodiode (corresponding to a 0). To collect measurements, a random number generator (RNG) sets the mirror orientations in a pseudorandom 1/0 pattern to create the measurement vector ϕ_j. The voltage at the photodiode then equals y_j, which is the inner product between ϕ_j and the desired image $\mathbf{x}$. The process is repeated M times to obtain all of the entries in $\mathbf{y}$.

(continued on page 124)

line structure in the spectrum of the recursive CORDIC, and the phase error correction is not applied to suppress phase error artifacts but rather to complete the phase rotation left incomplete due to the residual phase term in the angle accumulator. This is a very different DDS!

IMPLEMENTATION

As a practical note, there are truncating quantizers between the AGC multipliers and the feedback delay element registers. As such, the truncation error circulates in the registers and contributes an undesired dc component to the complex sinusoid output. This dc component can (and should) be suppressed by using a sigma delta-based dc cancellation loop between the AGC multipliers and the feedback delay elements [6].

CONCLUSIONS

We modified the traditional recursive DDS complex oscillator structure to a tangent/cosine configuration. The tan(θ) computations were implemented by CORDIC rotations avoiding the need for multiply operations. To minimize output phase angle error, we applied a post-CORDIC clean-up angle rotation. Finally, we stabilized the DDS output amplitude by an AGC loop. The phase-noise performance of the DDS is quite remarkable and we invite you, the reader, to take a careful look at its structure. A MATLAB-code implementation of the DDS is available at http://apollo.ee.columbia.edu/spm/?i=external/tipsandtricks.

ACKNOWLEDGMENT

Thanks to Rick Lyons for patience and constructive criticism above and beyond the call of duty.

AUTHOR

Fred Harris (fred.harris@sdsu.edu) teaches DSP and modem design at San Diego State University. He holds 12 patents on digital receivers and DSP technology. He has written over 140 journal and conference papers and is the author of the book *Multirate Signal Processing for Communication Systems* (Prentice Hall Publishing).

REFERENCES

[1] C. Dick, F. Harris, and M. Rice, "Synchronization in software defined radios—Carrier and timing recovery using FPGAs," in *Proc. IEEE Symp. Field-Programmable Custom Computing Machines*, Napa Valley, CA, pp. 195–204, Apr. 2000.

[2] J. Valls, T. Sansaloni, A. Perez-Pascual, V. Torres, and V. Almenar, "The use of CORDIC in software defined radios: A tutorial," *IEEE Commun. Mag.*, vol. 44, no. 9, pp. 46–50, Sept. 2006.

[3] F. Harris, C. Dick, and R. Jekel, "An ultra low phase noise DDS," presented at Software Defined Radio Forum Tech. Conf. (SDR-2006), Orlando FL, Nov. 2006.

[4] R. Lyons, *Understanding Digital Signal Processing*, 2nd ed. Upper Saddle River, NJ: Prentice Hall, pp. 576–578, 2004.

[5] C. Turner, "Recursive discrete-time sinusoidal oscillators," *IEEE Signal Processing Mag.*, vol. 20, no. 3, pp. 103–111, May 2003.

[6] C. Dick and F. Harris, "FPGA signal processing using sigma-delta modulation," *IEEE Signal Processing Mag.*, vol. 17, no. 1, pp. 20–35, Jan. 2000. SP

 continued from page 120

An image acquired with the single-pixel camera using about 60% fewer random measurements than reconstructed pixels is illustrated in Figure 3(c); compare to the target image in Figure 3(b). The reconstruction was performed via a total variation optimization [1], which is closely related to the ℓ_1 reconstruction in the wavelet domain. In addition to requiring fewer measurements, this camera can image at wavelengths where is difficult or expensive to create a large array of sensors. It can also acquire data over time to enable video reconstruction [10].

CONCLUSIONS: WHAT WE HAVE LEARNED

Signal acquisition based on compressive sensing can be more efficient than traditional sampling for sparse or compressible signals. In compressive sensing, the familiar least squares optimization is inadequate for signal reconstruction, and other types of convex optimization must be invoked.

ACKNOWLEDGMENTS

This work was supported by grants from NSF, DARPA, ONR, AFOSR, and the Texas Instruments (TI) Leadership University Program. Special thanks are due to TI for the DMD array used in the single-pixel camera. Thanks also to the Rice DSP group and Ron DeVore for many enlightening discussions and Justin Romberg for help with the reconstruction in Figure 3.

AUTHOR

Richard G. Baraniuk (richb@rice.edu) is the Victor E. Cameron Professor of Electrical and Computer Engineering at Rice University. His research interests include multiscale analysis, inverse problems, distributed signal processing, and sensor networks. He is a Fellow of the IEEE.

REFERENCES

Additional compressive sensing resources are available at dsp.rice.edu/cs.

[1] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[2] D. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[3] S. Mallat, *A Wavelet Tour of Signal Processing*. New York: Academic, 1999.

[4] R.G. Baraniuk, M. Davenport, R. DeVore, and M.B. Wakin, "A simple proof of the restricted isometry principle for random matrices (aka the Johnson-Lindenstrauss lemma meets compressed sensing)," *Constructive Approximation*, 2007 [Online]. Available: http://dsp.rice.edu/cs/jlcs-v03.pdf

[5] D. Baron, M.B. Wakin, M. Duarte, S. Sarvotham, and R.G. Baraniuk, "Distributed compressed sensing," 2005 [Online]. Available: http://dsp.rice.edu/cs/DCS112005.pdf

[6] J. Tropp and A.C. Gilbert, "Signal recovery from partial information via orthogonal matching pursuit," Apr. 2005 [Online]. Available: http://www-personal.umich.edu/~jtropp/papers/TG06-Signal-Recovery.pdf

[7] J. Haupt and R. Nowak, "Signal reconstruction from noisy random projections," *IEEE Trans. Inform. Theory*, vol. 52, no. 9, pp. 4036–4048, Sept. 2006.

[8] S. Kirolos, J. Laska, M. Wakin, M. Duarte, D. Baron, T. Ragheb, Y. Massoud, and R.G. Baraniuk, "Analog-to-information conversion via random demodulation," in *Proc. IEEE Dallas Circuits Systems Workshop*, Oct. 2006, pp. 71-74.

[9] M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation," *IEEE Trans. Signal Processing*, vol. 50, no. 6, pp. 1417–1428, June 2002.

[10] D. Takhar, V. Bansal, M. Wakin, M. Duarte, D. Baron, J. Laska, K.F. Kelly, and R.G. Baraniuk, "A compressed sensing camera: New theory and an implementation using digital micromirrors," in *Proc. Comput. Imaging IV SPIE Electronic Imaging*, San Jose, Jan. 2006. SP