Competitive optimization of compressed sensing

2007 J. Phys. A: Math. Theor. 40 F299

(http://iopscience.iop.org/1751-8121/40/16/F02)

IOP FTC
IOP PUBLISHING JOURNAL OF PHYSICS A: MATHEMATICAL AND THEORETICAL

J. Phys. A: Math. Theor. **40** (2007) F299–F305 doi:10.1088/1751-8113/40/16/F02

FAST TRACK COMMUNICATION

Competitive optimization of compressed sensing

M Andrecut, R A Este and S A Kauffman

Institute for Biocomplexity and Informatics, University of Calgary, 2500 University Drive NW, Calgary, Alberta, T2N 1N4, Canada

Received 27 December 2006, in final form 8 March 2007
Published 30 March 2007
Online at stacks.iop.org/JPhysA/40/F299

Abstract
The compressed sensing framework enables the recovery of a sparse signal from a small number of projections onto random vectors. Here, we propose a new extension method of compressed sensing, applied to signals sparse in the time domain. The method consists, in a competitive ensemble, of compressed sensing devices. We show that this approach leads to a substantial improvement in performance. Also, we propose a simplified version of compressed sensing, in which the random projections are replaced by cyclic correlations with a single random vector. Encouraged by the results obtained with the competitive ensemble approach, we show that the performance can be even more improved by employing a simulated annealing algorithm.

PACS numbers: 02.70.Rr, 05.10.−a, 02.60.Pn, 02.30.Zz, 07.05.Hd

(Some figures in this article are in colour only in the electronic version)

The recent evolution of sensing devices has triggered new research in the fields of compression and coding, in order to minimize storage, communication and bandwidth consumption. Current state-of-the-art compression algorithms exploit the fact that many signals have sparse representation in terms of some basis, meaning that a small number of coefficients can be transmitted or stored rather than the full signal. However, the complexity and computational cost of these algorithms limit the range of applications for many classes of sensing devices. Fortunately, the recently introduced framework of compressed sensing (CS) enables a reduction in the complexity and the computational cost [1–5]. The main idea of CS is that if the signal is compressible, then a small number of projections on random vectors contain sufficient information for approximate or exact reconstruction. This has promising implications for applications involving sparse signal acquisition, pattern detection and recognition. Also, CS strategy could potentially lead to interesting models for various sensing operations performed at the biological level.

The most remarkable aspect of CS is the role played by randomization. Here, we show how we can exploit even more this intriguing characteristic, and we propose a new method of CS for sparse discrete signals in the time domain. The method consists in a competitive ensemble of CS devices. A similar strategy has been successfully used before in other areas

such as neural networks and machine learning, where it has been shown that competitive models could give better results (see, e.g., committee networks, boosting) [6]. We show that this approach leads to a substantial improvement in the compression rate, error and probability of reconstruction of the CS method. Also, we propose a simplified version of CS, in which the random projections are replaced by cyclic correlations with a single random vector. This method is more easily implementable, and it is well suited for applications in which an approximate reconstruction of the original signal is still acceptable. Inspired by the results obtained with the competitive ensemble approach, we show that a major improvement in the performance of CS can be obtained by employing a simulated annealing algorithm.

Let us give a short description of the CS method. Assume that we acquire an N-sample signal $s \in \mathbf{R}^N$ and $\Psi = \{\psi_n | \psi_n \in \mathbf{R}^N, n = 0, \ldots, N-1\}$ is a basis of vectors spanning $\mathbf{R}^N$. We say that Ψ provides a sparse representation of s if s is well approximated by a linear combination of a small set of vectors from Ψ. That is, there exists a set of indices $\{n_1, \ldots, n_K\} \subset \{0, \ldots, N-1\}$ for small $K \ll N$ such that

$$s = \sum_{i=1}^{K} x_{n_i} \psi_{n_i}. \tag{1}$$

We say that s is K-sparse in Ψ and we call Ψ the sparse basis. Let $\Phi = \{\varphi_m | \varphi_m \in \mathbf{R}^N, m = 0, \ldots, M-1\}$ be a set of measurement vectors, where $M \ll N$. We use these vectors to collect M measurements of the signal:

$$y_m = \langle s, \varphi_m \rangle = \sum_{n=0}^{N-1} s_n \varphi_{nm}, \qquad m = 0, \ldots, M-1, \tag{2}$$

where $\langle \cdot, \cdot \rangle$ denotes the usual inner product. The CS theory states that it is possible to construct a set of vectors Φ, such that the measurements y preserve the essential information about the K-sparse signal s. Thus, the signal s can be reconstructed from the measurements y. The quantity $\rho = N/M$ corresponds to the compression ratio. The CS theory aims to answer three distinct questions: (i) what conditions must be satisfied by the measurement vectors Φ? (ii) how many measurements are necessary to reconstruct the signal? (iii) given these measurements, what algorithms can perform the reconstruction task?

If the two sets, Ψ and Φ, satisfy the *incoherence condition*, then it is indeed possible to recover the signal s from the measurements y [1–5]. The incoherence condition requires that the original basis set Ψ does not provide a sparse representation of the measurement basis set Φ. A measure of mutual incoherence of the two bases is given by [7]

$$\mu(\Psi, \Phi) = \max_{\psi_n \in \Psi, \varphi_m \in \Phi} |\langle \psi_n, \varphi_m \rangle|. \tag{3}$$

If the two bases are normalized, then the mutual incoherence always obeys

$$N^{-1/2} \leqslant \mu(\Psi, \Phi) \leqslant 1, \tag{4}$$

and measures how the two bases look alike. The smaller the incoherence, the stronger the recovery. This incoherence property holds for many pairs of bases, in particular a random measurement basis tends to be incoherent with any fixed basis. The two obvious distributions for the measurement basis set Φ are: (1) Gaussian ($\varphi_{nm} \in N(0, 1)$) and (2) Bernoulli ($\varphi_{nm} = \pm 1$ with equal probability). Indeed, either one of these distributions (normalized for convenience) can be used to collect measurements [1–5]. Because of the incoherence between Ψ and Φ, if the signal s is sparse in the original basis Ψ then no other set of sparse set coefficients can exist to explain the measurements y, performed in the basis Φ. According to the CS theory [1–5], the number of measurements M must have the size

$$M = \|y\|_0 \geqslant cK, \tag{5}$$

where c is dependent on the two bases $c = c(\Psi, \Phi)$. Several algorithms have been proposed for recovering s from the measurements y, each of them requiring a different constant c. Searching for the sparsest set of coefficients $\widehat{x}$ in the basis Ψ that matches the sensed values y leads to the l_0 optimization problem

$$\widehat{x} = \arg\min_x \|x\|_0 \qquad \text{subject to} \quad \Phi\Psi x = y. \tag{6}$$

This combinatorial optimization problem is NP-hard to solve and the convexification of the objective function is introduced by replacing the l_0 norm with the l_1 norm [1–3]. The resulting optimization problem is known as basis pursuit (BP) and it can be solved using linear programming techniques whose computational complexities are polynomial in N [7]. This problem requires $c \approx \log_2(1 + N/K)$ but still suffers from high computational complexity. At the expectation of slightly more measurements, iterative greedy algorithms such as matching pursuit (MP) [9] and orthogonal matching pursuit (OMP) [10] can also be used for the recovery problem. MP has been proven to achieve an accurate decomposition of the signal and it provides a low-complexity alternative to BP, but requires an unbounded number of iterations for convergence. OMP converges in a fixed number of iterations but requires the added complexity of the orthogonalization at each step. Our choice is MP because it is easily implementable in software and hardware. Given the measurements y, we solve for the sparsest approximation of y in the *dictionary* defined by the columns of

$$\Gamma = \Phi\Psi = [\gamma_1|\cdots|\gamma_N]. \tag{7}$$

MP is an iterative greedy algorithm which has the following pseudo-code:

(i) Initialize the dictionary $\Gamma \leftarrow \Phi\Psi$, the residual $r \leftarrow y \leftarrow \Phi s$, the approximation $\widehat{x} \leftarrow 0$ and set $t \leftarrow 1$.

(ii) Select the dictionary vector that maximizes the projection of the residual:

$$n \leftarrow \arg\max_{n\in\{0,\ldots,N-1\}} |\langle r, \gamma_n\rangle|/\|\gamma_n\|_2. \tag{8}$$

(iii) Update the estimate of the corresponding coefficient and the residual:

$$\widehat{x}_n \leftarrow \widehat{x}_n + \langle r, \gamma_n\rangle/\|\gamma_n\|_2^2, \qquad r \leftarrow r + \langle r, \gamma_n\rangle\gamma_n/\|\gamma_n\|_2^2. \tag{9}$$

(iv) If $\|r\|_2 > \varepsilon\|y\|_2$ and $t < MN$ then set $t \leftarrow t + 1$ and go to step (ii); otherwise terminate and return $\widehat{x}$.

The convergence criterion is the minimum proportion ε of y that can be left in the residual r. If the convergence is not achieved in MN steps then the algorithm stops. The reconstructed signal is then obtained as

$$\widehat{s} = \sum_{n=0}^{N-1} \widehat{x}_n \psi_n. \tag{10}$$

Let us now consider the case in which we have an ensemble of J CS devices, each of them characterized by its own dictionary:

$$\{\Gamma^{(j)}|\Gamma^{(j)} = \Phi^{(j)}\Psi,\ j = 0, \ldots, J-1\}. \tag{11}$$

All these CS devices have the same original sparse basis Ψ but they have different random measurement bases $\Phi^{(j)}$. Also, we assume that all of them are processing in a competitive way the same signal s. The competition consists in choosing the CS device which gives the best reconstruction in MN iteration steps of the MP algorithm. The algorithm returns the best

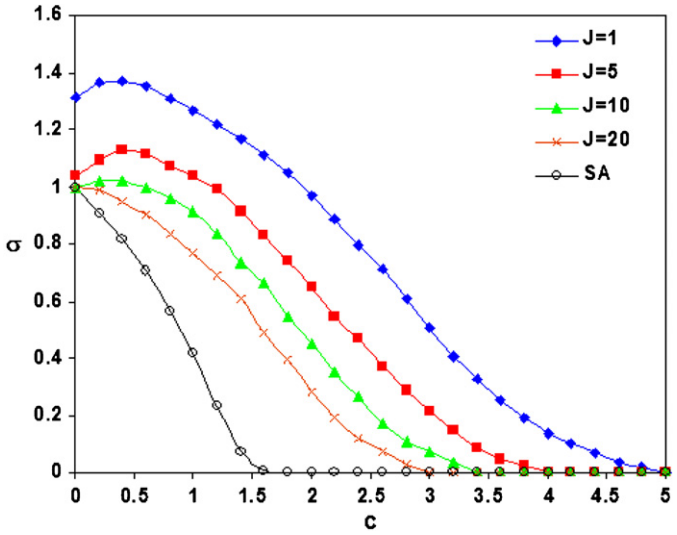

Figure 1. Competitive CS: the recovery error as a function of the overmeasuring parameter (c) (J = the number of competing CS devices, SA = simulated annealing).

approximation $\widehat{x}^{(k)}$ of the sparse coefficients and the index k of the corresponding CS device. This way, the signal can be recovered anytime by the specified CS device, k:

$$k = \arg \min_{j\in\{0,\ldots,J-1\}} \|\widehat{s}^{(j)} - s\|_2, \qquad \widehat{s}^{(k)} = \sum_{n=0}^{N-1} \widehat{x}_n^{(k)} \psi_n. \tag{12}$$

One may think that because the measurement basis set $\Phi^{(j)}$ is already random (for each CS device) one cannot improve the performance obtained by one CS device. However, our numerical experiments have shown that even for small values of J one already obtains a substantial improvement for the compression rate, error and probability of reconstruction.

In figure 1, we give the numerical results obtained for sparse discrete signals in the time domain. The length of the signals is $N = 100$ and the original sparse basis is the Dirac basis:

$$\Psi = \{\psi_n | \psi_{mn} = \delta_{mn}, n, m = 0, \ldots, N-1\}, \tag{13}$$
$$\delta_{mn} = \begin{cases} 1 & \text{if } m = n \\ 0 & \text{if } m \neq n. \end{cases}$$

The measurement bases $\Phi^{(j)}$ correspond to randomly drawn Bernoulli sets. The sparsity of the signal is fixed to $K = 10$, and the non-zero coefficients x are randomly generated uniformly in the interval $[-1, 1]$. As a measure of the performance of the algorithm we have used the relative error between the original signal s and the reconstructed approximation $\widehat{s}$:

$$\sigma = \|s - \widehat{s}\|_2 \|s\|_2^{-1}. \tag{14}$$

We let the number of measurements M to take values from 0 to N, and for each M we perform 10^2 independent signal acquisitions. One can see that for $J = 1$ (only one CS device) the signal is perfectly recovered ($\sigma \approx 0$) only for an overmeasuring parameter $c \geqslant 5$. By simply increasing J to $J = 5$ one obtains a substantial improvement in the error and the overmeasuring parameter, which becomes $c \geqslant 4$. By increasing J even more, one can obtain a higher rate of compression and a smaller error in the reconstruction (for $J = 20$ we already

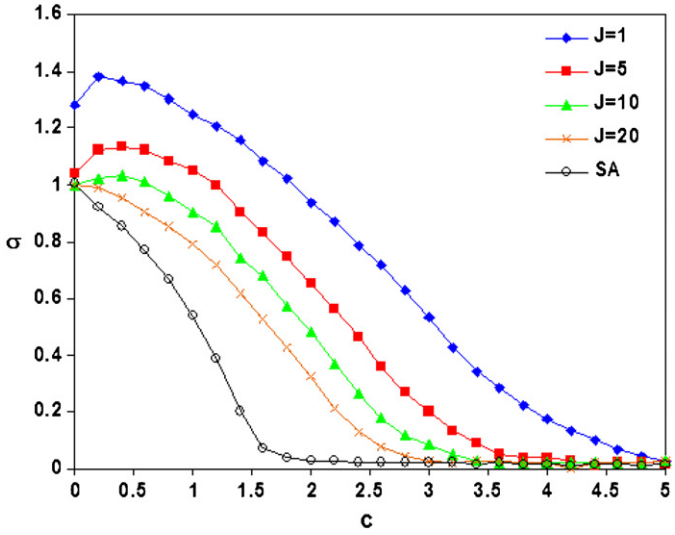

Figure 2. Simplified competitive CS: the recovery error as a function of the overmeasuring parameter (c) (J = the number of competing CS devices, SA = simulated annealing).

have $c \geqslant 3$). This competitive approach can be easily implemented in parallel software or hardware, by simply defining a separate thread for each CS device.

Now, let us consider the case in which the M random projections are performed using only one Bernoulli vector $\xi = (\xi_0, \ldots, \xi_{N-1}), \xi_n = \pm 1$, with equal probability. These measurements are defined by the cyclic correlations between ξ and the signal s:

$$y_m = s * \xi = \sum_{n=0}^{N-1} s_n \xi_{(n+m) \bmod(N)}, \qquad m = 0, \ldots, M-1. \tag{15}$$

So, instead of having M measurements defined by M independent random vectors, we have M measurements defined by only one vector. In this case, the measurement operator can be stored and applied more efficiently. For example, the M measurements require MN elementary operations. However, according to the correlation theorem [11], this number can be made independent of M by performing the measurements in the Fourier space, with a constant cost of $N \log N$ (using the fast Fourier transform, F):

$$y = s * \xi = F^{-1}(F(s)F(\xi)). \tag{16}$$

Also, if the original sparse basis Ψ is the Dirac basis, then the dictionary matrix $\Gamma = \Phi\Psi$ becomes a circulant matrix, because each row is a copy of ξ shifted right by m places. We use the same competitive MP algorithm for signal reconstruction and the same parameters. To our surprise, the reconstruction works almost as well as for the full matrix method described before. The only difference is that now the reconstruction is not perfect and a small error of $\sigma \approx 1$–2% is always present (figure 2). It is important to note that this error decreases if the length of the signal increases. This is due to the fact that the length of the random vector ξ increases also, and therefore the measurements becomes more reliable.

The results obtained suggest that a simulated annealing (SA) algorithm might perform even better than the simple competitive ensemble. SA is a generic probabilistic meta-algorithm for the global optimization problem, namely locating a good approximation to the global

optimum of a given function in a large search space [12, 13]. Each step of the SA algorithm replaces the current solution by a random solution, chosen with a probability that depends on the difference between the corresponding function values and on a global parameter θ (called the temperature), that is gradually decreased during the process. The dependency is such that the current solution changes almost randomly when θ is large, but increasingly downhill as θ goes to zero. The allowance for uphill moves saves the method from becoming stuck at local minima which are the bane of greedier methods. In order to apply the SA method to a specific problem, one must specify: (i) the state space; (ii) the candidate solution selection method (which enumerates the candidates for the next state); (iii) the annealing schedule; (iv) the probability transition function. These choices can have a significant impact on the method's effectiveness.

In our case, we observe that the CS method is based on pseudo-random number generation, and therefore we consider that the state space corresponds to the values taken by the seed of the random number generator. Therefore, in order to compress a signal and to reconstruct it, in the CS framework, one actually needs to know only the seed used to generate the measurement basis. We perform a random search on the space of seed values α by selecting a different seed each time we generate a new candidate measurement basis. If the new basis, generated with α', improves the performance of the MP algorithm then this seed becomes the best seed, for the given signal. The initial temperature must be large enough to make the uphill and downhill transition probabilities about the same. The temperature must then decrease so that it is nearly zero, at the end of the allotted time. A popular choice is the exponential schedule, where the temperature decreases by a fixed factor λ at each step. If θ_i and θ_f are the initial and, respectively, the final temperature then $\lambda = (\theta_f/\theta_i)^{1/T}$, where T is the final number of iteration steps. The function to be optimized is given by the norm $\|s - \widehat{s}\|_2$, where s is the given signal and $\widehat{s}$ is its approximation in the current measurement basis. The probability of making the transition from the current state α to a candidate new state α' is given by the Boltzmann distribution:

$$P(\alpha, \alpha', \theta) = \begin{cases} 1 & \text{if} \quad \|s - \widehat{s}'\|_2 \leqslant \|s - \widehat{s}\|_2 \\ \exp(-(\|s - \widehat{s}'\|_2 - \|s - \widehat{s}\|_2)/\theta) & \text{if} \quad \|s - \widehat{s}'\|_2 > \|s - \widehat{s}\|_2. \end{cases} \tag{17}$$

The pseudo-code of the SA method can be formulated as following:

(i) Initialize s; α; Ψ; T; θ_i; θ_f; $\lambda \leftarrow (\theta_f/\theta_i)^{1/T}$; $\theta \leftarrow \theta_i$; $\widehat{s} \leftarrow 0$.
(ii) Randomly select the new seed α'.
(iii) Generate a new measurement basis Φ.
(iv) Calculate the new dictionary $\Gamma \leftarrow \Phi\Psi$.
(v) Find the MP solution $\widehat{x} \leftarrow \mathrm{MP}(s, \Gamma)$.
(vi) Approximate the signal $\widehat{s}' \leftarrow \sum_{n=0}^{N-1} \widehat{x}_n \psi_n$.
(vii) Calculate the transition probability $P(\alpha, \alpha', \theta)$.
(viii) If $\mathrm{rnd}(1) < P(\alpha, \alpha', \theta)$ then $\alpha \leftarrow \alpha'$ and $\widehat{s} \leftarrow \widehat{s}'$.
(ix) Update the temperature $\theta \leftarrow \lambda\theta$.
(x) If $\theta < \theta_f$ then go to step (ii); otherwise terminate and return α and s.

Here, the function rnd(1) returns a uniform random number in [0, 1].

We have implemented the above algorithm for both cases, corresponding to the full matrix CS and the simplified version of CS (when the measurements are performed using the cyclic correlations with a single random vector). In both cases, the sparse basis is the Dirac basis. The parameter values are: $N = 100$, $K = 10$, $\theta_i = 1$, $\theta_f = 10^{-5}$, $T = 10^2$. One can see that the SA method provides a major improvement in the compression rate, error and probability of reconstruction. The overmeasuring parameter for the SA method is only $c \approx 1.5$, which is 3.33 better than for the original CS method.

In conclusion, we have presented a numerical investigation of a new extension method of the CS framework. The method consists in a competitive ensemble of CS devices. We have shown that this approach leads to a substantial improvement in the performance of CS applied to signals sparse in the time domain (13). We have proposed a simplified version of CS, in which the independent random projections are replaced by cyclic correlations with a single random vector. This method is more easily implementable in parallel software or hardware, and it is well suited for applications in which an approximate reconstruction of the original signal is still acceptable. Also, we have shown that the simulated annealing method can improve even more the performance of CS.

Acknowledgment

This research was supported by iCORE.

References

[1] Donoho D 2006 Compressed sensing *IEEE Trans. Inform. Theory* **52** 1289–306
[2] Candès E, Romberg J and Tao T 2006 Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information *IEEE Trans. Inform. Theory* **52** 489
[3] Candès E and Tao T 2006 Near optimal signal recovery from random projections: universal encoding strategies? *IEEE Trans. Inform. Theory* **52** 5406
[4] Donoho D and Tsaig Y 2006 Extensions of compressed sensing *Signal Process.* **86** 533
[5] Candès E and Romberg J 2005 Signal recovery from random projections *Computational Imaging III Proc. SPIE* 5674 ed C A Bouman and E L Miller 76–86
[6] Kittler J and Roli F (ed) 2000 *Proc. 1st Int. Workshop on Multiple Classifier Systems (Lecture Notes in Computer Science vol 1857)* (London: Springer)
[7] Candès E J and Romberg J 2004 Quantitative robust uncertainty principles and optimally sparse decompositions *Found. Comput. Math.* **6** 227
[8] Chen S S, Donoho D L and Saunders M A 1999 Atomic decomposition by basis pursuit *SIAM J. Sci. Comput.* **20** 1336
[9] Mallat S and Zhang Z 1993 Matching pursuits with time–frequency dictionaries *IEEE Trans. Signal Process.* **41** 3397
[10] Davis G, Mallat S and Avellaneda M 1997 Greedy adaptive approximation *J. Constr. Approx.* **13** 5798
[11] Brigham E O 1988 *The Fast Fourier Transform and Applications* (Englewood Cliffs, NJ: Prentice-Hall)
[12] Kirkpatrick S, Gelatt C D Jr and Vecchi M P 1983 Optimization by simulated annealing *Science* **220** 671
[13] Reeves C R (ed) 1995 *Modern Heuristic Techniques for Combinatorial Problems* (London: McGraw-Hill)