UNSCENTED COMPRESSED SENSING

Avishy Y. Carmi¹, Lyudmila Mihaylova², Dimitri Kanevsky³

 ¹ School of Mechanical and Aerospace Engineering Nanyang Technological University, Singapore
 ² School of Computing and Communications Lancaster University, United Kingdom
 ³ IBM T. J. Watson, Yorktown, NY, USA

ABSTRACT

In this paper we present a novel compressed sensing (CS) algorithm for the recovery of compressible, possibly time-varying, signal from a sequence of noisy observations. The newly derived scheme is based on the acclaimed unscented Kalman filter (UKF), and is essentially self reliant in the sense that no peripheral optimization or CS algorithm is required for identifying the underlying signal support. Relying exclusively on the UKF formulation, our method facilitates sequential processing of measurements by employing the familiar Kalman filter predictor corrector form. As distinct from other CS methods, and by virtue of its pseudomeasurement mechanism, the CS-UKF, as we termed it, is non iterative, thereby maintaining a computational overhead which is nearly equal to that of the conventional UKF.

Index Terms— Compressed sensing, Kalman filter, Unscented Kalman Filter, Sparse signal recovery, Sigma point filter

1. 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]. The essence of the new theory builds upon a new data acquisition formalism, in which compression plays a fundamental role. From a signal processing 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 Other applications consider model-reduction methods to enforce sparseness for preventing over-fitting and for reducing computational complexity and storage capacities.

The recovery of sparse signals is in general NP-hard [1]. State-of-the-art methods for addressing this optimization problem commonly utilize convex relaxations, non-convex local optimization and greedy search mechanisms. Convex relaxations are used in various methods such as LASSO, the Dantzig selector, basis pursuit and basis pursuit de-noising, and least angle regression. Non-convex optimization approaches include Bayesian methodologies such as the relevance vector machine, otherwise known as sparse Bayesian learning. Notable greedy search algorithms are the matching pursuit (MP), the orthogonal MP, and the orthogonal least squares.

CS theory has drawn much attention to the convex relaxation methods. It has been shown that the convex l_1 relaxation yields an

exact solution to the recovery problem provided two conditions are met: 1) the signal is sufficiently sparse, and 2) the sensing matrix obeys the so-called restricted isometry property at a certain level. A complementary result guarantees high accuracy when dealing with noisy observations, yielding recovery 'with overwhelming probability'. To put it informally, it is likely for the convex l_1 relaxation to yield an exact solution provided that the involved quantities, the sparseness degree s, and the sensing matrix dimensions $m \times n$ maintain relation of the type $s = O(m/\log(n/m))$.

The Bayesian CS (BCS) approach has been introduced in [2]. As opposed to the conventional non-Bayesian methods, the Bayesian CS has the advantage of providing the complete statistics of the estimate in the form of a posterior probability density function (pdf). Adopting this approach, however, suffers from the fact that rarely one can obtain a closed-form expression of the posterior and therefore approximation methods should be utilized.

1.1. Sequential and Dynamic CS

The basic CS framework is mainly concerned with parameter estimation, or time-invariant signals. A tremendous research effort is yet being made for developing efficient CS techniques that would be able to perform in high dimensional non dynamic settings. Only recently, CS has been applied for the recovery of time-varying sparse signals (i.e., sparse random processes). There is no wonder why there is such unbalance between the two realms of nondynamic and dynamic CS. The fundamentals of CS build upon convex optimization perspectives and as such it is conventionally assumed that the measurements are available in a batch form. This obviously restricts the theory to such signals for which the complexity does not considerably increase over time. Furthermore, the treatment of process dynamics, which are normally governed by probabilistic transition kernels, is not a straightforward task as far as optimization approaches are concerned.

In light of the above, a much more practical approach for treating dynamic sparse signals would be somehow based on state filtering methodologies. Followed by the pioneering works of Vaswani [3], and Carmi et al [4, 5], which show how the Kalman filter (KF) can be used in this respect, several dynamic CS schemes have been proposed over the last two years. Thus, the work in [6] derives a l_1 -regularized recursive least squares estimator. This type of estimator is capable of dealing with dynamic signals and support variations via the use of a "forgetting factor". In other works, the LASSO is amended for performing in dynamic settings with possible abrupt changes in the signal support [7, 8].

The KF algorithm constitutes a vital part in the works of [9], [10], and [11]. Indeed, the KF is elegant and simple and above all is the *linear* optimal minimum mean square error (MMSE) estimator irrespective of noise statistics. Despite its appealing features, rarely it is used in its standard formulation which is primarily designed for linear time-varying models. Modifying the KF structure and extending its capabilities have already become a common practice in many engineering and scientific fields. The resulting KF-based methods are vastly used for nonlinear filtering, constrained state estimation, distributed estimation, learning in neural networks, and fault-tolerant filtering.

The KF-based methodologies for dynamic CS can be divided into two broad classes: hybrid, and self-reliant. Whereas the former class refers to KF-based approaches involving the utilization of peripheral optimization schemes for handling sparseness and support variations, the latter class refers to methods that are entirely independent of any such scheme. Hybrid KF-based approaches refer to works such as [3, 9–11]. The only self-reliant KF method available to that end is the one of [4, 5].

The self-reliant KF method in [5] benefits from ease of implementation. It avoids intervening in the KF process which thereby maintains the filtering statistics as adequate as possible. The key idea behind it is to apply the KF in constrained filtering settings using the so-called pseudo-measurement technique. It may, however, exhibit an inferior performance when improperly tuned or when insufficient number of iterations had been carried out. In this work, we improve over [5] by employing the unscented KF (UKF) [12] for the pseudo-measurement update stage.

1.2. Why Unscented Compressed Sensing

The resulting UKF-based CS algorithm has the following benefits: 1) Self-reliant and easy to implement, 2) Recursively updates the mean and covariance of the filtering probability density function (pdf), 3) Facilitates sequential processing of measurements, 4) Non iterative - as opposed to [5] no reiterations are needed at any stage, 4) Its computational complexity is nearly equal to that of a standard UKF.

2. 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, $z_k = \psi^T x_k$, $s_k := ||z_k||_0 < n$, where $||z_k||_0$ denotes the cardinality of the signal support at time k, i.e., the number of non vanishing entries in z_k . Assume that z_k evolves according to

$$z_{k+1} = Az_k + w_k, \quad z_0 \sim \mathcal{N}(\mu_0, P_0)$$
 (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 $Q_k \succeq 0$. The signal x_k is measured by the \mathbb{R}^m -valued random process

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

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

Letting $\mathscr{Y}_k := [y_1, \ldots, y_k]$, our problem is defined as follows. We are interested in finding a \mathscr{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) $E [|| x_k - \hat{x}_k ||_2^2]$. 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 $\operatorname{rank}(H) < n$.

2.1. Compressed Sensing

It is well known 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 [1]

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

for a sufficiently small ϵ .

Following a similar rationale, in the stochastic case the soughtafter optimal estimator satisfies

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

Unfortunately, the above optimization problems are NP-hard and cannot be solved efficiently. Surprisingly enough, it has been shown in [1] that if the underlying signal is sufficiently sparse and the sensing matrix H' obeys the so-called *restricted isometry* property (RIP) to within a certain tolerance then the solution of the combinatorial problem (3) can almost always be obtained by solving the constrained convex relaxation

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

This is a fundamental result in CS theory [1]. The main idea is that the convex l_1 minimization problem can be efficiently solved using a myriad of existing methods. Additional 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.

3. UNSCENTED CS

The derivation of the improved Kalman filtering algorithm in this section is based on the notion of pseudo-measurement (PM) [5]. The key idea is fairly simple and has been vastly employed for constrained state estimation. Thus, instead of solving the l_1 -relaxed variant of the constrained problem (4), a non constrained minimization

$$\min_{\hat{z}} E_{z_k \mid \mathscr{Y}_k} \left[\parallel z_k - \hat{z}_k \parallel_2^2 \right] \tag{6}$$

is considered with the observation set \mathscr{Y}_k augmented by an additional fictitious measurement satisfying

$$0 = \|z_k\|_1 - v_k \tag{7}$$

where v_k is a Gaussian random variable with some predetermined mean and variance, μ_k and r_k , respectively. The above PM is in essence the stochastic analogous of the l_1 constrain imposed by the underlying dual problem [13].

The role of (7) can be better apprehended by realizing that v_k is aimed to capture the first two statistical moments of the random variable moderating the sparseness degree, $||z_k||_1$. In general, computation of the exact statistics of $||z_k||_1$ may be analytically intractable and therefore either approximations or tuning procedures should be utilized for determining preferred values for the mean and variance of v_k . We note, however, that the resulting method is rather robust to the underlying parameters as demonstrated in our previous work [5].

3.1. Adaptive Pseudo-Measurement Approximation

Equation (7) cannot be straightforwardly processed in the framework of Kalman filtering as it is nonlinear. In practice, this equation is substituted with the following approximation [5]

$$0 = \operatorname{sign}(\hat{z}_k)^T z_k - \bar{v}_k \tag{8}$$

where sign $(\hat{z}_k) \in \mathbb{R}^n$ is a \mathscr{Y}_k -measurable vector comprising of the sign of entries in \hat{z}_k . The effective measurement noise \bar{v}_k is zero-mean Gaussian with a variance obeying

$$\bar{r}_k = \mathcal{O}\left(\|\hat{z}_k\|_2^2 + g^T P_k g\right) + r_k \tag{9}$$

where $g \in \mathbb{R}^n$ is some (tunable) constant vector, and P_k is the estimation error covariance of the supposedly unbiased estimator \hat{z}_k , that is $P_k = E\left[(z_k - \hat{z}_k)(z_k - \hat{z}_k)^T\right]$. For brevity, the proof of the conjecture (9) is omitted here, and would be provided elsewhere.

3.2. Sigma Point Filtering

The UKF and its variants, which are broadly referred to as *sigma point Kalman filters*, parameterize the filtering pdf via the first two statistical moments, namely the mean and covariance, thus providing an approximation to the optimizer of (6) (i.e., the conditional mean). These methods amend the Kalman filter algorithm for handling generalized nonlinear process and measurement models. As distinct from the extended KF (EKF) which employs infamous linearization techniques, the UKF relies on the so-called unscented transformation (UT), otherwise known as statistical linearization. This approach is acclaimed for its ease of implementation and its improved estimation performance owing to a rather adequate computation of the underlying covariance matrix. By virtue of its mechanism, the UKF alleviates filtering inconsistencies which in most cases results in improved robustness to model nonlinearities and initial conditions.

The UT can be readily understood by considering a simple example. Let $z \sim \mathcal{N}(\mu, \Sigma)$ be a random vector of dimension n, and let also $f(\cdot) : \mathbb{R}^n \to \mathbb{R}^m$ be some function. Suppose that we are interested in computing the mean and covariance of f(z) to within a certain accuracy. It turns out that a fairly reasonable approximation of these quantities can be made by carefully choosing a finite set of L instrumental vectors $Z^i \in \mathbb{R}^n$, $i = 1, \ldots, L$, and corresponding weights w^i . The UT essentially provides a convenient deterministic mechanism for generating L = 2n + 1 such points which are known by the name sigma points. As Σ is a symmetric matrix it can be decomposed as $\Sigma = DD^T$ (e.g., Cholesky decomposition). The sigma points are then given as

$$Z^{i} = \mu + \sqrt{L}D^{i}, \qquad i = 0, \dots, n$$

$$Z^{i} = \mu - \sqrt{L}D^{L-i}, \qquad i = n+1, \dots, 2n$$
(10)

where D^i denotes the *i*th column of D, and $D^0 := 0$. Note that the sample mean and sample covariance of Z^i , i = 0, ..., 2n, are μ and Σ , respectively (i.e., this set of points captures the statistics of z). Now, the mean and covariance of f(z) can be approximated by

$$\hat{\mu}_f = \sum_{i=0}^{2n} w^i f(Z^i), \quad \hat{\Sigma}_f = \sum_{i=0}^{2n} w^i f(Z^i) f(Z^i)^T - \hat{\mu}_f \hat{\mu}_f^T \quad (11)$$

3.3. CS-UKF: Compressible Sigma Point Filter

In this work, we amend the UKF for handling sparse and compressible signals. The resulting algorithm, the CS-UKF as we termed it, is a Bayesian CS algorithm that is capable of estimating compressible dynamic signals sequentially in time. The sparseness constraint is imposed in a manner similar to [5] via the use of the PM approximation (8), however, without the need for reiterating the PM update. This in turn maintains a computational overhead similar to that of a standard UKF.

The CS-UKF consists of the two traditional UKF stages, the prediction and update, along with an additional refinement stage during which the sigma points gradually become compressible. In particular, after a single standard UKF cycle the sigma points are individually updated in a manner similar to the PM update stage in [5].

Let $P_{k|k}$ and $Z_{k|k}^i$ be the updated covariance and the *i*th sigma point at time k, respectively (i.e., after the measurement update). A set of compressible sigma points at time k is thus given by

$$\bar{Z}_{k|k}^{i} = Z_{k|k}^{i} - \frac{P_{k|k} \operatorname{sign}(Z_{k|k}^{i}) \|Z_{k|k}^{i}\|_{1}}{\operatorname{sign}(Z_{k|k}^{i})^{T} P_{k|k} \operatorname{sign}(Z_{k|k}^{i}) + \bar{r}_{k}^{i}}$$
(12)

$$\bar{r}_{k}^{i} = \alpha(\|Z_{k|k}^{i}\|_{2}^{2} + g^{T}P_{k}g) + r_{k}$$
(13)

for i = 0, ..., 2n, where α is some positive tuning parameter. Once the set $\{\overline{Z}_{k|k}^i\}_{i=0}^{2n}$ is obtained, its sample mean and sample covariance (see (11)) substitutes the updated mean and covariance of the UKF at time k.

We note that if the process and measurement models are linear then the prediction and update stages of the UKF can be substituted with those of the standard KF. In this case, the resulting CS-UKF algorithm would consist of the KF prediction and update stages together with a UKF-based PM refinement phase.

4. NUMERICAL STUDY

The CS-UKF is employed for tracking a time-invariant compressible signal with n = 200 entries out of which only 10 are significant in terms of their magnitude. The performance of the CS-UKF is compared with that of the CSKF [5], and of the BCS [2]. As the latter method, the BCS, is non sequential, in our experiments it is fed at any give time with the whole batch of measurements available up to the specific instance. Both KF variants process a single measurement at a time. The negligible entries of the signal are uniformly sampled from the interval [-0.02, 0.02].

The recovery performance based on 100 Monte Carlo (MC) runs of all methods over a time interval of 100 steps is depicted in Fig. 1. Thus, the MC-computed mean estimation errors and standard deviations (summed up for all entries) of the BCS and the CSKF are shown in the left panel in this figure. The KF-computed standard deviations are provided by the dashed lines. The right panel in this figure compares the same quantities, however for the BCS and the CS-UKF. The striking observation from this figure is that the mean performance of the BCS and the KF-based CS approaches nearly coincide. The KF-based approaches, however, process a single measurement at a time which thereby maintains around the same (low) computational overhead at every time step. The BCS, on the other hand, deals with an increased complexity over time. As opposed to the CSKF which employs 40 PM iterations and covariance updates per measurement, the CS-UKF uses only one iteration per sigma point and two covariance updates per measurement (which accounts for measurement update and PM refinement).

Not least important, Fig. 1 shows that both the CSKF and the CS-UKF are nearly consistent in the sense that their computed variances are greater than the MC-computed ones. In this respect, it seems that the CS-UKF manages to capture fairly accurately the actual estimation error variance from around the 65 time step, as

its computed standard deviations approximately concur with the MC-computed ones from that point onwards.

The recovery performance of the 3 methods is shown for various signal compressibility degrees in Fig. 2. The upper left and upper right panels in this figure show the normalized root MSE (RMSE) (i.e., $||z_k - \hat{z}_k||_2/||z_k||_2$) based on 20 MC runs of the 3 CS methods. The upper left panel illustrates the recovery performance in a compressible setting where the negligible entries of the signal are uniformly sampled from [-0.1, 0.1]. The upper right panel, on the other hand, depicts the recovery performance when the signal is perfectly sparse. Both these panels clearly show that the estimation accuracy of all methods nearly coincide after roughly 95 time steps (i.e., 95 observations). In the compressible case the KF methods attain accuracy which is either equal or better than that of the BCS.

This trend is further manifested in the bottom panel in Fig. 2 where the mean normalized RMSE (i.e., averaged over the entire running time) is shown for various compressibility levels. The negligible entries of the underlying compressible signal are uniformly sampled from an expanding interval $[-\epsilon, \epsilon]$, where $\epsilon = 0.02j$, $j = 1, \ldots, 10$, is referred to as the compressibility index. This panel clearly demonstrates the decisive superiority of the CS-UKF over the CSKF (in terms of computational efficiency), and over the BCS both in terms of computational efficiency and accuracy, for an increasing compressibility index starting at $\epsilon \approx 0.1$.



Fig. 1. Mean estimation errors and standard deviations of the BCS (red lines), CSKF (left panel only), and the CS-UKF (right panel only).

5. CONCLUDING REMARKS

A novel compressed sensing method is presented for the recovery of sparse/compressible, possibly time-varying, signals from a sequence of noisy observations. The newly derived scheme is based on the unscented Kalman filter (UKF) employing the socalled pseudo measurement technique for imposing the underlying sparseness constraint. By virtue of its UKF mechanism, the resulting CS-UKF method, is adequate for sequential processing of measurements and is capable of accurately capture the recovery error variances. The computational complexity of the CS-UKF is nearly equal to that of a standard UKF.

6. REFERENCES

- E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, pp. 489–509, 2006.
- [2] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Transactions on Signal Processing*, vol. 56, pp. 2346 – 2356, June 2008.
- [3] N. Vaswani, "Kalman filtered compressed sensing," Proceedings of the IEEE International Conference on Image Processing (ICIP), 2008.



Fig. 2. Normalized RMSE of the BCS, CSKF, and the CS-UKF for various compressibility indices.

- [4] A. Carmi, P. Gurfil, and D. Kanevsky, "A simple method for sparse signal recovery from noisy observations using kalman filtering," Tech. Rep. RC24709, Human Language Technologies, IBM, 2008.
- [5] A. Carmi, P. Gurfil, and D. Kanevsky, "Methods for sparse signal recovery using Kalman filtering with embedded pseudo-measurement norms and quasi-norms," *IEEE Transactions on Signal Processing*, vol. 58, no. 4, pp. 2405–2409, 2010.
- [6] D. Angelosante, J. A. Bazerque, and G. B. Giannakis, "Online adaptive estimation of sparse signals: Where RLS meets the l₁-norm," *IEEE Transactions on Signal Processing*, vol. 58, pp. 3436 – 3447, July 2010.
- [7] M. S. Asif and J. Romberg, "Dynamic updating for sparse time varying signals," Conference on Information Sciences and Systems, 2009.
- [8] D. Angelosante, G.B. Giannakis, and E. Grossi, "Compressed sensing of time-varying signals," 16th International Conference on Digital Signal Processing, 2009.
- [9] M. S. Asif, A. Charles, J. Romberg, and C. Rozell, "Estimation and dynamic updating of time-varying signals with sparse variations," International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2011.
- [10] A. Charles, M. S. Asif, J. Romberg, and C. Rozell, "Sparsity penalties in dynamical system estimation," Conference on Information Sciences and Systems, 2011.
- [11] N. Kalouptsidis, G. Mileounis, B. Babadi, and V. Tarokh, "Adaptive algorithms for sparse system identification," *Signal Processing*, vol. 91, pp. 1910 – 1919, August 2011.
- [12] S. J. Julier and J. K. Uhlmann, "A new extension of the kalman filter to nonlinear systems," Proceedings of the International Symposium on Aerospace/Defense Sensing, Simulation and Controls, 1997.
- [13] G. M. James, P. Radchenko, and J. Lv, "DASSO: connections between the Dantzig Selector and LASSO," *Journal of the Royal Statistical Society*, vol. 71, pp. 127–142, 2009.