Methods for Sparse Signal Recovery Using Kalman Filtering With Embedded Pseudo-Measurement Norms and Quasi-Norms

Avishy Carmi, Pini Gurfil, and Dimitri Kanevsky

***Abstract*—We present two simple methods for recovering sparse signals from a series of noisy observations. The theory of compressed sensing (CS) requires solving a convex constrained minimization problem. We propose solving this optimization problem by two algorithms that rely on a Kalman filter (KF) endowed with a pseudo-measurement (PM) equation. Compared to a recently-introduced KF-CS method, which involves the implementation of an auxiliary CS optimization algorithm (e.g., the Dantzig selector), our method can be straightforwardly implemented in a stand-alone manner, as it is exclusively based on the well-known KF formulation. In our first algorithm, the PM equation constrains the l_1 norm of the estimated state. In this case, the augmented measurement equation becomes linear, so a regular KF can be used. In our second algorithm, we replace the l_1 norm by a quasi-norm $l_p, 0 \leq p < 1$. This modification considerably improves the accuracy of the resulting KF algorithm; however, these improved results require an extended KF (EKF) for properly computing the state statistics. A numerical study demonstrates the viability of the new methods.**

***Index Terms*—Compressed sensing, Kalman filtering, quasi-norms.**

I. Introduction

RECENT studies have shown that sparse signals can be recovered accurately using less observations than what is considered necessary by the Nyquist/Shannon sampling principle; the emergent theory that brought this insight into being is known as compressed sensing (CS) [1], [2]. The essence of the new theory builds upon a new data acquisition formalism, in which compression plays a fundamental role. From a filtering standpoint, one can think about a procedure in which signal recovery and compression are carried out simultaneously, thereby reducing the amount of required observations. Sparse, and more generally, compressible signals arise naturally in many fields of science and engineering. A typical example is the reconstruction of images from under-sampled Fourier data as encountered in radiology, biomedical imaging and astronomy [3], [4]. Other applications consider model-reduction methods to enforce sparseness for preventing over-fitting and for reducing computational complexity and storage capacities. The reader is referred to the seminal work reported in [1] and [2] for an extensive overview of the CS theory.

The recovery of sparse signals consists of solving an NP-hard minimization problem [1], [5]. This, however, can be relaxed under some restrictions by resorting to a convex l_1 minimization as suggested by the new CS theory [1]. The constrained l_1 minimization can be solved using various methods, such as the methods suggested by [6]–[9].

Manuscript received January 27, 2009; accepted November 09, 2009. First published December 18, 2009; current version published March 10, 2010. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Deniz Erdogmus.

A. Carmi is with the Signal Processing Group, Department of Engineering, University of Cambridge, U.K.

P. Gurfil is with the Faculty of Aerospace Engineering, Technion—Israel Institute of Technology, Haifa 32000, Israel (e-mail: pgurfil@technion.ac.il).

D. Kanevsky is with IBM T. J. Watson Research Center, Yorktown, NY 10598 USA.

Color versions of one or more of the figures in this correspondence are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2009.2038959

The first successful attempt of Kalman filtering-based compressed sensing was presented in [10], where the traditional Kalman filter (KF) algorithm was endowed with the Dantzig selector of [6]. This approach divides the sparse recovery problem into two interlaced subproblems: 1) support extraction and 2) reduced-order recovery. This separation roughly specifies a two phase algorithm in which some CS method (in this case the Dantzig selector) identifies the subset of elements in the support of the signal while the ordinary KF is applied to the reduced order system corresponding to the obtained subset.

In this correspondence, we propose a new Kalman filtering approach for recovery of possibly time-varying sparse signals. Our method is based on a pseudo-measurement formulation of the underlying constrained optimization problem. Compared to the algorithm in [10], our method can be straightforwardly implemented in a stand-alone manner, as it is exclusively based on the well-known KF formulation.

II. Sparse Signal Recovery

Consider an $\mathbb{R}^n$-valued random discrete-time process $\{x_k\}_{k=1}^{\infty}$ that is sparse in some known orthonormal sparsity basis $\psi \in \mathbb{R}^{n \times n}$, that is $z_k = \psi^T x_k$, $\#\{\text{supp}(z_k)\} < n$, where $\text{supp}(z_k)$ and $\#$ denote the support of z_k and the cardinality of a set, respectively (i.e., $\#\{\text{supp}(z_k)\}$ denotes the number of non-zero elements of z_k). Assume that z_k evolves according to

$$z_{k+1} = A z_k + w_k, \quad z_0 \sim \mathcal{N}(\mu_0, P_0) \tag{1}$$

where $A \in \mathbb{R}^{n \times n}$ is the state transition matrix and $\{w_k\}_{k=1}^{\infty}$ is a zero-mean white Gaussian sequence with covariance $Q_k \succeq 0$. Note that (1) does not necessarily imply a change in the support of the signal. For example, A can be a block-diagonal matrix decomposed of A^d and A^n corresponding to the statistically independent elements $z^d \notin \text{supp}(z_k)$ and $z^n \in \text{supp}(z_k)$, where the respective noise covariance submatrices satisfy $Q^d = 0$ and $Q^n \succeq 0$. The process x_k is measured by the $\mathbb{R}^m$-valued random process

$$y_k = H x_k + \zeta_k = H' z_k + \zeta_k \tag{2}$$

where $\{\zeta_k\}_{k=1}^{\infty}$ is a zero-mean white Gaussian sequence with covariance $R_k > 0$, and $H := H' \psi^T \in \mathbb{R}^{m \times n}$.

Letting $y^k := \{y_1, \ldots, y_k\}$, our problem is defined as follows. We are interested in finding a y^k-measurable estimator, $\hat{x}_k$, that is optimal in some sense. Often, the sought-after estimator is the one that minimizes the mean square error (MSE) $E\left[\|x_k - \hat{x}_k\|_2^2\right]$. It is well-known that if the linear system (1), (2) is observable then the solution to this problem can be obtained using the KF. On the other hand, if the system is unobservable, then the regular KF algorithm is useless; if, for instance, $A = I_{n \times n}$, then it may seem hopeless to reconstruct x_k from an under-determined system in which $m < n$ and $\text{rank}(H) < n$. Surprisingly, this problem may be circumvented by taking into account the fact that z_k is sparse.

A. The Combinatorial Problem and Compressed Sensing

References [1] and [5] have shown that in the deterministic case (i.e., when z is a parameter vector), one can accurately recover z (and therefore also x, i.e., $x = \psi z$) by solving the optimization problem

$$\min \|\hat{z}\|_0 \text{ s.t. } \sum_{i=1}^{k} \|y_i - H'\hat{z}\|_2^2 \leq \epsilon \tag{3}$$

for a sufficiently small ϵ, where $\|v\|_p = \left(\sum_{j=1}^n v_j^p\right)^{1/p}$ is the l_p-norm of v, and the zero-norm, $\|v\|_0$, is defined as[1] $\|v\|_0 := \#\{\text{supp}(v)\}$.

[1]For $0 \leq p < 1$, $\|v\|_p$ is not a norm; the common terminology is *zero norm* for $p = 0$ and *quasi-norm* for $0 < p < 1$.

Following a similar rationale, in the stochastic case the sought-after optimal estimator satisfies [2]

$$\min \|\hat{z}_k\|_0 \text{ s.t. } E_{z_k|y^k}\left[\|z_k - \hat{z}_k\|_2^2\right] \le \epsilon. \tag{4}$$

Unfortunately, the above optimization problems are NP-hard and cannot be solved efficiently. Recently, it has been shown that if the sensing matrix H' obeys the restricted isometry property (RIP), then the solution of the combinatorial problem (3) can almost always be obtained by solving the constrained convex optimization [1], [2]

$$\min \|\hat{z}\|_1 \text{ s.t. } \sum_{i=1}^{k} \|y_i - H'\hat{z}\|_2^2 \le \epsilon. \tag{5}$$

This is a fundamental result in a new emerging theory known as compressed sensing [1], [2]. The main idea is that the convex l_1 minimization problem can be efficiently solved using a myriad of existing methods, such as LASSO [7]. Other insights provided by CS are related to the construction of sensitivity matrices that satisfy the RIP. These underlying matrices are random by nature, which sheds a new light on the way observations should be sampled. For an extensive review of CS the reader is referred to [1] and [2].

III. CS-Embedded Kalman Filtering

For the system described by (1) and (2), the classical KF provides an estimate $\hat{z}_k$ that is a solution to the unconstrained l_2 minimization problem

$$\min_{\hat{z}_k} E_{z_k|y^k}\left[\|z_k - \hat{z}_k\|_2^2\right]. \tag{6}$$

Inspired by the CS approach—while retaining the KF objective function—we replace (4) with the dual problem [11]

$$\min_{\hat{z}_k} E_{z_k|y^k}\left[\|z_k - \hat{z}_k\|_2^2\right] \text{ s.t. } \|\hat{z}_k\|_1 \le \epsilon'. \tag{7}$$

The constrained optimization problem (7) can be solved in the framework of Kalman filtering using the pseudo-measurement (PM) technique [12]. Following the PM technique, the inequality constraint $\|z_k\|_1 \le \epsilon'$ is incorporated into the filtering process using a fictitious measurement $0 = \|z_k\|_1 - \epsilon'$, where ϵ' serves as a measurement noise. This PM can be rewritten as

$$0 = \bar{H}z_k - \epsilon', \quad \bar{H} := [\text{sign}(z_k(1)), \ldots, \text{sign}(z_k(n))] \tag{8}$$

where $\text{sign}(z_k(i))$ denotes the sign function of the i^{th} element of z_k (i.e., $\text{sign}(z_k(i)) = 1$ if $z_k(i) \ge 0$ and $\text{sign}(z_k(i)) = -1$ otherwise).

In this setting, the covariance R_ϵ of ϵ' is regarded as a tuning parameter that regulates the tightness of the constraint. An ideal value for this parameter may be chosen in much the same manner as the process noise covariance is determined in a common extended KF (EKF) tuning procedure (further discussions about various PM tuning approaches can be found in [12]). In the examples considered herein, it turns out that taking ϵ' to be of relatively high intensity, e. g. $R_\epsilon = \gamma^2 I_{m\times m}$ with $\gamma \ge 100$, yields satisfactory performance in terms of estimation accuracy.

It is well known that the PM stage can be iterated to better enforce the constraint [12]. Regulating the tradeoff between estimation accuracy and computational complexity, the number of PM iterations is essentially application dependent.

To summarize the CS-embedded KF with the l_1 norm constraint (CSKF-1), we have provided Algorithm 1, listing a single iteration of the CS-embedded KF. This is an unusual implementation of the KF, as the matrix $\bar{H}_\tau$ is state-dependent.

Algorithm 1: CSKF-1

1: *Prediction*

$$\hat{z}_{k+1|k} = A\hat{z}_{k|k} \tag{9a}$$
$$P_{k+1|k} = AP_{k|k}A^T + Q_k. \tag{9b}$$

2: *Measurement Update*

$$K_k = P_{k+1|k}H'^T\left(H'P_{k+1|k}H'^T + R_k\right)^{-1} \tag{10a}$$
$$\hat{z}_{k+1|k+1} = \hat{z}_{k+1|k} + K_k\left(y_k - H'\hat{z}_{k+1|k}\right) \tag{10b}$$
$$P_{k+1|k+1} = (I - K_kH')P_{k+1|k}. \tag{10c}$$

3: *CS Pseudo Measurement:* Let $P^1 = P_{k+1|k+1}$ and $\hat{z}^1 = \hat{z}_{k+1|k+1}$.
4: **for** $\tau = 1, 2, \ldots, N_\tau - 1$ iterations **do**
5:

$$\bar{H}_\tau = [\text{sign}(\hat{z}^\tau(1)), \ldots, \text{sign}(\hat{z}^\tau(n))] \tag{11a}$$
$$K^\tau = P^\tau\bar{H}_\tau^T\left(\bar{H}_\tau P^\tau\bar{H}_\tau^T + R_\epsilon\right)^{-1} \tag{11b}$$
$$\hat{z}^{\tau+1} = (I - K^\tau\bar{H}_\tau)\hat{z}^\tau \tag{11c}$$
$$P^{\tau+1} = (I - K^\tau\bar{H}_\tau)P^\tau. \tag{11d}$$

6: **end for**
7: Set $P_{k+1|k+1} = P^{N_\tau}$ and $\hat{z}_{k+1|k+1} = \hat{z}^{N_\tau}$.

IV. Quasi-Norm Constrained Kalman Filtering

We have shown above that an approximate solution to the optimization problem (4) can be iteratively obtained using a KF augmented with a PM equation. This was accomplished by resorting to an equivalent convex formulation based on the l_1 norm, as suggested by the theory of CS. A different approach for approximately solving the combinatorial problem in (4) is to replace $\|\cdot\|_0$ by a quasi-norm $\|\cdot\|_p$ with $0 < p < 1$. This approach has already been shown to yield better accuracy compared to the l_1 norm [5].

Following the previous section's methodology, the PM technique is used here to incorporate the quasi-norm inequality constraint $\|z_k\|_p \le \epsilon'$ by producing the fictitious measurement

$$0 = \left(\sum_{i=1}^{n} |z_k(i)|^p\right)^{1/p} - \epsilon' \tag{12}$$

where ϵ' serves as a zero-mean Gaussian measurement noise with covariance R_ϵ. In practice, the above PM is linearized about some nominal state z_k^* to yield

$$0 = \left(\sum_{i=1}^{n} |z_k^*(i)|^p\right)^{1/p} + \bar{H}\delta z_k - \epsilon' + \mathcal{O}(\|\delta z_k\|_2^2) \tag{13}$$

where $\delta z_k := z_k - z_k^*$ and (14), shown at the bottom of the next page, is the ith element of $\bar{H}$. This formulation facilitates the implementation of an EKF stage for incorporating the PM. Following this, the nominal state z_k^* is set as the updated estimate at time k.

A. CS-Embedded KF With l_p Quasi-Norm Constraint (CSKF-p)

A single iteration of the resulting KF algorithm with the linearized PM stage is similar to Algorithm 1 with a slight modification in the PM implementation as described in Algorithm 2.

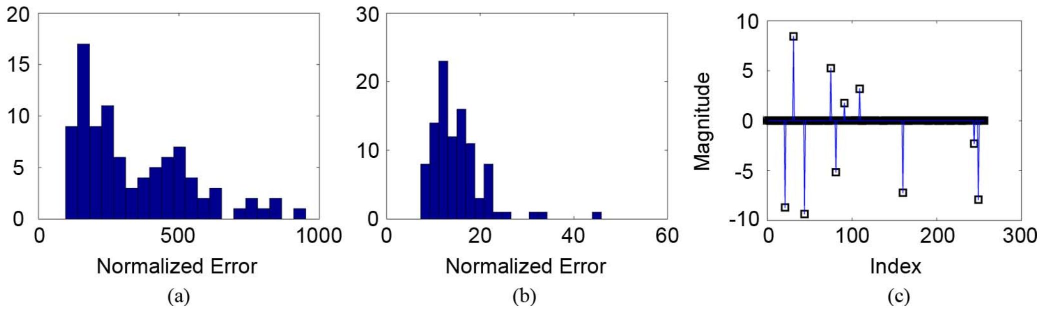


Fig. 1. Estimation performance of the CSKF-1 and the Dantzig selector. (a) Dantzig selector. (b) CSKF-1. (c) CSKF-1.

Algorithm 2: PM Stage of The CSKF-p

1: *Pseudo Measurement:* Let $P^1 = P_{k+1|k+1}$ and $\hat{z}^1 = \hat{z}_{k+1|k+1}$.
2: **for** $\tau = 1, 2, \ldots, N_\tau - 1$ iterations **do**
3: Compute $\bar{H}_\tau$ using (14) with $z_k^* = \hat{z}^\tau$.

$$K^\tau = P^\tau \bar{H}_\tau^T \left(\bar{H}_\tau P^\tau \bar{H}_\tau^T + R_\epsilon\right)^{-1} \tag{15a}$$

$$\hat{z}^{\tau+1} = \hat{z}^\tau - K^\tau \|\hat{z}^\tau\|_p \tag{15b}$$

$$P^{\tau+1} = (I - K^\tau \bar{H}_\tau)P^\tau. \tag{15c}$$

4: **end for**
5: Set $P_{k+1|k+1} = P^{N_\tau}$ and $\hat{z}_{k+1|k+1} = \hat{z}^{N_\tau}$.

B. Approximate l_0 Norm

The l_0 norm can alternatively be approximated by

$$\|z_k\|_0 = n - \sum_{i=1}^{n} \exp\left(-\alpha |z_k(i)|\right) \tag{16}$$

for large enough $\alpha > 0$. The corresponding PM stage in this case consists of the same steps (15) where (15b) is replaced by

$$\hat{z}^{\tau+1} = \hat{z}^\tau + K^\tau \left[n - \sum_{i=1}^{n} \exp\left(-\alpha|\hat{z}^\tau(i)|\right)\right] \tag{17}$$

(i.e., the PM is $n = \sum_{i=1}^{n} \exp\left(-\alpha|z_k(i)|\right) + \epsilon'$) where $\bar{H}$ is given by

$$\bar{H}(i) = \begin{cases} -\alpha \exp\left(-\alpha z_k^*(i)\right), & \text{if } z_k^*(i) > 0 \\ \alpha \exp\left(\alpha z_k^*(i)\right), & \text{if } z_k^*(i) \leq 0, \end{cases} \quad i = 1, \ldots, n. \tag{18}$$

V. Numerical Study

The performance of the CS-embedded KF algorithms is demonstrated using several examples in which sparse signals are recovered from a series of noisy observations. Two cases are examined: 1) recovery of a static parameter vector z, and 2) recovery of a stochastic process z_k of a fixed support. We shall first study the performance of the CSKF-1 algorithm and then compare the results to the CSKF-p algorithm.

Remark 1 ("Ordinary" KF Implementation): Implementing an ordinary KF without the PM extension for the following examples is useless as the underlying systems are unobservable.

A. Static Case—CSKF-1 Algorithm

The estimation performance of the CSKF-1 is demonstrated using a simple example similar to the one in [6]. The new algorithm is compared with the Dantzig Selector (DS) [6], which forms the core of the method in [10]. The DS was proposed as an ideal scheme by the authors of [1] for solving the CS problem (5). In this example, the signal $z \in \mathbb{R}^{256}$ is assumed to be a sparse parameter vector (i.e., $A = I_{256\times256}$, $Q_k = 0$). The signal support consists of total of 10 elements $z(i) \neq 0$ of which both the index and value are uniformly sampled over[2] $i \sim U_i[1, 256]$ and $z(i) \sim U[-10, 10]$, respectively. The sensing matrix $H \in \mathbb{R}^{72\times256}$ consists of entries that are sampled according to $\mathcal{N}(0, 1/72)$. This type of matrix has been shown to satisfy the RIP with overwhelming probability for sufficiently sparse signals (other random constructions that satisfy the RIP with overwhelming probability consist of either Bernoulli or Fourier entries; see [2] and [6] for further details). The observation noise covariance is set as $R_k = 0.01^2 I_{72\times72}$ where the tuning parameter $R_\epsilon = 200^2$. This example consists of a single observation, i.e., $k = 1$.

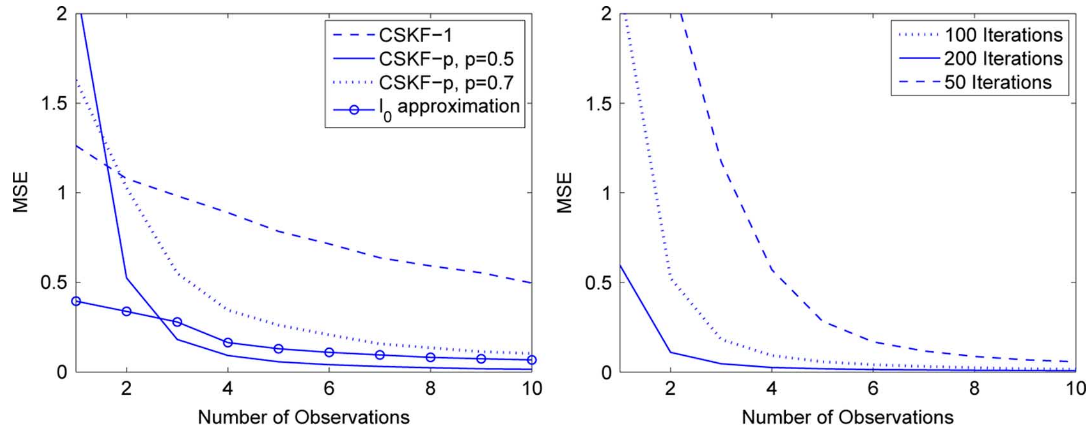
The distributions of the estimation errors over 100 Monte Carlo runs of both the DS and the CSKF-1 are depicted in Fig. 1(a) and (b), respectively. These figures show the histograms of the normalized errors, defined in [6] as

$$\frac{\|z - \hat{z}_k\|_2^2}{\sum_{i=1}^{n} \min\left(z(i)^2, \mathrm{Tr}(R_k)/m\right)} \tag{19}$$

where $\mathrm{Tr}(\cdot)$ denotes the matrix trace. Both histograms were obtained after approximately 1500 iterations. From these figures, it can be clearly recognized that the CSKF-1 outperforms the DS in terms of

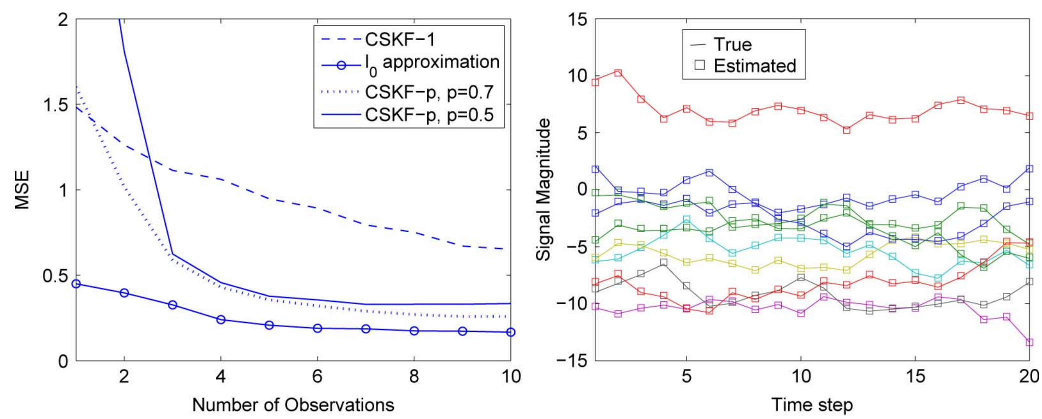
[2] $U_i[a, b]$ denotes a discrete uniform distribution of which the support constitutes all the integers in the interval $[a, b]$.

$$\bar{H}(i) = \begin{cases} \left(\sum_{i=1}^{n} |z_k^*(i)|^p\right)^{1/p-1} [z_k^*(i)]^{p-1}, & \text{if } z_k^*(i) > 0 \\ -\left(\sum_{i=1}^{n} |z_k^*(i)|^p\right)^{1/p-1} [-z_k^*(i)]^{p-1}, & \text{if } z_k^*(i) \leq 0, \end{cases} \quad i = 1, \ldots, n \tag{14}$$



(a) (b)

Fig. 2. Static case: (a) Mean square estimation error based on 50 runs of the CSKFs using 100 pseudo-measurement (PM) iterations. Showing the CSKF-1 (dashed line), the CSKF-p with $p = 0.7$ (dotted line), $p = 0.5$ (solid line), and a CSKF-p aided with the alternative l_0 approximation (circles). (b) Mean square estimation error based on 50 runs of the CSKF-p with $p = 0.5$ for various number of PM iterations. The lines, top to bottom, correspond to 50, 100, and 200 iterations. (a) MSE: CSKF variants comparison. (b) MSE: PM iterations.



(a) (b)

Fig. 3. Dynamic case: (a) Mean square estimation error based on 50 runs of the CSKFs using 100 pseudo-measurement (PM) iterations. Showing the CSKF-1 (dashed line), the CSKF-p with $p = 0.7$ (dotted line), $p = 0.5$ (solid line), and a KF aided by the l_0 approximation (circles). (b) A typical single run of the CSKF-1 for $N_\tau = 200$ PM iterations. Showing the true (lines) and estimated (squares) nonzero signals. (a) MSE: CSKF variants comparison. (b) Single run: Time history.

estimation accuracy. While the average normalized error of the DS is around 300, the CSKF-1 attains an average of approximately 16.

The actual (lines) and the CSKF-1 estimated (squares) signals are shown for a typical run in Fig. 1(c). In this figure, the magnitudes of the 256 entries of z and $\hat{z}_k$ are shown along the abscissa.

B. Static Case—CSKF-p Algorithm

The same static example of the previous subsection is used for comparing the variants, CSKF-1, CSKF-p and the l_0 approximate norm, using not more than ten observations (i.e., $k = 1, \ldots, 10$). In all runs, the tuning covariance R_ϵ of the CSKF-p is 1000^2 and 20000^2 for $p = 0.7$ and $p = 0.5$, respectively. The alternative l_0 approximation (16) is implemented using $\alpha = 1$ and $R_\epsilon = 100^2$ (these values were chosen based on tuning runs for achieving best performance in terms of accuracy).

The performance of the CSKF-p Algorithm is depicted by Fig. 2(a). This figure shows the mean square estimation error based on $N = 50$ Monte Carlo runs for various quasi-norms. It can be clearly seen that the best estimation accuracy is attained while using the quasi-norm l_p with $p = 0.5$. The alternative l_0 approximation is slightly less accurate but tends to converge faster. The CSKF-1 algorithm exhibits inferior estimation accuracy compared to the other variants.

In the current example, the CSKF-p Algorithm with $p = 0.5$ provides superior performance. However, the value of p is generally problem-dependant, and should be viewed as a tuning parameter along with the process and pseudo-measurement noise covariances. We already pointed out that a smaller p yields improved estimation accuracy, as the sparseness constraint tends to better approximate the l_0 norm [5]; in practice, however, decreasing p may deteriorate the filtering performance as the constraint becomes highly nonlinear.

Fig. 2(b) shows the effect of the number of PM iterations on the estimation performance of the CSKF-p algorithm with $p = 0.5$. Clearly, increasing the number of iterations reduces the mean square estimation error.

C. Dynamic Case

In this example, the sparse signal consists of exactly ten nonzero elements that behave as a random walk process. The driving white noise covariance of the elements in the support of z_k is taken as $Q = 1$. This process can be described by

$$z_{k+1}(i) = \begin{cases} z_k(i) + w_k(i),\ w_k(i) \sim \mathcal{N}(0, Q), & \text{if } z_k(i) \in \operatorname{supp}(z_k) \\ 0, & \text{otherwise} \end{cases} \tag{20}$$

where $i \sim U_i[1, 256]$ and $z_0(i) \sim U[-10, 10]$. The sensitivity matrix H' is chosen as in the static case and the observation noise covariance is taken as $R_k = 0.01^2 I_{72\times 72}$.

The estimation performance of the CSKF-p Algorithm in the dynamic case is presented in Fig. 3(a). This figure depicts the mean square estimation error based on $N = 50$ Monte Carlo runs for $N_\tau = 100$ PM iterations. As can be seen, the best estimation performance in this case is attained by the CSKF-p aided by the approximate l_0 norm (16). Again, the CSKF-1 exhibits the worst performance compared to the other filters. The attainable estimation errors in this case are slightly higher than in the static case. Nevertheless, it seems that the algorithms manage to adequately estimate the behavior of the nonzero processes as shown for the CSKF-1 in Fig. 3(b).

D. Application to Image Classification

The CSKF method was applied to image classification in [13]. The classifier derived in [13] utilizes a static version of the CSKF-1 (i.e., with $A = I_{n\times n}$ and $Q_k = 0$) for learning a sparse random field model individually for each class. The new classifier, which is applied for prediction of mental states based on fMRI scans, is shown to outperform other schemes that rely on state-of-the-art methods such as naive Bayes and LASSO.

VI. Conclusion

We have presented simple Kalman-filtering-based algorithms for sparse signal recovery from a series of noisy observations. The proposed methods utilized a pseudo-measurement technique for enforcing an l_p norm constraint. In the first approach, the Kalman filter was augmented by an l_1 norm constraint, which retained the linearity of the filter. The simplicity of this method stems from the fact that no considerable modifications are required in the basic Kalman filter formulation. In the second approach, a quasi-norm constraint was used, giving rise to an extended Kalman filter. Both formulations were inspired by recent results in compressed sensing theory, albeit the new algorithms are much simpler for implementation than existing algorithms. The simulation study showed that the static version of the l_1-constrained filter outperforms the Dantzig selector, which was suggested as an ideal scheme for solving the compressed sensing problem. Moreover, it was shown that using a quasi-norm-constrained filter outperforms the l_1-constrained filter both in the static and dynamic settings.

References

[1] E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.
[2] E. J. Candes, "Compressive sampling," presented at the Int. Congr. Mathematicians, Madrid, Spain, 2006.
[3] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magn. Resonance Med.*, vol. 58, pp. 1182–1195, 2007.
[4] U. Gamper, P. Boesiger, and S. Kozerke, "Compressed sensing in dynamic MRI," *Magn. Resonance Med.*, vol. 59, pp. 365–373, 2008.
[5] R. Chartrand, "Exact reconstruction of sparse signals via nonconvex minimization," *IEEE Signal Process. Lett.*, vol. 14, no. 10, pp. 707–710, Oct. 2007.
[6] E. Candes and T. Tao, "The Dantzig selector: Statistical estimation when p is much larger than n," *Ann. Stat.*, vol. 35, pp. 2313–2351, 2007.
[7] R. Tibshirani, "Regression shrinkage and selection via the LASSO," *J. Roy. Stat. Soc. B (Methodologic.)*, vol. 58, no. 1, pp. 267–288, 1996.
[8] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 1998.
[9] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *Ann. Stat.*, vol. 32, no. 2, pp. 407–499, 2004.
[10] N. Vaswani, "Kalman filtered compressed sensing," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, Oct. 2008, pp. 893–896.
[11] G. M. James, P. Radchenko, and J. Lv, "DASSO: Connections between the Dantzig selector and LASSO," *J. Roy. Stat. Soc.*, vol. 71, pp. 127–142, 2009.
[12] S. J. Julier and J. J. LaViola, "On Kalman filtering with nonlinear equality constraints," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2774–2784, 2007.
[13] A. Carmi, P. Gurfil, D. Kanevsky, and B. Ramabhadran, "ABCS: Approximate Bayesian compressed sensing," Human Language Technologies, IBM, RC24816, 2009.

On Entropy Rate for the Complex Domain and Its Application to i.i.d. Sampling

Wei Xiong, Hualiang Li, Tülay Adalı, *Fellow, IEEE*, Yi-Ou Li, and Vince D. Calhoun

***Abstract*—We derive the entropy rate formula for a complex Gaussian random process by using a widely linear model. The resulting expression is general and applicable to both circular and noncircular Gaussian processes, since any second-order stationary process can be modeled as the output of a widely linear system driven by a circular white noise. Furthermore, we demonstrate application of the derived formula to an order selection problem. We extend a scheme for independent and identically distributed (i.i.d.) sampling to the complex domain to improve the estimation performance of information-theoretic criteria when samples are correlated. We show the effectiveness of the approach for order selection for simulated and actual functional magnetic resonance imaging (fMRI) data that are inherently complex valued.**

***Index Terms*—Complex-valued signal processing, entropy rate, order selection.**

I. Introduction

Entropy rate is a measure of the average information carried by each sample in a random sequence, and is a useful measure with applications in spectral estimation, coding, and biomedical signal processing among others [1], [2]. For example, it has been used to quantify heart variability for anomaly detection [3] and to determine independent samples in functional magnetic resonance data [4].

The entropy rate formula for a real-valued stationary Gaussian random process is developed by Kolmogorov [2], and a practical derivation using a linear model is given by Papoulis in [1]. As complex-valued signals arise typically in many signal processing problems such as communications and medical imaging, it is desirable to extend the formula to the complex domain. The representation can be done

Manuscript received February 03, 2009; accepted December 19, 2009. First published January 12, 2010; current version published March 10, 2010. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Danilo P. Mandic. This work is supported by the NSF Grants NSF-IIS 0612076 and NSF-CCF 0635129.

W. Xiong, H. Li, T. Adalı, and Y.-O. Li are with the Department of Computer Science and Electrical Engineering, University of Maryland Baltimore County, Baltimore, MD 21250, USA (e-mail: xiongw1@umbc.edu; lihua1@umbc.edu; adali@umbc.edu; liyiou1@umbc.edu).

V. D. Calhoun is with the Department of Electrical and Computer Engineering, University of New Mexico, and the MIND Research Network, Albuquerque, NM 87131 USA (e-mail: vcalhoun@unm.edu).

Color versions of one or more of the figures in this correspondence are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2010.2040411