Distributed processing in frames for sparse approximation

Christopher Rozell
Redwood Center for Theoretical Neuroscience
School of Optometry
University of California, Berkeley
Berkeley, CA 94720
Email: crozell@berkeley.edu

Abstract—**Beyond signal processing applications, frames are also powerful tools for modeling the sensing and information processing of many biological and man-made systems that exhibit inherent redundancy. In many cases, these systems are required to use distributed computational strategies to analyze and process the sensory information. In this talk, I will review the use of frames to model distributed sensing systems with a particular focus on sensory neural systems. In light of the evidence that many of these systems employ sparse codes, I will describe our Locally Competitive Algorithms (LCAs) that use a dynamical system to solve many sparse approximation problems. These LCAs employ a parallel computational architecture with simple analog components. I will show numerical simulation results for these systems and describe their relationship to the many recently-proposed iterative thresholding algorithms. Our LCA approach also demonstrates potential advantages in coding time-varying signals (e.g., video) by reflecting the smooth signal changes in smooth coefficient variations. Finally, I will highlight some future directions where we hope to impact areas such as efficient analog signal processing devices, fast discrete approximation algorithms, and video processing and computer vision in complex temporal environments.**

I. Introduction

Sensing systems are the window that allows both man-made and biological systems to view the outside world. The ability of a machine or a biological organism to understand its local environment relies on the methods used to collect and process this sensory information. Understanding more about optimal information processing strategies in sensing systems is a critical goal both for science and engineering.

Unfortunately, many sensing systems have two characteristics that are not uniformly well understood analytically: redundancy and distributed processing. In particular, these characteristics appear to be ubiquitous in biological sensory neural systems and their roles are increasingly important in modern signal processing systems. Understanding the performance of these sensing systems will be aided by having good analytic models that can incorporate both informational redundancy and the distributed computation of (perhaps complex) processing tasks.

Conventional signal processing wisdom associates redundancy with robustness. This viewpoint states that redundancy is useful to recover from corruption but should be removed in all other cases for maximum efficiency. This conflict appears in the classic information theory result known as the "separation theorem" [1], which divides the fundamental communication pathway into two stages. The first stage (source coding) removes as much redundancy as possible from a signal representation. The second stage (error control coding) introduces redundancy back into the signal representation to counter the inevitable communication errors.

While the communications paradigm describes efficient information transmission, it does not address the task of trying to efficiently *understand* the content of a sensed signal (i.e., trying to infer the environmental conditions responsible for the signal). When considering this broader task, a second advantage of redundancy comes to light: redundancy allows *flexibility* in the representation. Critically sampled representations (e.g., an orthonormal basis) contain enough information to reconstruct a signal but they must use a single fixed strategy for the encoding. In a redundant representation, the redundancy allows us to choose from among many possible encodings to find one that not only communicates the sensed signal but also includes other desirable properties. In particular, there are many reasons to desire the sparsest encoding, using the fewest number of non-zero coefficients.

Many modern signal processing problems rely on distributed processing, either because the data was originally gathered in that form (e.g., from a physically dispersed collection of sensors) or because the available processing resources are inadequate for centralized computation on a full dataset. Again, conventional wisdom says to remove the distributed processing constraints as quickly as possible in favor of centralized computations. The most favorable situation distills the data down to a single sufficient statistic that can be evaluated in one central location.

This desire for data centralization is understandable. Managing the redundancy in a representation requires coordination that is most easily achieved with centralized oversight. In contrast, neural systems appear to take an opposite strategy. Rather than centralize information down to a single decision-making cell, neural systems appear to use the joint activity in successively larger neural populations to analyze a sensory stimulus. The example of neural computation indicates that there exist efficient strategies for managing redundant representations even without centralized control.

This paper will address the use of frames as a platform to model the sensing and distributed processing performed in a variety of sensing systems. In particular, we will review this modeling approach in both a man-made and a biological system. Section III will use the example of a wireless sensor network as a man-made system with distributed processing constraints to reduce communication costs. Section IV will use the example of sparse coding in primary visual cortex as a biological system with an architectural constraint that requires distributed processing. In the context of this model, we will review our recent development of a novel class of neurally plausible *locally competitive algorithms* (LCAs) that solve a collection of sparse coding principles.

II. Background

A. Frames

In general, measurements made by nearby sensors are highly correlated because the sensor receptive fields are not orthogonal and may be linearly dependent. When a representation system has vectors that are linearly dependent, the collection of vectors technically no longer forms a basis. A collection of K vectors $\{\boldsymbol{\phi}_k\}$ is said to form a *frame* [2]–[4] for the space $\mathcal{H}$ if there exist constants $0 < A \leq B < \infty$ such that Parseval's relation is bounded for any $\boldsymbol{x} \in \mathcal{H}$,

$$A \|\boldsymbol{x}\|^2 \leq \sum_{k \in K} |\langle \boldsymbol{\phi}_k, \boldsymbol{x} \rangle|^2 \leq B \|\boldsymbol{x}\|^2 .$$

Because of the dependency present between frame vectors, a set of dual vectors are used for the reconstruction,

$$\boldsymbol{x} = \sum_{k \in K} \langle \boldsymbol{x}, \boldsymbol{\phi}_k \rangle \widetilde{\boldsymbol{\phi}}_k.$$

While there are an infinite number of sets of dual vectors that will work for reconstruction, the canonical dual set is calculated to reduce the error due to corrupted coefficients as much as possible. One would expect that noise reduction would be a significant advantage of using a redundant frame.

B. Fusion frames

To begin analyzing the reconstruction from redundant elements in a distributed way, we turn to a new theory of fusion frames described in [5]. Here, signals are decomposed in terms of overlapping subspaces. A family of closed subspaces $\{W_i\}_{i=1}^L$ is a *fusion frame* for $\mathcal{H}$ if for every signal $\boldsymbol{x} \in \mathcal{H}$,

$$A^s \|\boldsymbol{x}\|^2 \leq \sum\nolimits_i \|\pi_i(\boldsymbol{x})\|^2 \leq B^s \|\boldsymbol{x}\|^2 , \qquad (1)$$

where $0 < A^s \leq B^s < \infty$ are the frame bounds and $\pi_i(\cdot)$ is the orthogonal projection onto W_i. A fusion frame is in many ways analogous to a frame. In frame theory, an input signal is represented by a collection of scalar coefficients that measure the projection of that signal onto each frame vector. In a fusion frame, an input signal is represented by a collection of *vector* coefficients that represent the projection (not just the projection energy) onto the each subspace. Formally, for a fusion frame, the representation space is defined as $\mathcal{V} = \left\{\{\boldsymbol{x}_i\} | \boldsymbol{x}_i \in W_i \text{ and } \boldsymbol{x}_i \in l^2\right\}$, the space of all collections of finite-energy vectors containing one representative from each subspace.

Analogous to frame theory, a fusion frame has an analysis operator, $\mathbf{W} : \mathcal{H} \to \mathcal{V}$, $\mathbf{W}\boldsymbol{x} = \{\pi_i(\boldsymbol{x})\} = \{\boldsymbol{x}_i\}$. The adjoint of the analysis operator is the synthesis operator, $\mathbf{W}^* : \mathcal{V} \to \mathcal{H}$, $\mathbf{W}^*\{\boldsymbol{x}_i\} = \sum_i \boldsymbol{x}_i$. The composition of the analysis and synthesis operators form the frame operator, $(\mathbf{W}^*\mathbf{W}) : \mathcal{H} \to \mathcal{H}$, $(\mathbf{W}^*\mathbf{W})\, \boldsymbol{x} = \mathbf{W}^*\mathbf{W}\boldsymbol{x} = \sum_i \pi_i(\boldsymbol{x})$. The frame operator is bounded and invertible, with bounds $A^s \leq \|(\mathbf{W}^*\mathbf{W})\, \boldsymbol{x}\| \leq B^s$ and $\frac{1}{B^s} \leq \left\|(\mathbf{W}^*\mathbf{W})^{-1} \boldsymbol{x}\right\| \leq \frac{1}{A^s}$ for all $\|\boldsymbol{x}\| = 1$. As with a standard frame, the unique pseudoinverse for $\mathbf{W}$ is given by $\mathbf{W}^\dagger = (\mathbf{W}^*\mathbf{W})^{-1} \mathbf{W}^*$.

One of the fundamental results of [5] shows that fusion frames also provide a link between the properties of a global frame and local collections of frame elements. If each subspace W_i has an associated family of vectors that locally form a frame for W_i with bounds (A_i, B_i), then the total collection of vectors globally form a frame for $\mathcal{H}$ with bounds (A, B).

Because of this link, we can use fusion frames to model the effects of distributed signal reconstruction in a sensing system. In this model, the frame vectors local to one subspace are used to reconstruct the orthogonal projection of the signal into that subspace while maximizing the noise reduction properties of the local frame. The collection of reconstructed signals within each subspace are then used to reconstruct the original input signal while maximizing the noise reduction properties of the frame of subspaces. This subspace-based reconstruction distributes the processing by only requiring knowledge of the frame vectors within a subspace to do a local reconstruction. Such a scheme adds a level of robustness to the system because only vectors in one local subspace are affected if a frame vector is added or removed.

C. Sparse approximation

Given an N-dimensional stimulus $\boldsymbol{x} \in \mathbb{R}^N$ (e.g., an N-pixel image), we seek a representation in terms of a dictionary $\mathcal{D}$ composed of K unit-length vectors $\{\boldsymbol{\phi}_k\}$ that span the space $\mathbb{R}^N$. When the dictionary is overcomplete ($K > N$), there are an infinite number of ways to choose coefficients $\{a_k\}$ such that $\boldsymbol{x} = \sum_{k=1}^{K} a_k \boldsymbol{\phi}_k$. In sparse approximation, we seek the coefficients having the fewest number of non-zero entries by solving the minimization problem

$$\min_{\boldsymbol{a}} \sum_k C(a_k) \quad \text{subject to} \quad \left\| \boldsymbol{x} - \sum_{k=1}^{K} a_k \boldsymbol{\phi}_k \right\|^2 < \epsilon, \tag{2}$$

where $C(\cdot)$ is a sparsity-inducing cost function.

In the signal processing community, two approaches are typically used to find acceptable solutions to this problem. The first general approach, known as Basis Pursuit De-Noising (BPDN) [6], uses the convex ℓ^1 norm as a cost function. There are many algorithms that can be used to solve the BPDN optimization problem, with interior point-type methods being the most common choice. The second general approach employed by signal processing researchers uses iterative greedy algorithms to constructively build up a signal representation [7] in the hopes of approximating a solution with the ℓ^0 norm as a cost function. The canonical example of a greedy algorithm is known in the signal processing community as Matching Pursuit (MP) [8]. MP simply selects the dictionary element with the largest (magnitude) inner product with the current residual, remembers the coefficient it selected, and updates the residual. Though they may not be optimal in general, greedy algorithms often efficiently find good sparse signal representations in practice.

III. Man-made sensing: wireless sensor networks

Wireless sensor networks are a canonical example of a man-made sensing system that is constrained to use distributed processing strategies. In this setting, sensor measurements, represented by inner products of sensor receptive fields with an environmental signal $\langle \boldsymbol{\phi}_k, \boldsymbol{x} \rangle$, must be processed without sending all of the data to a central fusion center. While the redundancy in the sensor network will still provide robustness to measurement noise, some of these noise reduction properties will certainly be sacrificed because of the distributed processing constraint. This distributed processing restriction can be modeled using the concept of a fusion frame [9].

Using the tools of fusion frame, we can consider the noise reduction capability of a redundant expansion under distributed processing requirements. Let $\{W_i\}_{i=1}^{L}$ be a frame of subspaces for the space $\mathcal{H}$, with frame bounds (A^s, B^s). Let each subspace be spanned by a collection of K_i vectors that locally form a frame for W_i with frame bounds (A_i, B_i). When taken together, these vectors form a frame for $\mathcal{H}$ with bounds (A, B). A signal $\boldsymbol{x}$ is represented by all of the vectors in the global frame, and coefficients are corrupted by additive noise with subspace-dependent variance σ_i^2.

For distributed reconstruction, local frame vectors are first used to reconstruct the projection of the signal onto each subspace, $\widehat{\boldsymbol{x}}_i$. In [10] we derive a bound on the total MSE when reconstructing each $\boldsymbol{x}_i = \pi_i(\boldsymbol{x})$, $\frac{\sigma_i^2 N}{B_i} \leq \mathcal{E}\left[\|\widehat{\boldsymbol{x}}_i - \boldsymbol{x}_i\|^2\right] \leq \frac{\sigma_i^2 N}{A_i}$. Distributed reconstruction $\widehat{\boldsymbol{x}}^d$ is completed by using each $\widehat{\boldsymbol{x}}_i$ as a corrupted coefficient in the fusion frame. The previous bound tells us that we can view the subspace projections $\boldsymbol{x}_i$ as being corrupted independently by an additive noise vector $\boldsymbol{n}_i \in W_i$, with covariance matrix $\boldsymbol{\Gamma}_i$ that has bounded total variance $\frac{\sigma_i^2 N}{B_i} \leq \mathrm{Tr}\left[\boldsymbol{\Gamma}_i\right] \leq \frac{\sigma_i^2 N}{A_i}$. If we define $\widetilde{\boldsymbol{\Gamma}} = \sum_i \boldsymbol{\Gamma}_i$, we can also bound the MSE per signal dimension [10],

$$\begin{aligned} \frac{L\sigma_{\min}^2}{B_{\max} {B^s}^2} \leq \frac{1}{{B^s}^2} \sum\nolimits_i \frac{\sigma_i^2}{B_i} \leq \frac{\mathcal{E}\left[\left\|\widehat{\boldsymbol{x}}^d - \boldsymbol{x}\right\|^2\right]}{N} \\ \cdots \leq \frac{1}{{A^s}^2} \sum\nolimits_i \frac{\sigma_i^2}{A_i} \leq \frac{L\sigma_{\max}^2}{A_{\min} {A^s}^2}. \end{aligned} \tag{3}$$

Frame reconstruction with our distributed processing constraint has the power to reduce noise in the reconstructed signal by an amount that depends on the minimum redundancy of both the frame of subspaces, and the individual local frames that span the subspaces. It is interesting to consider how much of a penalty in noise reduction a distributed processing scheme incurs compared to centralized processing. Using the properties of fusion frames, considering the collection of local frames together yields a global frame for $\mathcal{H}$ with lower and upper bounds $A \geq A^s A_{\min}$ and $B \leq B^s B_{\max}$. From there we can determine [10] that a centralized reconstruction of $\boldsymbol{x}$ using the global frame directly, $\widehat{\boldsymbol{x}}^c$, would yield an upper bound on the MSE per signal dimension of $\frac{\mathcal{E}\left[\|\widehat{\boldsymbol{x}}^c-\boldsymbol{x}\|^2\right]}{N} \leq \frac{\sigma^2_{\max}}{A} \leq \frac{\sigma^2_{\max}}{A^s A_{\min}}$. Comparing this bound to the MSE upper bound in (3) when using the distributed scheme, we see that (because $L \geq A^s$) the upper bound using the centralized approach is better than the distributed reconstruction by a factor of $\frac{L}{A^s}$. While these are not tight bounds on the error, they hint at the potential for the distributed reconstruction to perform worse than the centralized scheme and give a bound on the performance reduction. Finally, we have also shown that there will be no noise-reduction penalty for distributed reconstruction when all of the local frames are tight frames with the same number of vectors and the same frame bounds, and the fusion frame is also tight. Though the conditions proposed here for equal noise reduction may seem restrictive, a result from [11] regarding random frames indicates that frames will become tight asymptotically as more random vectors are added. Therefore, systems where random vectors are randomly assigned to a local subspace will asymptotically meet the conditions for achieving the optimal (centralized) noise reduction.

As a further note, we have also used frame theory to model the distributed interactions necessary to realize action directly from sensed data in a distributed system. In the setting of wireless sensor and actuator networks, the analytic tools of frame theory have pointed the way to efficient [12] (and sometimes optimal [13]) methods for reducing the communication between sensors and actuators in such a system.

IV. Biological sensing: sensory neural systems

Despite not knowing the inner workings of sensory neural systems, redundancy clearly plays a crucial role in representing information. Anatomical observations indicate that shortly after the transduction front-end, sensory systems process data in successive stages each composed of a neural population [14]. Examples from both the early visual pathway [15] and the early auditory pathway [16] indicate that these populations are highly redundant, using many more neurons than the previous stage. This observation raises the question "Why do neural systems use so much redundancy?"

One possible answer to this question is that neural systems use the redundancy provided by a frame to leverage the flexibility possible in the encoding. In particular, several recent results indicate that neurons in primary visual cortex may try to sparsify their responses [15], [17], [18]. Using this flexibility to achieve sparse codes might offer many advantages to sensory neural systems, including enhancing the performance of subsequent processing stages, increasing the storage capacity in associative memories, and increasing the energy efficiency of the system [18]. However, sparse approximation algorithms from the signal processing community, such as Matching Pursuit (MP) [8] and Basis Pursuit De-Noising (BPDN) [6], do not have implementations that correspond both naturally and efficiently to the distributed parallel computational architectures used by neural systems [19].

In a search for a distributed processing algorithm that can calculate sparse codes in a frame, we have recently developed a system called locally competitive algorithms (LCAs) [19]. Our LCAs associate each neuron with an element of the dictionary $\boldsymbol{\phi}_k \in \mathcal{D}$. The system is presented with a (possibly time-varying) input image $\boldsymbol{x}(t)$. The collection of nodes evolve according to fixed dynamics (described below) and settle on a collective output $\{a_k(t)\}$, corresponding to the short-term average firing rate of the neurons. The goal is to converge to coefficients that have few non-zero values while approximately reconstructing the input, $\widehat{\boldsymbol{x}}\,(t) = \sum_k a_k(t)\,\boldsymbol{\phi}_k \approx \boldsymbol{x}(t)$.

The LCA dynamics follow several properties observed in neural systems: inputs cause an internal state to "charge up" like a leaky integrator; states over a threshold produce non-trivial outputs; and these super-threshold outputs inhibit neighboring units through lateral connections. We represent each unit's sub-threshold value by a time-varying *internal state* $u_k(t)$. When u_k becomes significantly large, the node becomes "active" and produces an output signal a_k used to represent the stimulus and inhibit other nodes. This output coefficient is the result of an activation function applied to the internal state, $a_k = T_\lambda(u_k)$, parameterized by the system threshold λ. We primarily consider activation functions that operate as thresholding devices — they are essentially zero for values below λ and essentially linear for values above λ.

The internal states evolve according to the dynamical system equation

$$\dot{u}_k(t) = \frac{1}{\tau}\left[m_k(t) - u_k(t) - \sum_{n\neq k} \mathbf{G}_{k,n} a_n(t) \right], \quad (4)$$

where τ is the membrane time-constant, $m_k(t) = \langle \boldsymbol{\phi}_k, \boldsymbol{x}(t) \rangle$ is the unit's excitatory input current, and $\mathbf{G}_{k,n} = \langle \boldsymbol{\phi}_k, \boldsymbol{\phi}_n \rangle$ is the inner product between the node receptive fields. The nodes best matching the stimulus will have internal state variables that charge at the fastest rates. These nodes will cross threshold and become "active" first, inducing an inhibition signal to every other node. The possibility of unidirectional inhibition gives strong nodes a chance to prevent weaker nodes from becoming active and initiating counter-inhibition, thus making the search for a sparse solution more efficient.

In [19] we show that the LCA architecture described by (4) solves a family of sparse approximation problems by descending an energy function that combines reconstruction MSE and a sparsity-inducing cost $C(\cdot)$,

$$E(t) = \frac{1}{2} \left\| \boldsymbol{x}(t) - \widehat{\boldsymbol{x}}\,(t) \right\|^2 + \lambda \sum_k C(a_k(t))\,.$$

The specific form of the cost function $C(\cdot)$ is determined by the form of the thresholding activation function $T_\lambda(\cdot)$. For a given threshold function, the cost function is specified (up to a constant) by

$$\lambda \frac{dC(a_k)}{da_k} = u_k - a_k = u_k - T_\lambda(u_k). \quad (5)$$

This correspondence between the thresholding function and the cost function can be seen by computing the derivative of E with respect to the active coefficients, $\{a_k\}$. If (5) holds, then letting the internal states $\{u_k\}$ evolve according to $\dot{u}_k \propto -\frac{\partial E}{\partial a_k}$ yields the equation for the internal state dynamics in (4). As long as a_k and u_k are related by a monotonically increasing function, the coefficients will also descend the energy function E. Note that the system is not performing direct gradient descent because the dynamics on u_k correspond to the gradient of E with respect to a_k. The system is performing a gradient descent that has been warped by $T_\lambda(\cdot)$. This warping allows the internal state variables to determine an effective direction even when they are near zero (and the gradient of energy function looks flat) and it favors nodes that are already active.

We are most interested in thresholding functions

$$T_{(\alpha,\gamma,\lambda)}(u_k) = \begin{cases} u_k - (\alpha\lambda)\cdot \mathrm{sign}(u_k) & \text{if} \quad |u_k| \geq \lambda \\ 0 & \text{if} \quad |u_k| < \lambda \end{cases} \quad (6)$$

called a "hard" threshold when $\alpha = 0$ and "soft" threshold when $\alpha = 1$. By using smooth analytic functions and taking the limit, we can calculate the corresponding cost functions in the special case of hard and soft thresholding [19]. The soft-thresholding locally competitive algorithm (SLCA) applies the ℓ^1 norm as a cost function on the active coefficients, $C_{(1,\infty,\lambda)}(a_k) = |a_k|$. Thus, the SLCA is solving the BPDN optimization in. Alternately, the hard-thresholding locally competitive algorithm (HLCA) applies an ℓ^0 -like cost function by using a constant penalty regardless of magnitude, $C_{(0,\infty,\lambda)}(a_k) = \frac{\lambda}{2}\,\mathbf{1}_{(|a_k|>\lambda)}$, where $\mathbf{1}_{(\cdot)}$ is the indicator function evaluating to 1 if the argument is true and 0 if the argument is false. Because the ℓ^0 -penalty is non-convex, the HLCA will only find a local minimum of its desired objective function. Many sparse approximation methods have been reported recently, and several of these algorithms are closely related to discrete approximations of our LCAs. In particular, the family of "iterative thresholding algorithms" [20]–[24] can be viewed as discrete approximations to the continuous-time LCAs.

Numerical experiments have shown that the HLCA produces ℓ^0 sparsity-MSE tradeoffs comparable to MP. Furthermore, SLCA produces coefficients with significant ℓ^0 sparsity (comparable to an oracle threshold applied to the results of an interior point BPDN solver). Note that SLCA keeps many coefficients zero throughout the calculation and has no need for an oracle threshold. This sparsity-MSE tradeoff is demonstrated in Figure 1. The LCAs have also demonstrated extremely efficient encodings for time-varying signals (e.g., video sequences), as discussed thoroughly in [19].

V. Conclusions

We have seen two scenarios where frame theory has provided an analytic framework to understand the information processing in a sensing system with distributed constraints. In the case of wireless sensor networks, we see a man-made system where we can characterize how the noise robustness is affected by a requirement to reconstruct the sensed data in a particular way. In the case of neural systems performing sparse approximation, we see a biological system performing information processing that can be modeled as a dynamical system used to calculate coefficients in a frame expansion. In

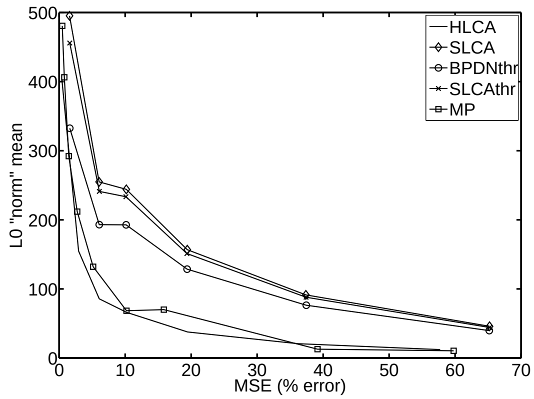

Fig. 1. Sparsity vs. MSE tradeoff calculated on 32×32 bandpassed image patches in a dictionary that was four-times overcomplete. HLCA and SLCA represent the hard and soft thresholding locally competitive algorithms. BPDNthr represents an interior point BPDN solver with an oracle threshold applied. SCLAthr represents the SLCA system with the same threshold applied.

both cases, the notion of an overcomplete basis was a powerful tool to analytically characterize the system. While frames have found great utility in modern signal processing applications, this provides strong evidence that they can also be effectively used as models for many common architectures found in sensing systems.

ACKNOWLEDGMENT

I am grateful to Don H. Johnson, Richard Baraniuk and Bruno Olshausen for collaborating on the research described in this overview. The original work highlighted here appeared previously in [10], [19].

REFERENCES

[1] C. Shannon, “A mathematical theory of communication,” *Bell System Technical Journal*, vol. 27, pp. 379–423, 623–656, 1948.

[2] O. Christensen, *An Introduction to Frames and Riesz Bases*. Boston, MA: Birkhauser, 2002.

[3] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA: Society for Industrial and Applied Mathematics, 1992.

[4] R. Duffin and A. Schaeffer, “A class of nonharmonic Fourier series,” *Transactions of the American Mathematical Society*, vol. 72, no. 2, pp. 341–366, March 1952.

[5] P. Casazza, G. Kutyniok, and S. Li, “Fusion frames and distributed processing,” *Preprint*, 2006.

[6] S. Chen, D. Donoho, and M. Saunders, “Atomic decomposition by basis pursuit,” *SIAM Journal on Scientific Computing*, vol. 43, no. 1, pp. 129–159, 2001.

[7] J. Tropp, “Greed is good: Algorithmic results for sparse approximation,” *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2231–2242, 2004.

[8] S. Mallat and Z. Zhang, “Matching pursuits with time-frequency dictionaries,” *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, December 1993.

[9] P. Casazza, G. Kutyniok, S. Li, and C. Rozell, “Modeling sensor networks with fusion frames,” in *Proceedings of SPIE, Wavelets XII at SPIE Optics and Photonics*, vol. 6701, San Diego, CA, August 2007.

[10] C. Rozell and D. Johnson, “Analysis of noise reduction in redundant expansions under distributed processing requirements,” in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Philadelphia, PA, March 2005.

[11] V. Goyal, J. Kovačević, and J. Kelner, “Quantized frame expansions with erasures,” *Applied and Computational Harmonic Analysis*, vol. 10, pp. 203–233, 2001.

[12] C. Rozell and D. Johnson, “Evaluating local contributions to global performance in wireless sensor and actuator networks,” *Lecture Notes in Computer Science*, vol. 4026, pp. 1–16, 2006, *Proceedings of the International Conference on Distributed Computing in Sensor Systems (DCOSS)*, San Francisco, CA, June 2006.

[13] ——, “Power scheduling for wireless sensor and actuator networks,” in *Proceedings of the International Conference on Information Processing in Sensor Networks (IPSN)*, Cambridge, MA, April 2007.

[14] E. Kandel, J. Schwartz, and T. Jessell, *Principles of Neural Science*, 3rd ed. Norwalk, CT: Appleton & Lange, 1991.

[15] B. Olshausen, “Principles of image representation in visual cortex,” in *The Visual Neurosciences*, L. Chalupa and J. Werner, Eds. MIT Press, 2003, pp. 1603–1615.

[16] A. Popper and R. Fay, *The Mammalian Auditory Pathway: Neurophysiology*. New York: Springer-Verlag, 1992.

[17] W. Vinje and J. Gallant, “Natural stimulation of the nonclassical receptive field increases information transmission efficiency in V1,” *Journal of Neuroscience*, vol. 22, pp. 2904–2915, 2002.

[18] B. Olshausen and D. Field, “Sparse coding of sensory inputs,” *Current Opinion in Neurobiology*, vol. 14, pp. 481–487, 2004.

[19] C. Rozell, D. Johnson, R. Baraniuk, and B. Olshausen, “Sparse coding via thresholding and local competition in neural circuits,” *Neural Computation*, 2008, in press.

[20] K. Herrity, A. Gilbert, and J. Tropp, “Sparse approximation via iterative thresholding,” in *Proceedings of the 2006 IEEE International Conference on Acoustics, Speech, and Signal Processing*, Toulouse, France, May 2006.

[21] E. Hale, W. Yin, and Y. Zhang, “A fixed-point continuation method for l1-regularized minimization with applications to compressed sensing,” Department of Computational and Applied Mathematics, Rice University, Tech. Rep. TR07-07, 2007.

[22] M. Figueiredo and R. Nowak, “An EM algorithm for wavelet-based image restoration,” *IEEE Transactions on Image Processing*, vol. 12, pp. 906–916, 2003.

[23] I. Daubechies, M. M. Defrise, and C. De Mol, “An iterative thresholding algorithm for linear inverse problems with a sparsity constraint,” *Communications on Pure and Applied Mathematics*, vol. 57, no. 11, pp. 1413–1457, 2004.

[24] T. Blumensath and M. Davies, “Iterative thresholding for sparse approximations,” *The Journal of Fourier Analysis and Applications*, 2008, in Press.