

Contents lists available at [ScienceDirect](http://ScienceDirect.com)

## Digital Signal Processing

[www.elsevier.com/locate/dsp](http://www.elsevier.com/locate/dsp)

## Compressive sensing using the modified entropy functional

Kivanc Kose<sup>a,b,\*</sup>, Osman Gunay<sup>a</sup>, A. Enis Cetin<sup>a</sup><sup>a</sup> Electrical and Electronics Engineering Department, Bilkent University, Turkey<sup>b</sup> Dermatology Service, Memorial Sloan-Kettering Cancer Center, USA

## ARTICLE INFO

## Article history:

Available online 2 October 2013

## Keywords:

Compressive sensing  
Modified entropy functional  
Projection onto convex sets  
Iterative row-action methods  
Bregman-projection  
Proximal splitting

## ABSTRACT

In most compressive sensing problems,  $\ell_1$  norm is used during the signal reconstruction process. In this article, a modified version of the entropy functional is proposed to approximate the  $\ell_1$  norm. The proposed modified version of the entropy functional is continuous, differentiable and convex. Therefore, it is possible to construct globally convergent iterative algorithms using Bregman's row-action method for compressive sensing applications. Simulation examples with both 1D signals and images are presented.

© 2013 Elsevier Inc. All rights reserved.

## 1. Introduction

The Nyquist–Shannon sampling theorem [1] is one of the fundamental theorems in signal processing literature. It specifies the conditions for perfect reconstruction of a continuous signal from its samples. If a signal is sampled with a sampling frequency that is at least two times larger than its bandwidth, it can be perfectly reconstructed from its samples. However in many applications of signal processing including waveform compression, perfect reconstruction is not necessary. In this article, a modified version of the entropy functional is proposed. The functional is defined for both positive and negative real numbers and it is continuous, differentiable and convex everywhere. Therefore it can be used as a cost function in many signal processing problems including the compressive sensing problem.

The most common method used in compression applications is transform coding. The signal  $\mathbf{x}[n]$  is transformed into another domain defined by the transformation matrix  $\psi$ . The transformation procedure is simply finding the inner product of the signal  $\mathbf{x}[n]$  with the rows  $\psi_i$  of the transformation matrix  $\psi$  represented as follows:

$$s_l = \langle \mathbf{x}, \psi_l \rangle, \quad l = 1, 2, \dots, N, \quad (1)$$

where  $\mathbf{x}$  is a column vector, whose entries are samples of the signal  $\mathbf{x}[n]$ .

The digital signal  $\mathbf{x}[n]$  can be reconstructed from its transform coefficients  $s_l$  as follows:

$$\mathbf{x} = \sum_{l=1}^N s_l \psi_l \quad \text{or} \quad \mathbf{x} = \psi \mathbf{s}, \quad (2)$$

where  $\mathbf{s}$  is a vector containing the transform domain coefficients  $s_l$ .

The basic idea in digital waveform coding is that the signal should be approximately reconstructed from only a few of its non-zero transform coefficients. In most cases, including the JPEG image coding standard, the transform matrix  $\psi$  is chosen in such a way that the new signal  $\mathbf{s}$  is efficiently represented in the transform domain with a small number of coefficients. A signal  $\mathbf{x}$  is compressible, if it has only a few large amplitude  $s_l$  coefficients in the transform domain and the rest of the coefficients are either zeros or negligibly small-valued.

In a compressive sensing framework, the signal is assumed to be  $K$ -sparse in a transformation domain, such as the wavelet domain or the DCT (Discrete Cosine Transform) domain. A signal with length  $N$  is  $K$ -sparse if it has at most  $K$  non-zero and  $(N - K)$  zero coefficients in a transform domain. The case of interest in CS problems is when  $K \ll N$ , i.e., sparse in the transform domain.

The CS theory introduced in [2–6] provides answers to the question of reconstructing a signal from its compressed measurements  $\mathbf{y}$ , which is defined as follows

$$\mathbf{y} = \phi \mathbf{x} = \phi \cdot \psi \cdot \mathbf{s} = \theta \cdot \mathbf{s}, \quad (3)$$

where  $\phi$  is the  $M \times N$  measurement matrix and  $M \ll N$ . The reconstruction of the original signal  $\mathbf{x}$  from its compressed measurements  $\mathbf{y}$  cannot be achieved by simple matrix inversion or inverse transformation techniques. A sparse solution can be obtained by solving the following optimization problem:

$$\mathbf{s}_p = \arg \min \|\mathbf{s}\|_0 \quad \text{such that} \quad \theta \cdot \mathbf{s} = \mathbf{y}. \quad (4)$$

However, this problem is an NP-complete optimization problem; therefore, its solution cannot be found easily. It is also shown in [2–4] that it is possible to construct the  $\phi$  matrix from random numbers, which are i.i.d. Gaussian random variables. In this case, the number of measurements should be chosen as

\* Corresponding author.

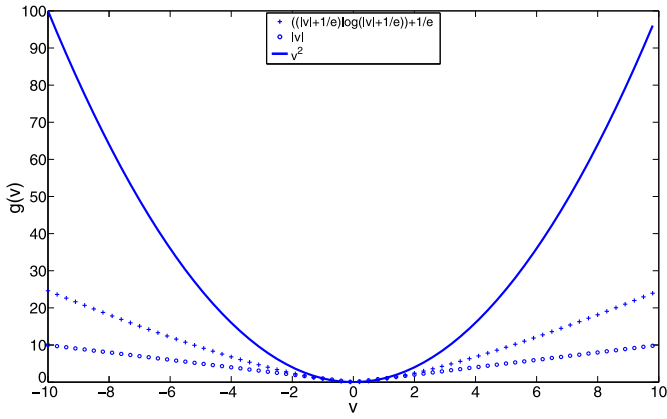


Fig. 1. Entropy functional  $g(v)$  (+),  $|v|$  (o) that is used in  $\ell_1$  norm and the Euclidean cost function  $v^2$  (–) that is used in  $\ell_2$  norm.

$cK \log(\frac{N}{K}) < M \ll N$  to satisfy the conditions for perfect reconstruction [2], and [3]. With this choice of the measurement matrix, the optimization problem (4) can be approximated by  $\ell_1$  norm minimization as:

$$\mathbf{s}_p = \arg \min \|\mathbf{s}\|_1 \quad \text{such that} \quad \theta \cdot \mathbf{s} = \mathbf{y}. \quad (5)$$

Instead of solving the original CS problem in (4) or (5), several researchers reformulate them to approximate the solution. For example, in [15], the authors developed a Bayesian framework and solved the CS problem using Relevance Vector Machines (RVM). In [7,8] the authors replaced the objective function of the CS optimization in (4), (5) with a new objective function to solve the sparse signal reconstruction problem. One popular approach is replacing  $\ell_0$  norm with  $\ell_p$  norm, where  $p \in (0, 1)$  [7,9] or even with the mix of two different norms as in [10]. However, in these cases, the resulting optimization problems are not convex. Several studies in the literature addressed  $\ell_p$  norm based non-convex optimization problems and applied their results to the sparse signal reconstruction example [11–14].

The entropy functional  $g(v) = v \log v$  is also used to approximate the solution of  $\ell_1$  optimization and linear programming problems in signal and image reconstruction by Bregman [16], and others [17–23]. In this article, we propose the use of a modified version of the entropy functional as an alternative way to approximate the CS problem. In Fig. 1, plots of the different cost functions including the proposed modified entropy function

$$g(v) = \left( |v| + \frac{1}{e} \right) \log \left( |v| + \frac{1}{e} \right) + \frac{1}{e}, \quad (6)$$

as well as the absolute value  $g(v) = |v|$  and  $g(v) = v^2$  are shown. The modified entropy functional (6) is convex, continuous and differentiable, it slowly increases compared to  $g(v) = v^2$ , because  $\log(v)$  is much smaller than  $v$  for high  $v$  values as seen in Fig. 1. The convexity proof for the modified entropy functional is given in Appendix A.

Bregman also developed iterative row-action methods to solve the global optimization problem by successive local Bregman-projections. In each iteration step, a Bregman-projection, which is a generalized version of the orthogonal projection, is performed onto a hyperplane representing a row of the constraint matrix  $\theta$ . In [16], Bregman proved that the proposed iterative method is guaranteed to converge to the global minimum, given that there is a proper choice of the initial estimate (e.g.,  $\mathbf{v}_0 = 0$ ).

An interesting interpretation of the row-action approach is that it provides an on-line solution to the CS problem. Each new measurement of the signal adds a row to the matrix  $\theta$ . In the iterative

row-action method, a Bregman-projection is performed onto the new hyperplane formed by the new measurement. In this way, the currently available solution is updated without solving the entire CS problem. The new solution can be further updated by using past or new measurements in an iterative manner by performing other Bregman-projections. Therefore, it is possible to develop a real-time on-line CS method using the proposed approach.

In Section 2 of this paper, we review the Bregman-projection concept and define the modified entropy functional and related Bregman-projections. We generalize the entropy function based convex optimization method introduced by Bregman because the ordinary entropy function is defined only for positive real numbers. On the other hand, transform domain coefficients can be both positive and negative.

Section 2 also contains the Bregman-projection definition, and formulation of the entropy functional based CS reconstruction problem. We define the iterative CS algorithm in Section 2.1, and provide experimental results in Section 4.

## 2. Bregman-projection based algorithm

The  $\ell_0$  and  $\ell_1$  norm based cost functions (4) and (5) used in compressive sensing problems are not differentiable everywhere. Therefore, it is not possible to use convex optimization algorithms to solve the CS problems in (4) and (5). Besides, as the size of the problem increases, solving these optimization problems becomes very compelling. As the original CS problem given in (4) and (5) involves non-convex  $\ell_0$  and  $\ell_1$  cost functions, it cannot be divided into simpler subproblems for convex optimization.

In this article, we replace  $\ell_0$  or  $\ell_1$  norms in the original CS problem with a new cost function called modified entropy function. In this way, it becomes possible to utilize Bregman's iterative convex optimization methods. Bregman's algorithms have been widely used in many signal processing applications such as signal reconstruction and inverse problems [17,18,22–31]. Here, we introduce an entropy based cost function that leads to an iterative solution of the CS problem by dividing it into simpler convex subproblems.

Assume that the original signal  $\mathbf{x}$  can be represented by a  $K$ -sparse length- $N$  vector  $\mathbf{s}$  in a transform domain characterized by the transform matrix  $\psi$ . In CS problems, the original signal  $\mathbf{x}$  is not available. However  $M$  measurements  $\mathbf{y} = [y_1, \dots, y_M]^T = \phi \mathbf{x}$  of the original signal are observable via the measurement matrix  $\phi$ , and the relations between  $\mathbf{y}$  and  $\mathbf{s}$  are described as in Eq. (3). CS theory suggests that we can find  $\mathbf{x}$  using  $\ell_1$  minimization if certain conditions hold, such as the Restricted Isometry Property [3].

Bregman's method provides globally convergent iterative algorithms to solve optimization problems with convex, continuous and differentiable cost functionals  $g(\cdot)$ :

$$\min_{\mathbf{s} \in \mathcal{C}} g(\mathbf{s}), \quad (7)$$

such that

$$\theta_i \cdot \mathbf{s} = y_i \quad \text{for } i = 1, 2, \dots, M, \quad (8)$$

where  $\theta_i$  is the  $i$ th row of the matrix  $\theta$ . In [16], Bregman showed that optimization problems with continuous and differentiable cost functionals can be divided into subproblems, which can be solved in an iterative manner, to approximate the solution of the original problem. Each equation in (8) represents a hyperplane  $H_i \in \mathbb{R}^N$ , which are closed and convex sets in  $\mathbb{R}^N$ . In Bregman's method, the iterative reconstruction algorithm starts with an arbitrary initial estimate and successive Bregman-projections are performed onto the hyperplanes  $H_i$ ,  $i = 1, 2, \dots, M$ , in each step of the iterative algorithm.

The Bregman-projection onto a closed and convex set is a generalized version of the orthogonal projection onto a convex set [16]. Let  $\mathbf{s}^o$  be an arbitrary vector in  $\mathbb{R}^N$ . Its Bregman-projection  $\mathbf{s}^p$  onto a closed convex set  $C$  with respect to a cost functional  $g(\mathbf{s})$  is defined as follows:

$$\mathbf{s}^p = \arg \min_{\mathbf{s} \in C} D(\mathbf{s}, \mathbf{s}^o), \quad (9)$$

where

$$D(\mathbf{s}, \mathbf{s}^o) = g(\mathbf{s}) - g(\mathbf{s}^o) - \langle \nabla g(\mathbf{s}^o), \mathbf{s}^o - \mathbf{s} \rangle, \quad (10)$$

and  $D$  is the distance function related with the cost function  $\|\cdot\|_d$ ,  $g$  is the distance measure and  $\nabla$  is the gradient operator.

In CS problems, we have  $M$  hyperplanes  $H_i : \theta_i \mathbf{s} = y_i$  for  $i = 1, 2, \dots, M$ . For each hyperplane  $H_i$ , the Bregman-projection (9) is equivalent to

$$\nabla g(\mathbf{s}^p) = \nabla g(\mathbf{s}^o) + \lambda \theta_i, \quad (11)$$

$$\theta_i \mathbf{s}^p = y_i \quad (12)$$

where  $\lambda$  is the Lagrange multiplier. As pointed out above, the Bregman-projection is a generalization of the orthogonal projection. When the cost functional is the Euclidean cost functional  $g(\mathbf{s}) = \sum_n s[n]^2$ , the distance  $D(\mathbf{s}^1, \mathbf{s}^2)$  becomes the  $\ell_2$  norm of difference vector  $(\mathbf{s}^1 - \mathbf{s}^2)$ , and the Bregman-projection simply becomes the well-known orthogonal projection onto a hyperplane.

When the cost functional is the entropy functional, which is defined as

$$g(\mathbf{s}) = \sum_n s(n) \log(s(n)), \quad (13)$$

the Bregman-projection onto the hyperplane  $H_i$  leads to the following equation

$$s^p(n) = s^o(n) \cdot e^{\lambda \cdot \theta_i(n)}, \quad n = 1, 2, \dots, N, \quad (14)$$

where the Lagrange multiplier  $\lambda$  is obtained by inserting (14) into the hyperplane equation given in (8). The previous set of equations are used in signal reconstruction from Fourier Transform samples [23] and the tomographic reconstruction problem [17].

However, in its original form, entropy is only defined for positive real numbers. In CS problem entries of vector  $\mathbf{s}$  can take both positive and negative values. We thus modify the original entropy function and extend it to negative real numbers as follows:

$$g_e(\mathbf{s}) = \sum_{n=1}^N \left( \left( |s[n]| + \frac{1}{e} \right) \cdot \left( \log \left( |s[n]| + \frac{1}{e} \right) \right) \right) + \frac{1}{e}, \quad (15)$$

where the subscript  $e$  represents the term entropy. The modified entropy function satisfies the following conditions:

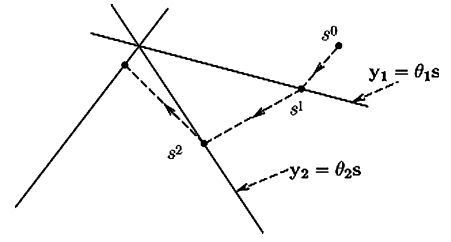
- (i)  $\frac{\partial g_e}{\partial s[n]}(0) = 0, n = 1, 2, \dots, N$ , and
- (ii)  $g_e$  is strictly convex and continuously differentiable.

Therefore the following optimization problem

$$\min g_e(\mathbf{s}) \quad \text{s.t.} \quad \theta_i \mathbf{s} = y_i \quad (16)$$

can be solved using Bregman's convex optimization method.

On the other hand, the  $\ell_1$  norm is not a continuously differentiable function; therefore, non-differentiable minimization techniques such as sub-gradient methods [32] should be used for solving  $\ell_1$  based optimization problems. Another way of approximating the  $\ell_1$  penalty function using an entropy functional is also available in [33].



**Fig. 2.** Geometric interpretation of the entropic projection method: Sparse representation  $\mathbf{s}^i$  corresponding to decision functions at each iteration are updated so as to satisfy the hyperplane equations defined by the measurements  $y_i$  and the measurement vector  $\theta_i$ . Lines in the figure represent hyperplanes in  $\mathbb{R}^N$ . Sparse representation vector  $s_i$  converges to the intersection of the hyperplanes. Notice that Bregman-projections are not orthogonal projections.

To obtain the Bregman-projection of  $\mathbf{s}^o$  onto a hyperplane  $H_i$  with respect to the entropic cost functional (16), we need to minimize the generalized Bregman distance  $D(\mathbf{s}, \mathbf{s}^o)$  between  $\mathbf{s}^o$  and the hyperplane  $H_i$ :

$$D(\mathbf{s}, \mathbf{s}^o) = g_e(\mathbf{s}^o) - g_e(\mathbf{s}) - \langle \nabla g_e(\mathbf{s}), \mathbf{s}^o - \mathbf{s} \rangle \quad (17)$$

with the condition that  $\theta_i \mathbf{s} = y_i$ . Using the modified entropy cost functional (15) in (11), entries of the projection vector  $\mathbf{s}^p$  can be obtained as:

$$\begin{aligned} \text{sgn}(s^p[n]) \cdot \left[ \log \left( |s^p[n]| + \frac{1}{e} \right) + 1 \right] \\ = \text{sgn}(s^o[n]) \cdot \left[ \log \left( |s^o[n]| + \frac{1}{e} \right) + 1 \right] + \lambda \theta_i[n], \\ n = 1, 2, \dots, N, \end{aligned} \quad (18)$$

where  $\lambda$  is the Lagrange multiplier, which can be obtained from  $\theta_i \mathbf{s} = y_i$ .

The Bregman-projection vector  $\mathbf{s}^p$  is the solution that satisfies the set of equations (18), and the hyperplane equation  $H_i : \theta_i \mathbf{s} = y_i$ .

### 2.1. Iterative reconstruction algorithm

The global convex optimization problem defined in (16) is solved by performing successive local Bregman-projections onto hyperplanes defined by the rows of the matrix  $\theta$ .

The iterations start with an initial estimate  $\mathbf{s}^0 = 0$ . In the first iteration cycle, this vector is Bregman-projected onto the hyperplane  $H_1$  and  $\mathbf{s}_1$  is obtained. The iterate  $\mathbf{s}^1$  is projected onto the next hyperplane  $H_2$  (see Fig. 2). This iterative process continues until the  $(N - 1)$ st estimate  $\mathbf{s}^{N-1}$  is Bregman-projected onto  $H_N$  and  $\mathbf{s}^N$  is obtained. In this way the first iteration cycle is completed. In the next cycle, the vector  $\mathbf{s}^N$  is projected onto the hyperplane  $H_1$  and  $\mathbf{s}^{N+1}$  is obtained, and so on. Bregman proved that  $\mathbf{s}_i$  defined in (16) converges to the solution of the optimization problem. In Appendix A, the proof of convergence of the proposed algorithm is provided.

Bregman-projection method can handle inequality constraints as well. The iterative algorithm is still globally convergent (Appendix A) when the equality constraints in (8) are relaxed by  $\epsilon_i$

$$y_i - \epsilon_i \leq \theta_i \mathbf{s} \leq y_i + \epsilon_i, \quad i = 1, 2, \dots, N. \quad (19)$$

This is because hyperslabs defined by (19) are closed and convex sets. In each step of the iterative algorithm the current iterate is projected onto the closest boundary hyperplane defined by one of the inequality signs in (19). If the iterate satisfies the current inequality, it is simply projected onto the next hyperslab.

It is also possible to use block iterative projection methods, which converge faster than single projection based methods [28,34]. Usually block iterative algorithms handling a block

of hyperplanes at the same time converge faster than the single hyperplane based algorithm that we described above. Since the compressive sensing problems described in this paper are offline problems, the speed of convergence is not that important.

### 3. Proximal splitting based algorithm

In [35–38] proximity operators of convex functions and their signal processing applications are reviewed. Various proximal splitting algorithms for convex optimization problems including the forward–backward splitting (FBS) algorithm are also presented. In [39], the proof of convergence of the FBS algorithm, and a framework, which uses Bregman distance function (17) as the cost function, is presented. This framework can be used to solve convex optimization problems involving modified entropy function. The FBS based iterative algorithm has the following update equation:

$$\mathbf{s}^{n+1} = P_{c_i}(\mathbf{s}^n - \gamma_n \nabla g_e(\mathbf{s}^n)), \quad (20)$$

where  $P_{c_i}$  is the orthogonal projection operator onto the  $i$ th hyperplane,  $\gamma_n$  is the step size satisfying  $0 < \gamma \leq 1/L$ , and  $\nabla g_e(\cdot)$  is  $L$ -Lipschitz continuous gradient of the modified entropy function [35,39]. As given in (18), we have an analytic expression for the gradient of the modified entropic cost function, therefore we do not need to solve any non-linear equations to obtain the next iterate  $\mathbf{s}^{n+1}$ . The algorithm is summarized as follows:

```

Begin  $\mathbf{s}^0 \in \mathbb{R}^N$ 
For  $n = 0, 1, \dots$ 
   $\gamma_n \in [0, 1/L]$ 
   $\mathbf{v}^n = \mathbf{s}^n - \gamma_n \nabla g_e(\mathbf{s}_n)$ 
   $\lambda_n \in (0, 1)$ 
   $\mathbf{s}^{n+1} = \mathbf{s}^n + \lambda_n (P_{c,n}(\mathbf{v}_n) - \mathbf{s}^n)$ 

```

where the  $P_{c,n}$  operation in the last step is the orthogonal projections onto the  $n$ th hyperplane defined in (8). Proximity splitting method reduces the computational cost significantly because there is no need to solve any non-linear equations and we have an analytic expression for the gradient of the cost function. The convergence of the algorithm is proved in [39]. It is also possible to obtain a block iterative version of the FBS algorithm as described in [28,34].

### 4. Experimental results

For the validation and testing of the entropic minimization method, experiments with four different one-dimensional (1D) signals, and 30 different images are carried out. The *cusp* signal, which consists of 1024 samples, and the *hisine* signal, which consists of 256 samples are shown in Figs. 3 and 4, respectively. The cusp and hisine signals can be sparsely approximated in the DCT domain. The 4 and 25 sparse random signals are composed of 128 and 256 samples, respectively, and they consist of 4 and 25 randomly located non-zero samples, respectively. The measurement matrices  $\phi$  are chosen as Gaussian random matrices.

In the first set of experiments, the original signal is reconstructed, when  $M = 204$ , 717 measurements are taken from the cusp signal and, when  $M = 24$ , 40 measurements are taken from the  $S = 5$  random signal. The reconstructed signals using the entropy based cost functional based algorithm are shown in Figs. 5(a), 5(b), 6(a), and 6(b). The cusp signal has 76 DCT coefficients, whose magnitudes are larger than  $10^{-2}$ . Therefore, it can be approximated as  $S = 76$  sparse signal in the DCT domain. The performance of the reconstruction is measured using the SNR criterion, which is defined as follows

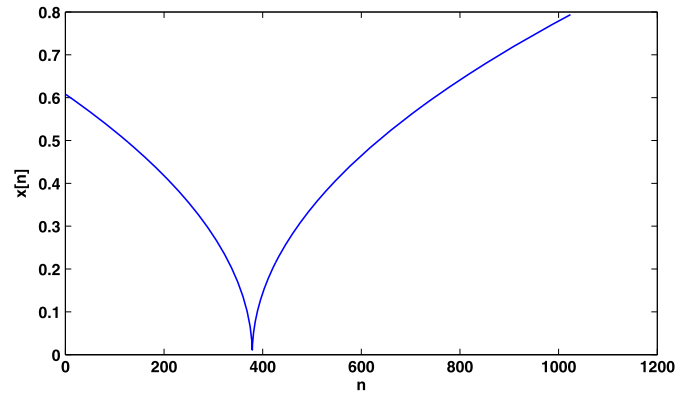


Fig. 3. The cusp signal with  $N = 1024$  samples.

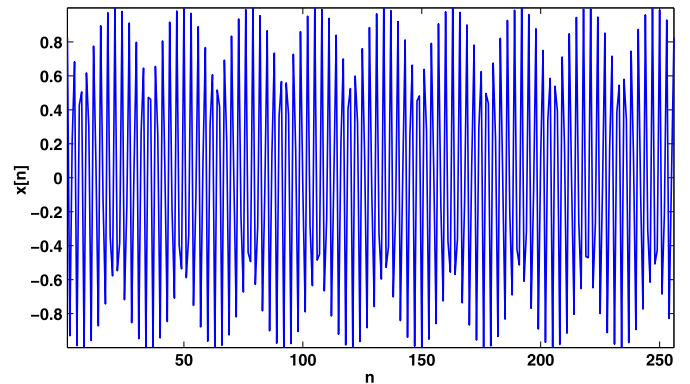


Fig. 4. Hisine signal with  $N = 256$  samples.

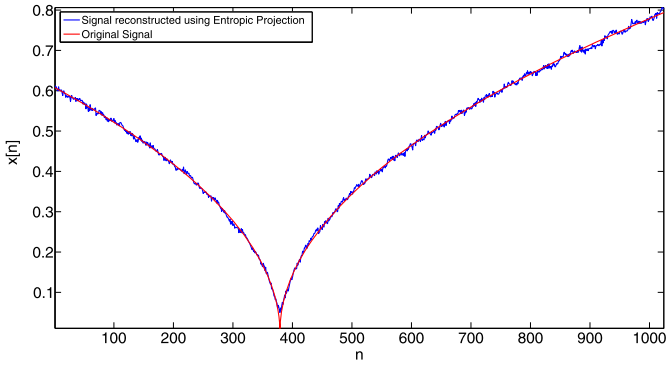
$$\text{SNR} = 20 \log_{10} \left( \frac{\|\mathbf{x}\|_2}{\|\mathbf{x} - \mathbf{x}_{\text{rec}}\|_2} \right), \quad (21)$$

where  $\mathbf{x}$  is the original signal and  $\mathbf{x}_{\text{rec}}$  is the reconstructed signal. The results of 39 and 44 dB SNR are achieved by reconstructing the original signal using the proposed method from  $M = 204$ , and 717 measurements, respectively. In the case of the experiment with random signals, the proposed method missed one sample from the original signal when using 30 measurements and perfectly reconstructed the original signal when using 50 measurements.

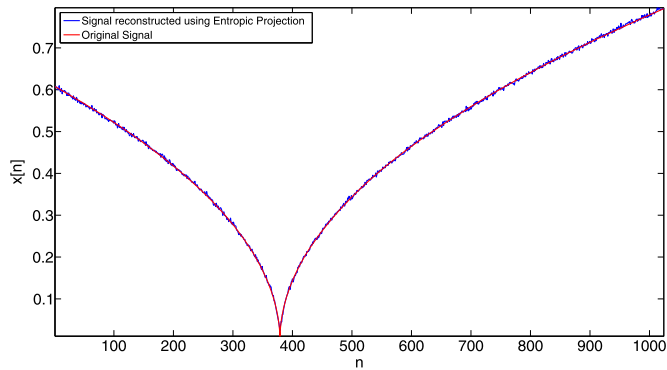
In the next set of experiments we compared our reconstruction results with 4 well-known CS reconstruction algorithms from the literature; CoSamp [40], Matching Pursuit (MP) [41],  $\ell_1$ magic [42], and  $\ell_p$  optimization based CS reconstruction [9] algorithms. We tested the proposed method against the  $\ell_p$  optimization based CS reconstruction algorithm with 3 different  $p$  values;  $p = [0.8, 1, 1.7]$ . When  $p = 1$ , the algorithm solves the  $\ell_1$  norm optimization problem given in (5). The reason why we choose  $p = 1.7$  to test against the proposed algorithm is that the  $\ell_{1.7}$  norm curve is very similar to the curve of the modified entropic functional (Fig. 1).

In this set of experiments, by taking different amounts of measurements ranging from 10% to 80% of the total number of the samples of the 1D signal, we tried to reconstruct the original signal. Then, we measured the SNR between the original and the reconstructed signals. In these tests, the main region of interest is 20–60% range.

We present results of the tests with the cusp signal in Fig. 7. The proposed algorithm performed better than  $\ell_1$ magic, CoSamp algorithms in the 20–50% measurements range. It also has a comparable performance with the rest of the algorithms. In Fig. 8, the results of the tests with hisine signal are presented. The proposed algorithm performed significantly better than  $\ell_1$ magic and CoSamp, and marginally outperformed the rest of the algorithms.



(a)  $N = 1024$  length cusp signal reconstructed from 204 measurements



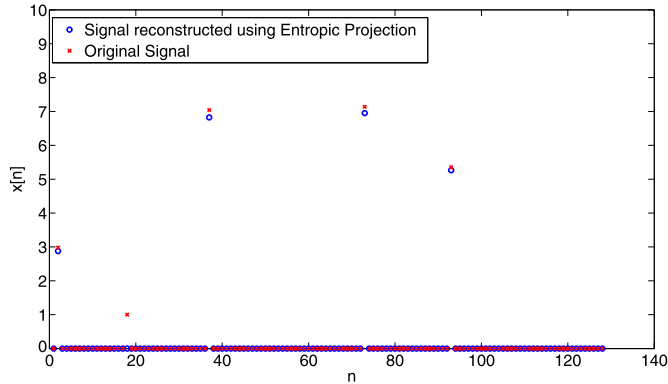
(b)  $N = 1024$  length cusp signal reconstructed from 716 measurements

**Fig. 5.** The cusp signal with 1024 samples reconstructed from  $M = 204$  (a) and  $M = 716$  (b) measurements using the iterative, entropy functional based method.

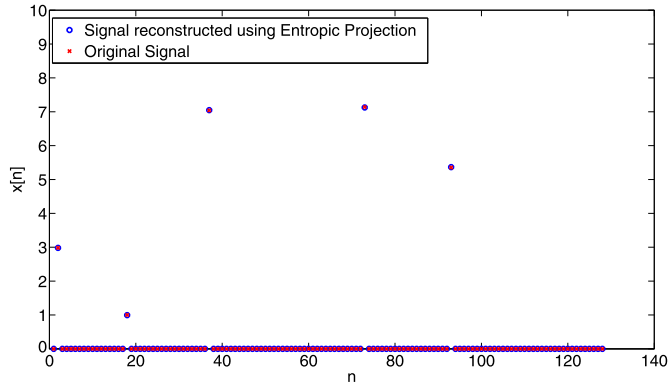
It is important to note that, both the cusp and the hisine signals are not sparse but compressible in the sense that most of the transform domain coefficients are not zero but close to zero [43]. Therefore, the sparsity level of the test signals are not known exactly beforehand.

In case of the tests with the 25-sparse impulse signal, which consists of isolated impulses, CoSamp outperformed all the other algorithms. In Fig. 9, we presented the results of reconstructing the signal from measurements that as many as 20–40% of the signal samples. Above 40%, all the algorithms except Matching Pursuit and  $\ell_{1,7}$  norm based reconstruction algorithms, achieved more than 50 dB SNR. We believe that above 40–50 dB of SNR, the signal reconstruction can be counted as a perfect reconstruction. Therefore, we compared the percentage of measurements at which the individual algorithms achieved 50 dB SNR. The proposed algorithm achieved 50 dB SNR around 30% of the measurements. Due to numerical imprecision in the calculation of the alternating entropic projections, the proposed algorithm achieved approximately 50 dB SNR. Only  $\ell_{0,8}$  norm based reconstruction algorithm and CoSamp achieved this bound at lower measurement rates compared to the proposed algorithm. The entropic projection based method outperformed the rest of the algorithms.

In the last set of experiments, we implemented the proposed algorithm in 2-dimension (2D) and applied it to six well-known images from the image processing literature and 24 images from the Kodak dataset [44]. The images in Kodak dataset are 24 bit per pixel color images. We first transformed all the color images into YUV color space and used the 8 bit per pixel luminance component (Y channel) of the images in our tests. We compared our results with the block based compressed sensing algorithm given in [45]. As in [45], we divided the image into blocks and reconstructed those blocks individually. We tested both the proposed algorithm and Fowler et al.’s method using random measurements,



(a)  $N = 128$  length random sparse signal reconstructed from  $3S = 15$  measurements



(b)  $N = 128$  length random sparse signal reconstructed from  $4S = 20$  measurements

**Fig. 6.** Random sparse signal with 128 samples is reconstructed from (a)  $M = 3S$  and (b)  $M = 4S$  measurements using the iterative, entropy functional based method.

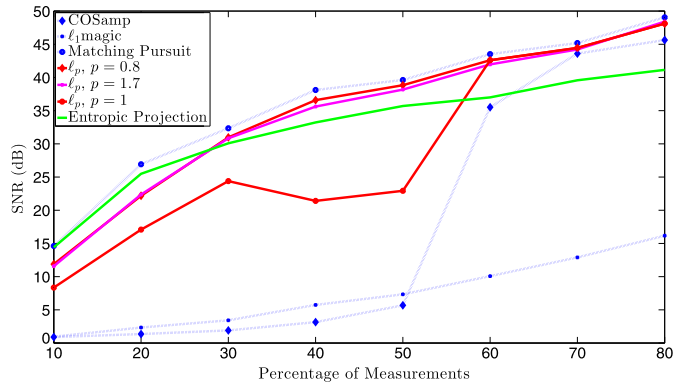


Fig. 7. The reconstructed cusp signal with  $N = 256$  samples.

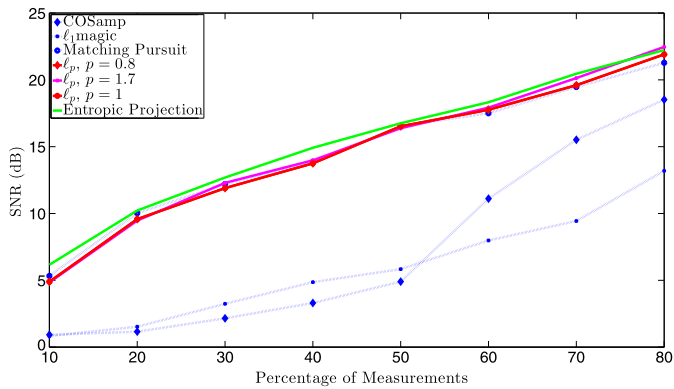


Fig. 8. The reconstruction error for a hisine signal with  $N = 256$  samples.

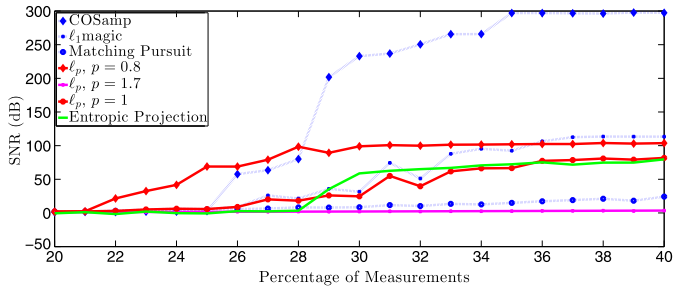


Fig. 9. The impulse signal with  $N = 256$  samples. The signal consists of 25 random amplitudes that are located at random locations in the signal.

Table 1

Image reconstruction results. The images are reconstructed using measurements that are 30% of the total number of the pixels in the image.

Images	Fowler's method [45] SNR in dB	Proposed method SNR in dB
Barbara	19.412	18.528
Mandrill	16.822	17.401
Lenna	26.516	26.806
Goldhill	22.473	23.857
Fingerprint	20.171	22.205
Peppers	26.831	25.854
Kodak (average)	21.51	21.98
Average	21.615	22.072

that are as many as 30% of the total number of pixels in the image. On average, we achieved approximately a 0.4 dB higher SNR compared to the algorithm given in [45], as shown in Table 1. In both methods, the images are processed using a Wiener filter to smooth out the blocking artifacts caused by block processing.

In all of the previous examples, the entropic projection algorithm is implemented in the following way. The algorithm starts with an initial estimate of the signal, such as a zero amplitude signal or a signal reconstructed using a pseudo inversion. The choice of the initial estimate of the signal may affect the speed of convergence. Then in the first iteration cycle, the estimated signal is entropically projected onto the hyperplanes defined by the measurements, one after the other. At the end of the iteration cycle, the transform domain coefficients of the resulting estimate are rank-ordered according to their magnitude values with only the significant coefficients being kept and the rest set to zero. After each iteration cycle, the number of retained transform domain coefficients that are kept is increased by one. The number of the coefficients that are kept does not exceed the number of measurements. If the initial signal is known to be exactly  $K$ -sparse, then only  $K$  transform domain coefficients, which have the largest amplitude, are kept.

## 5. Conclusion

In this article, we introduced an iterative CS reconstruction algorithm that uses an entropy based cost functional. The proposed entropy based cost functional is convex, continuous and differentiable everywhere and approximates  $\ell_1$  norm in the original CS formulation. It is convex, continuous and differentiable, therefore, it enables the user to divide the large and complex CS problems into smaller and simpler subproblems. We developed a globally convergent algorithm, that solves these individual subproblems in an iterative manner to approximate the original problem solution. It is experimentally observed that the entropy based cost function and the iterative row-action method can be used for reconstructing both sparse and compressible signals from their compressed measurements. Since most practical signals are not completely sparse but compressible, the proposed algorithm is suitable for compressive sensing applications of real life signals.

The proposed method is capable of dividing large CS reconstruction problems into smaller and simpler parts. Therefore, it can be used to simplify large scale problems and provide computationally efficient ways to solve those problems. Moreover it is also shown that the row-action methods provide an on-line solution to the CS problem. The reconstruction result can be updated on-line according to the new measurements without solving the entire optimization problem again in real time.

## Acknowledgment

This work was supported by Scientific and Technical Research Council of Turkey, TUBITAK with a grant number 111E057 (Compressive Sensing Algorithms Using Entropy Functional and Total Variation Constraints).

## Appendix A. Proof of convergence of the iterative algorithm

The problem described in (8) and (9) is a convex programming problem:

$$\begin{aligned} \min_{\mathbf{s} \in H} \quad & g(\mathbf{s}) \\ \text{subject to} \quad & \theta_i \cdot \mathbf{s} = y_i \quad \text{for } i = 1, 2, \dots, M, \end{aligned} \quad (\text{A.1})$$

where  $g(\mathbf{s})$  is a strictly convex and differentiable cost function in  $\mathbb{R}^N$ ,  $H$  is the intersection of  $M$  hyperplanes  $\theta_i \cdot \mathbf{s} = y_i$ , and  $\mathbf{s} \in \mathbb{R}^N$ . In [16], Bregman solved the convex optimization problem (A.1) using Bregman-projections. He proved in [16, Theorem 3] that starting from an initial point  $\mathbf{s}_0 = 0$ , and making successive Bregman-projections on convex hyperplanes as defined by  $\theta_i \cdot \mathbf{s} = y_i$  (Section 2.1), converges to the solution of the convex optimization problem, provided that  $H$  is non-empty.

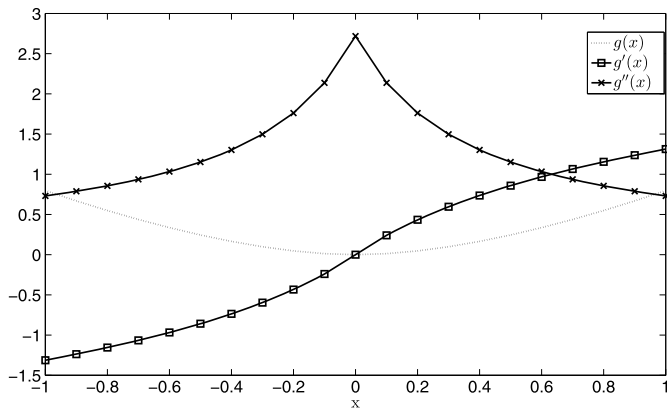


Fig. 10. The plot of the entropic cost function, its first, and second derivatives.

**Statement 1.** The function  $g(x) = (|x| + \frac{1}{e}) \log(|x| + \frac{1}{e}) + \frac{1}{e}$  is continuously differentiable in  $\mathbb{R}$ .

**Proof.** The derivative of the cost function  $g(x)$  can be computed using the chain rule. The first derivative of the cost function  $g(x)$  is

$$g'(x) = \text{sgn}(x) \left[ \log\left(|x| + \frac{1}{e}\right) + 1 \right], \quad (\text{A.2})$$

which is a continuous function in  $\mathbb{R}$ . The plot of the function is shown in Fig. 10. Extension to  $\mathbb{R}^N$  is straightforward.  $\square$

**Statement 2.** The function  $g(x)$  is a strictly convex function.

**Proof.** The second derivatives of the cost function  $g(x) = (|x| + \frac{1}{e}) \log(|x| + \frac{1}{e}) + \frac{1}{e}$  is

$$g''(x) = \frac{1}{|x| + \frac{1}{e}} > 0, \quad (\text{A.3})$$

where  $g(x) > 0, \forall x \in \mathbb{R}$ . The one-dimensional plot of the function is shown in Fig. 10. The cost function is strictly convex because its second derivative is non-negative  $\forall x \in \mathbb{R}$ .

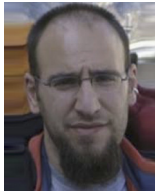
The problem described in (19) is also a convex programming problem. The convergence of this optimization problem can also be proved using Theorem 4 of [16] because  $g(\mathbf{s})$  is a strictly convex and differentiable function in  $\mathbb{R}^N$ .  $\square$

## References

- [1] C.E. Shannon, Communication in the presence of noise, Proc. Inst. Radio Eng. 37 (1) (January 1949) 10–21.
- [2] R.G. Baraniuk, Compressed sensing [Lecture Notes], IEEE Signal Process. Mag. 24 (4) (July 2007) 118–124.
- [3] E. Candes, J. Romberg, T. Tao, Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information, IEEE Trans. Inform. Theory 52 (2) (February 2006) 489–509.
- [4] E. Candes, Compressive sampling, in: International Congress of Mathematics, vol. 3, 2006, pp. 1433–1452.
- [5] E. Candes, T. Tao, Near optimal signal recovery from random projections: Universal encoding strategies?, IEEE Trans. Inform. Theory 52 (12) (December 2006) 5406–5425.
- [6] D. Donoho, Compressed sensing, IEEE Trans. Inform. Theory 52 (4) (April 2006) 1289–1306.
- [7] R. Chartrand, Exact reconstruction of sparse signals via nonconvex minimization, IEEE Signal Process. Lett. 14 (10) (Oct. 2007) 707–710.
- [8] T.Y. Rezaei, M.A. Tinati, S. Beheshti, Sparsity aware consistent and high precision variable selection, Signal Image Video Process. (ISSN 1863-1703) (2012) 1–12, <http://dx.doi.org/10.1007/s11760-012-0401-6>.

- [9] R. Chartrand, Wotao Yin, Iteratively reweighted algorithms for compressive sensing, in: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2008, pp. 3869–3872.
- [10] M. Kowalski, B.A. Torrèsani, Sparsity and persistence: mixed norms provide simple signal models with dependent coefficients, Signal Image Video Process. 3 (3) (2009) 251–264.
- [11] M. Ehler, Shrinkage rules for variational minimization problems and applications to analytical ultracentrifugation, J. Inverse Ill-Posed Probl. 19 (2011) 593–614.
- [12] Kristian Bredies, Dirk A. Lorenz, Minimization of non-smooth, non-convex functionals by iterative thresholding, April 2009 (DFG SPP 1324 Preprint 10), submitted for publication, preprint available at <http://www.uni-graz.at/~bredies/publications.html>.
- [13] A. Achim, B. Buxton, G. Tzagkarakis, P. Tsakalides, Compressive sensing for ultrasound RF echoes using  $\alpha$ -stable distributions, in: IEEE Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference, September 2010, pp. 4304–4307.
- [14] G. Tzagkarakis, P. Tsakalides, Greedy sparse reconstruction of non-negative signals using symmetric  $\alpha$ -stable distributions, in: Proceedings of European Signal Processing Conference (EUSIPCO), 2010.
- [15] S. Ji, Y. Xue, L. Carin, Bayesian compressive sensing, IEEE Trans. Signal Process. 56 (6) (June 2008) 2346–2356.
- [16] L.M. Bregman, The relaxation method of finding the common point of convex sets and its applications to the solution of the problems in convex programming, USSR Comput. Math. Math. Phys. 7 (1967) 200–217.
- [17] G.T. Herman, Image reconstruction from projections, Real-Time Imag. 1 (1) (1995) 3–18.
- [18] Y. Censor, A. Lent, An iterative row-action method for interval convex programming, J. Optim. Theory Appl. 34 (3) (1981) 321–353.
- [19] A. Lent, H. Tuy, An iterative method for the extrapolation of band-limited functions, J. Math. Anal. Appl. 83 (2) (1981) 554–565.
- [20] S. Jafarpour, V. Cevher, R.E. Schapire, A game theoretic approach to expander-based compressive sensing, in: IEEE International Symposium on Information Theory Proceedings (ISIT), 2011, pp. 464–468.
- [21] S. Jafarpour, R.E. Schapire, V. Cevher, Compressive sensing meets game theory, in: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2011, pp. 3660–3663.
- [22] A.E. Cetin, An iterative algorithm for signal reconstruction from bispectrum, IEEE Trans. Signal Process. 39 (12) (December 1991) 2621–2628.
- [23] A.E. Cetin, R. Ansari, Convolution-based framework for signal recovery and applications, J. Opt. Soc. Amer. A 5 (8) (1988) 1193–1200.
- [24] D.C. Youla, H. Webb, Image restoration by the method of convex projections: Part 1—Theory, IEEE Trans. Med. Imag. 1 (2) (1982) 81–94.
- [25] H. Trussell, M.R. Civanlar, The Landweber iteration and projection onto convex sets, IEEE Trans. Acoust. Speech Signal Process. 33 (6) (December 1985) 1632–1634.
- [26] M.I. Sezan, H. Stark, Image restoration by the method of convex projections: Part 2—Applications and numerical results, IEEE Trans. Med. Imag. 1 (2) (October 1982) 95–101.
- [27] P.L. Combettes, The foundations of set theoretic estimation, Proc. IEEE 81 (2) (February 1993) 182–208.
- [28] S. Theodoridis, K. Slavakis, I. Yamada, Adaptive learning in a world of projections, IEEE Signal Process. Mag. 28 (1) (January 2011) 97–123.
- [29] A.E. Cetin, R. Ansari, Signal recovery from wavelet transform maxima, IEEE Trans. Signal Process. 42 (1) (1994) 194–196.
- [30] J.F. Cai, S. Osher, Z. Shen, Linearized Bregman iterations for compressed sensing, Math. Comput. 78 (2009) 1515–1536.
- [31] P.L. Combettes, J.C. Pesquet, Proximal thresholding algorithm for minimization over orthonormal bases, SIAM J. Optim. 18 (2007) 1351–1376.
- [32] J.-B. Hiriart-Urruty, C. Lemarichal, Convex Analysis and Minimization Algorithms II, Springer-Verlag, Berlin, 1993.
- [33] M.C. Pinar, S.A. Zenios, An entropic approximation of  $\ell_1$  penalty function, Trans. Oper. Res. 7 (1995) 101–120.
- [34] R. Davidi, G. Herman, Y. Censor, Perturbation-resilient block-iterative projection methods with application to image reconstruction from projections, Int. Trans. Oper. Res. 16 (2009) 505–524.
- [35] P.L. Combettes, J.-C. Pesquet, Proximal splitting methods in signal processing, in: Fixed-Point Algorithms for Inverse Problems in Science and Engineering, in: Springer Optim. Appl., Springer, 2011, pp. 185–212.
- [36] M. Teboulle, Entropic proximal mappings with applications to nonlinear programming, Math. Oper. Res. 17 (3) (1992) 670–690.
- [37] A. Chambolle, Total variation minimization and a class of binary MRF model, in: Lecture Notes in Comput. Sci., vol. 3757, 2005, pp. 136–152.
- [38] M.A.T. Figueiredo, R.D. Nowak, S.J. Wright, Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems, IEEE J. Sel. Top. Signal Process. 1 (2007) 586–597.
- [39] C.L. Byrne, An elementary proof of convergence of the forward–backward splitting algorithm, J. Nonlinear Convex Anal. 190 (2013).

- [40] D. Needell, J.A. Tropp, CoSamp: Iterative signal recovery from incomplete and inaccurate samples, Technical report, California Institute of Technology, Pasadena, 2008.
- [41] S. Mallat, S. Zhang, Matching pursuits with time–frequency dictionaries, *IEEE Trans. Signal Process.* 41 (12) (1993) 3397–3415.
- [42] E. Candes, J. Romberg,  $\ell_1$ -magic: Recovery of sparse signals using convex optimization, <http://users.ece.gatech.edu/~justin/l1magic/>.
- [43] V. Cevher, M.F. Duarte, C. Hegde, R.G. Baraniuk, Sparse signal recovery using Markov random fields, in: Proceedings of the Workshop on Neural Information Processing Systems (NIPS), December 2008.
- [44] Kodak lossless true color image suite, <http://r0k.us/graphics/kodak/>, last visited on 05/05/2013.
- [45] J.E. Fowler, S. Mun, E.W. Tramel, Block-based compressed sensing of images and video, *Found. Trends Signal Process.* 4 (4) (March 2012) 297–416.



**Kivanc Kose** received his B.Sc., M.Sc., Ph.D. degrees in the Electrical and Electronics Engineering from Bilkent University. Since December 2012, he has been working as a post-doctoral research fellow at Dermatology Service in Memorial Sloan Kettering Cancer Center, New York, USA. He has been serving as an associate editor for *Signal, Image and Video Processing Journal* since August 2013. His research interests are multimedia signal analysis, computer vision, machine learning, and their applications in various fields such as medical imaging.



**Osman Gunay** received his B.Sc. and M.S. degrees in Electrical and Electronics Engineering from Bilkent University, Ankara, Turkey. He is currently working towards his Ph.D. in Electrical and Electronics Engineering at Bilkent University. His research interests include image analysis, computer vision, video segmentation, dynamic texture recognition and image retrieval.



**A. Enis Cetin** got his B.Sc. degree from Electrical Engineering at the Middle East Technical University; and got his M.S.E. and Ph.D. degrees in Systems Engineering from the Moore School of Electrical Engineering at the University of Pennsylvania in Philadelphia. Between 1987–1989, he was Assistant Professor of Electrical Engineering at the University of Toronto, Canada. Since then he has been with Bilkent University, Ankara, Turkey. Currently he is a full professor.

He has involved in more than ten EU and Nationally funded research projects as a researcher and/or project coordinator. He was in the editorial boards of *IEEE Transactions on Image Processing*, and several *EURASIP* journals. He has been the Editor in Chief of *Signal, Image and Video Processing* journal since July 2013. His research interests are Multimedia Signal Analysis and Applications, Computer Vision, Biomedical Image Processing, Human-Computer Interaction using vision and speech, Speech Processing, Digital Coding of Waveforms (Image, Video, Speech, and Biomedical Signals), Adaptive Filtering and Adaptive Subband Coding, Time Series Analysis and Stochastic Processes.