Digital Signal Processing 24 (2014) 63–70

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

# Compressive sensing using the modified entropy functional

Kivanc Kose [a,b,*], Osman Gunay [a], A. Enis Cetin [a]

[a] Electrical and Electronics Engineering Department, Bilkent University, Turkey
[b] 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.



## 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:

$$s_l = \langle \mathbf{x}, \psi_l \rangle, \quad l = 1, 2, \ldots, N, \tag{1}$$

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:

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

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 nonzero 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

$$\mathbf{y} = \phi \mathbf{x} = \phi . \psi . \mathbf{s} = \theta . \mathbf{s}, \tag{3}$$

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:

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

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.

1051-2004/$ – see front matter 
http://dx.doi.org/10.1016/j.dsp.2013.09.010

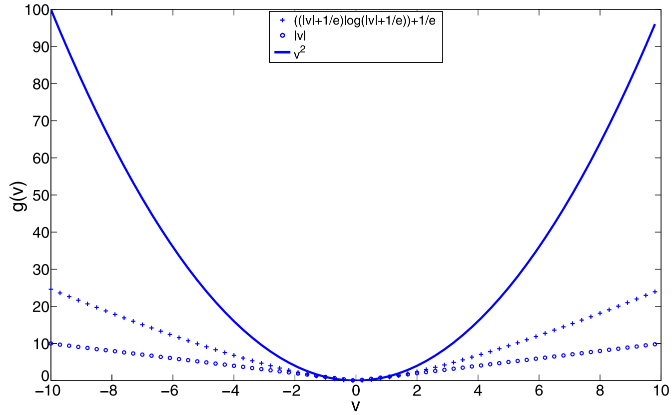


**Fig. 1.** Entropy functional $g(v)$ (+), $|v|$ (∘) that is used in $\ell_1$ norm and the Euclidean cost function $v^2$ (−) that is used in $\ell_2$ norm.

$cK\log(\frac{N}{K}) < M \ll N$ 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:

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

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

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

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 $\ell_o$ 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 $\mathbf{y} = [y_1, \ldots, y_M]^T = \phi\mathbf{x}$ 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 $g(.)$:

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

such that

$$\theta_i.\mathbf{s} = y_i \quad \text{for } i = 1, 2, \ldots, M, \tag{8}$$

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 $H_i \in R^N$, which are closed and convex sets in $R^N$. In Bregman's method, the iterative reconstruction algorithm starts with an arbitrary initial estimate and successive Bregman-projections are performed onto the hyperplanes $H_i$, $i = 1, 2, \ldots, M$, in each step of the iterative algorithm.