RC24711 (W0812-084) December 15, 2008
Computer Science

IBM Research Report

A Simple Method for Sparse Signal Recovery from Noisy Observations Using Kalman Filtering. Embedding Approximate Quasi-Norms for Improved Accuracy

Avishy Carmi
Cambridge University
UK

Pini Gurfil
Technion
Haifa, Israel

Dimitri Kanevsky
IBM Research Division
Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598
USA

Research Division
Almaden - Austin - Beijing - Cambridge - Haifa - India - T. J. Watson - Tokyo - Zurich

Abstract

In a primary research report we have introduced a simple method for recovering sparse signals from a series of noisy observations using a Kalman filter (KF). The new method utilizes a so-called pseudo-measurement technique for optimizing the convex minimization problem following from the theory of compressed sensing (CS). The CS-embedded KF approach has shown promising results when applied for reconstruction of linear Gaussian sparse processes.

This report presents an improved version of the KF algorithm for the recovery of sparse signals. In this work we substitute the l_1 norm by an approximate quasi-norm l_p, $0 \leq p < 1$. This modification, which better suits the original combinatorial problem, greatly improves the accuracy of the resulting KF algorithm. This, however, involves the implementation of an extended KF (EKF) stage for properly computing the state statistics.

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

$$z_k = \psi^T x_k, \quad |\text{supp}(z_k)| << n \tag{1}$$

where $\text{supp}(z_k)$ denotes the support of z_k. Assume that z_k evolves according to

$$z_{k+1} = Az_k + w_k \tag{2}$$

where $A \in \mathbb{R}^{n\times n}$ and $\{w_k\}_{k=1}^{\infty}$ is a zero-mean white Gaussian sequence with covariance Q_k. The process x_k is measured using a sequence of noisy observations given by

$$y_k = Hx_k + \zeta_k = H' z_k + \zeta_k \tag{3}$$

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

Letting $y^k := [y_1, \dots, y_k]$, our problem is defined as follows. We are interested in a y^k-measurable estimator $\hat{x}_k$ such that the minimum mean square error (MMSE) $E\left[\| x_k - \hat{x}_k \|_2^2\right]$ is minimized. It is well known that if the above linear system (2),(3) is observable, i.e.,

$$\mathscr{O} := \begin{bmatrix} H \\ HA \\ \vdots \\ HA^{n-1} \end{bmatrix}, \quad \text{rank}(\mathscr{O}) = n \tag{4}$$

then the solution to this problem can be obtained using the Kalman filter (KF). On the other hand, if the system is unobservable then the regular KF algorithm is useless in this case.

When the above system is unobservable, if for instance $A = I_{n\times n}$, then it may look hopeless to try reconstructing x_k from an under-determined system in which $\text{rank}(H) < n$. Surprisingly, this problem may be circumvented by taking into account the fact that z_k is sparse.

The Combinatorial Problem and Compressed Sensing

It has already been shown that in the deterministic case (i.e., when z is a parameter vector) one can recover z (and therefore also x, i.e., $x = \psi z$) with high accuracy by solving the optimization problem [1, 2]

$$\min \| \hat{z} \|_0 \ \text{ s.t. } \sum_{i=1}^{k} \| y_i - H'\hat{z} \|_2^2 \leq \epsilon \tag{5}$$

for sufficiently small ϵ. Following a similar approach, in the stochastic case it can be shown that the sought-after optimal estimator satisfies

$$\min \| \hat{z}_k \|_0 \ \text{ s.t. } E_{z_k|y^k}\left[\| z_k - \hat{z}_k \|_2^2\right] \leq \epsilon \tag{6}$$

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 a so-called *restricted isometry hypothesis* (RIH) then the solution of the combinatorial problem (5) can almost always be obtained by solving the convex optimization [1, 3]

$$\min \| \hat{z} \|_1 \ \text{ s.t. } \sum_{i=1}^{k} \| y_i - H'\hat{z} \|_2^2 \leq \epsilon \tag{7}$$

This is a fundamental result in a new emerging theory known as compressed sensing (CS) [1,3]. The essential idea here is that the convex l_1 minimization problem can be efficiently solved. Another insights provided by CS are related to the construction of sensitivity matrices that satisfy the RIH. These underlying matrices are random by nature which in turn has shed a new light on the way observations should be sampled. For an extensive review of CS the reader is referred to [1, 3].

Quasi-Norm Constrained Kalman Filtering

As we have shown in [4] an approximate solution to the optimization problem in (6) can be iteratively obtained using a pseudo-measurement (PM) aided Kalman filtering (KF). This is accomplished by resorting to an equivalent convex formulation based on the l_1 norm following the theory of CS. A rather different approach for approximately solving the combinatorial problem in (6) is based on replacing $\| \cdot \|_0$ 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 [2].

Following the argument in [4], the PM technique is used here to incorporate the quasi-norm inequality constraint

$$\| z_k \|_p \leq \epsilon' \tag{8}$$

by producing the fictitious measurement

$$0 = \| z_k \|_p - \epsilon' \tag{9}$$

where ϵ' serves as a zero-mean Gaussian measurement noise with covariance R_ϵ. Equation (9) can be written explicitly as

$$0 = \left(\sum_{i=1}^{n} |z_k(i)|^p \right)^{1/p} - \epsilon' \tag{10}$$

where $z_k(i)$ denotes the ith element of z_k. In practice, the above PM is linearized around some nominal state z_k^* to yield

$$0 = \left(\sum_{i=1}^{n} |z_k^*(i)|^p \right)^{1/p} + \bar{H} \delta z_k - \epsilon' + \mathcal{O}(\| \delta z_k \|_2^2) \tag{11}$$

where $\delta z_k := z_k - z_k^*$ and

$$\bar{H}(i) = \begin{cases} \left(\sum_{i=1}^{n} |z_k^*(i)|^p\right)^{1/p-1} [z_k^*(i)]^{p-1}, & \text{if } z_k^*(i) > 0 \\ -\left(\sum_{i=1}^{n} |z_k^*(i)|^p\right)^{1/p-1} [-z_k^*(i)]^{p-1}, & \text{if } z_k^*(i) \leq 0 \end{cases}, \qquad i = 1, \dots, n \tag{12}$$

is the ith element of $\bar{H}$. This formulation facilitates the implementation of an extended KF (EKF) stage for incorporating the PM. Following this, the nominal state z_k^* is set as the updated estimate at time k.

Algorithm Summary

A single iteration of the resulting KF algorithm with the linearized PM stage is summarized below.

1. *Prediction*

$$\hat{z}_{k+1|k} = A \hat{z}_{k|k} \tag{13a}$$

$$P_{k+1|k} = A P_{k|k} A^T + Q_k \tag{13b}$$

2. *Measurement Update*

$$K_k = P_{k+1|k} H'^T \left(H' P_{k+1|k} H'^T + R_k \right)^{-1} \tag{14a}$$

$$\hat{z}_{k+1|k+1} = \hat{z}_{k+1|k} + K_k \left(y_k - H' \hat{z}_{k+1|k} \right) \tag{14b}$$

$$P_{k+1|k+1} = (I - K_k H') P_{k+1|k} \tag{14c}$$

3. *Pseudo Measurement:* Let $P^1 = P_{k+1|k+1}$ and $\hat{z}^1 = \hat{z}_{k+1|k+1}$. For $\tau = 1, 2, \dots, N_\tau - 1$ iterations do the following. Compute $\bar{H}_\tau$ using (12) 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. Set $P_{k+1|k+1} = P^{N_\tau}$ and $\hat{z}_{k+1|k+1} = \hat{z}^{N_\tau}$.

Approximate l_0 Norm

The l_0 norm can alternatively be approximated by

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

for large enough $\alpha > 0$. The corresponding PM stage in that 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

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

Numerical Study

The performance of the quasi-norm constrained KF algorithm is demonstrated using several simple 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. In both cases the new KF algorithm is compared to the CS-embedded KF in [4].

Static Case

In this example the signal $z \in \mathbb{R}^{256}$ is assumed to be a sparse parameter vector (i.e., $A = I_{256\times 256}$, $Q_k = 0$). The signal support consists of total of 10 elements $z(i) \neq 0$ of which both the index and value are uniformly sampled over $i \sim U_i[1, 256]$ [1] and $z(i) \sim U[-10, 10]$, respectively. The sensitivity matrix H' consists of 72 rows in which the elements are sampled from a Guassian distribution. The columns of H' are normalized following the example in [5] (this matrix has been shown to satisfy the RIH, see [3, 5]). The observation noise covariance is set as $R_k = 0.001^2 I_{72\times 72}$. In all runs the tuning covariance R_ϵ of the quasi-norm KFs was set as 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 ideal performance in terms of accuracy).

The estimation performance of the quasi-norm constrained KF algorithm is presented in Fig. 1. This figure depicts the mean square estimation error based on $N = 50$ Monte Carlo runs for various quasi-norm approximations. Thus, 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. Unsurprisingly, the CS-embedded KF of [4] exhibits inferior estimation accuracy compared to the other KF algorithms.

Figure 2 shows the effect of the number of PM iterations on the estimation performance of the quasi-norm constrained KF with $p = 0.5$.

Dynamic Case

In this example the sparse signal consists of exactly 10 non-zero 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

$$z_{k+1}(i) = \begin{cases} z_k(i) + w_k(i), \quad w_k(i) \sim \mathcal{N}(0, Q), & \text{if } z_k(i) \in \text{supp}(z_k) \\ 0, & \text{otherwise} \end{cases} \tag{19}$$

where $i \sim U_i[1, 256]$ and $z_0(i) \sim U[-10, 10]$ as before. The sensitivity matrix H' is set 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 quasi-norm constrained KF algorithm in the dynamic case is presented in Fig. 3. This figure depicts the mean square estimation error based on $N = 50$ Monte

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

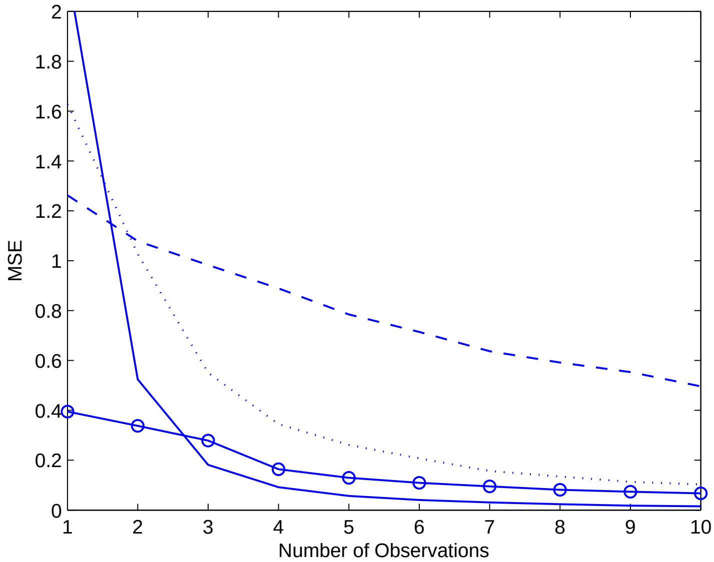

Figure 1. Mean square estimation error based on 50 runs of the sparse KFs using 100 PM iterations. Showing the CS-embedded KF (dashed line), the quasi-norm constrained KF with $p = 0.7$ (dotted line), $p = 0.5$ (solid line), and a KF aided with the alternative l_0 approximation (circles). Static case.

Carlo runs for $N_\tau = 100$ PM iterations. As it can be seen, the best estimation performance in this case is attained by the KF that is aided with the approximate l_0 norm (16). Unexceptionally, the CS-embedded KF exhibits the worst performance compared to the other filters.

Figure 4 shows the effect of the number of PM iterations on the estimation performance of the KF using the approximate l_0 norm.

Conclusions

We have presented a quasi-norm constrained Kalman filtering-based method for sparse signal recovery from a series of noisy observations. This algorithm is an improved version of the CS-embedded KF that was derived in our previous work. It is shown via simulations that the new algorithm outperforms the CS-embedded KF in both static and dynamic scenarios.

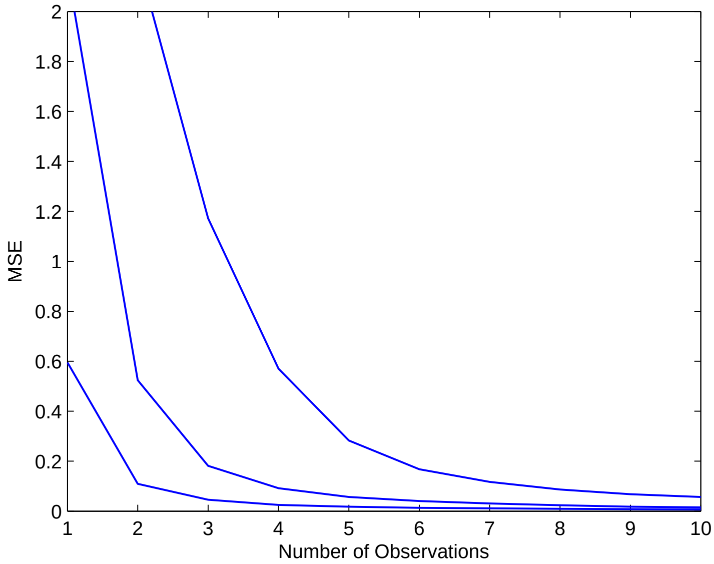

Figure 2. **Mean square estimation error based on 50 runs of the quasi-norm constrained KF with $p = 0.5$ for various number of PM iterations. The lines top to bottom correspond to 50, 100 and 200 PM iterations. Static case.**

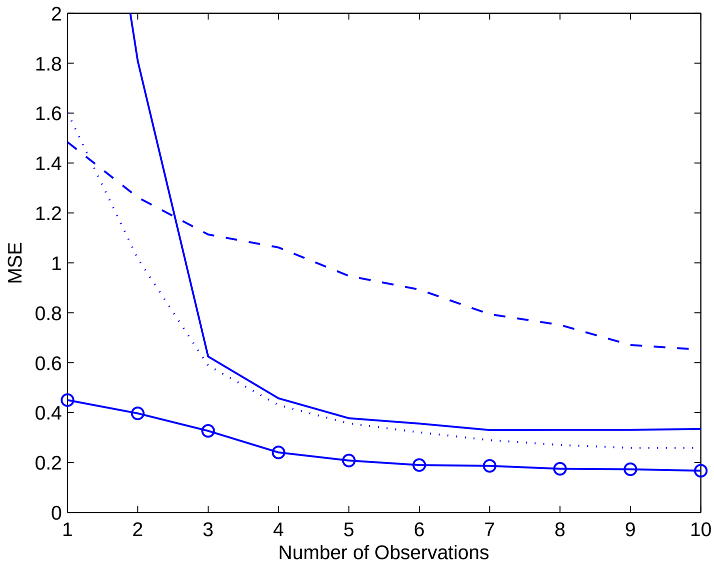

Figure 3. **Mean square estimation error based on 50 runs of the sparse KFs using 100 PM iterations. Showing the CS-embedded KF (dashed line), the quasi-norm constrained KF with $p = 0.7$ (dotted line), $p = 0.5$ (solid line), and a KF aided with the alternative l_0 approximation (circles). Dynamic case.**

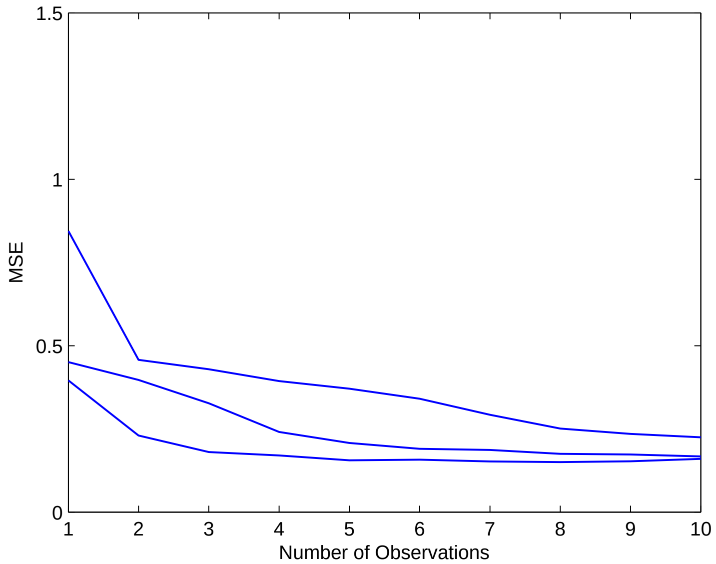

Figure 4. Mean square estimation error based on 50 runs of the KF using the approximate l_0 norm for various number of PM iterations. The lines top to bottom correspond to 50, 100 and 200 PM iterations. Dynamic case.

Bibliography

[1] Candes, E. J., Romberg, J., and Tao, T., "Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information," *IEEE Transactions on Information Theory*, Vol. 52, 2006.

[2] Chartrand, R., "Exact Reconstruction of Sparse Signals via Nonconvex Minimization," *IEEE Signal Processing Letters*, Vol. 14, 2007, pp. 707–710.

[3] Candes, E. J., "Compressive Sampling," European Mathematical Society, Proceedings of the International Congress of Mathematicians, Madrid, Spain, 2006.

[4] Carmi, A., Gurfil, P., and Kanevsky, D., "A Simple Method for Sparse Signal Recovery from Noisy Observations Using Kalman Filtering," Tech. Rep. RC24709, Human Language Technologies, IBM, 2008.

[5] Vaswani, N., "Kalman Filtered Compressed Sensing," Proceedings of the IEEE International Conference on Image Processing (ICIP), October 2008.