Bayesian Compressive Sensing in Radar Systems

Radmila Pribić

Sensors Advanced Developments Thales Nederland Delft, The Netherlands radmila.pribic@nl.thalesgroup.com

Abstract— Compressive Sensing (CS) is presented in a Bayesian framework for realistic radar cases whose likelihood or priors are usually non-Gaussian. Its sparse-signal processing is modelbased and detection-driven, and also done numerically using Monte-Carlo methods. This approach aims for the stochastic description of sparse solutions, and the flexibility to use any prior information on signals or on data acquisition, as well as any distribution of noise or clutter, without the need for a closed analytic form of the Bayesian solution. This flexible Bayesian CS is shown by comparison with its closed-form predecessors.

Keywords: Bayesian compressive sensing; radar; Monte-Carlo methods; non-Gaussian distribution; numerical sparse recovery

I. INTRODUCTION

Compressive Sensing (CS) is refining a radar system utterly in the mathematics and engineering (e.g. [7], [9], [12] and [15]). It is optimized to information in received data rather than only to the sensing bandwidth (what is also being emphasized in the information geometry, e.g. [5]). Sparse-signal recovery (SSR) and analogue-to-information conversion (AIC) are major issues in CS, which are still being extensively investigated (e.g. [21] and [22]). Crucial mathematical and stochastic refinements that CS shall bring in radar are still to be revealed, especially regarding their practical aspects. Pursuing that practical CS in radar, Bayesian CS is preferred in radar because it is stochastic when treating noise, prior knowledge on signals or their data acquisition, and when providing estimation results ([19]). In a way, Bayesian CS is giving a fresh boost to exploring the prior knowledge in radar that started in the 50-ties ([23]).

In existing Bayesian CS, the Gaussian distribution is often assumed because of the simplicity and the feasibility of a closed-form solution. In most real-world cases it is not realistic as e.g. in high-resolution radar that is especially relevant with CS. High-resolution echoes such as e.g. sea clutter can better be described with a long-tail distribution (e.g. [10]). Moreover in CS, even originally Gaussian measurements can easily become non-Gaussian because of the random acquisition. Such non-Gaussian essentials lead us to a Bayesian approach based on Monte-Carlo (MC) methods, as already matured in particle filtering (PF, e.g. [13] and [15]). The MC approach in CS has also been studied in theory, e.g. for a universal goal in [3], for sparsity in [4] and [8], and for a combinatorial recovery in [16].

In this study, we introduce a practical MC-based Bayesian CS in radar. The numeric approach shall accommodate any prior knowledge on signals or their data acquisition, as well as any distribution of noise or clutter. Finally, the stochastic behavior of the estimates shall be available so that the radar back-end can be completed. The proposed Bayesian framework Ioannis Kyriakides Department of Electrical Engineering University of Nicosia, Cyprus kyriakides.i@unic.ac.cy

is presented after typical radar processing is recalled first at the system level in II. In III, their performance is compared on simulated data in a range-only case (kept basic for clarity). In IV, conclusions are drawn, and future work is indicated.

II. CS IN RADAR PROCESSING

Typical radar processing starts with matched filtering (MF) as in Figure 1. MF maximizes the signal-to-noise ratio (SNR) of a single target in radar echoes. When measurements y are described in a model: y = s + z, by nonrandom signals s in complex Gaussian noise z with zero mean and equal variances γ , $p(z|\gamma) \propto \exp(-|z|^2/\gamma)$, MF correlates y with a model of s.



Figure 1. Typical array-radar receiver without CS

If s is described as in CS, by a linear model: s = Ax, A is a sensing matrix and x is a target profile. Thus, A contains a radar-signal model with a desired profile x. Such a model is known from the physics, and initially nonrandom. E.g., in range estimation, columns of A are time-delayed replicas of the transmitted signal (usually a linear chirp because of the optimal gain, e.g. [6]). A remains an echo-model with no compression in data acquisition from Figure 1. The MF profile $x_{\rm MF}$, $x_{\rm MF}$ = $A^{H}y$ (H denoting hermitian) is computed using FFT and also decoupled for angle(s), range and doppler. x_{MF} is the statistics in detection and estimation, based on the likelihood-ratio test (LRT) and the maximum likelihood (ML), respectively. Thus, this stochastic processing is formally based on the likelihood $p(\mathbf{y}|\mathbf{x})$ of the data y whose stochastic behavior comes from the Gaussian noise z (describing the receiver thermal noise). This basic model is further extended with many empirical stochastic techniques such as e.g. CFAR for detection in clutter.

Finally, MF followed by detection, extraction/estimation and tracking, together with proper waveform design, large antenna arrays and a long observation time, can make very small targets at very long ranges be visible to radar. In such cases, the input SNR (at each antenna element) can be very low.



Figure 2. Array-radar receiver with CS

CS will effect both front-end and back-end as indicated in Figure 2. The origin of CS is the assumed sparsity of a profile xbecause there are only some targets expected in x. Accordingly, its recovery can be done with less measurements in y as long as the sensing matrix A remains incoherent enough what means that the maximum inner product between different columns of A is low (e.g. [7]). The physical nature of A suits the incoherence well (e.g. [24]). In the front-end, there may be random or deterministic AIC. This information conversion is also legitimate in the digital domain although more tempting if done sooner. The conversion works as a fat compression matrix **B** that shrinks y (both A and z) into y = By = BAx + Bz = Ax + zwhere an Arial letter indicates the compressed version. The ultimate compression would apply before the reception, so that it cannot degrade SNR by effecting z. Input SNR is very low (even below 0dB, what is often ignored in CS). For the optimal processing gain, measurements y gathered over the whole observation time and antenna elements, support the sparsesignal processing of a profile x over all its parameters: range, doppler and angle(s). The matrix A as well as the SSR size becomes huge but well-arranged because of the straight signal models what makes SSR even achievable in real-time ([19]).

A. Bayesian Approach in CS radar

In the linear model (with no random compression), y = Ax + z, a Bayesian solution for x relies on the posterior p(x|y), i.e. the probability of unknown x given data y, p(x|y)=p(y|x)p(x)/p(y), built from a prior p(x) available on the unknown x, together with the likelihood p(y|x) and the evidence p(y). A Bayesian solution is usually hard to compute in a closed form because it involves many complicated integrals. In Bayesian CS, the sparsity of a target profile x delivers a prior p(x). When the sparsity of an Nx1 vector x is formalized by a multivariate Laplace prior $p(x|\lambda)$, $p(x|\lambda) \propto \exp(-\lambda|x|_1)$, the maximum a posteriori (MAP) estimator of x, written as:

$$\boldsymbol{x}_{\text{MAP}} = \min_{\boldsymbol{x}} \{ |\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}|^2 + h|\boldsymbol{x}|_1 \}, \qquad (1)$$

is typical SSR (e.g. [1] and [21]) with the Manhattan norm $|\mathbf{x}|_1$ for promoting the sparsity and the Euclidean norm $|\mathbf{y}-\mathbf{A}\mathbf{x}|$ for minimizing the Gaussian noise, together with a threshold *h* that balances between the two tasks. Although a Laplace prior creates the typical SSR in (1), realizations from a Laplace distribution are hardly compressible, i.e. their values decrease purely and moreover, even worse with more realizations ([2]). In general, a MAP estimator can also be interpreted as the minimum mean square error (MMSE) estimator (being the mean with a posterior) with a possibly different prior. With a Gaussian prior, the two Bayesian estimators are equivalent.

In CS, an underdetermined linear system: y = Ax + z, can be solved by SSR in (1), i.e. *M* input observations in *y* could be enough for *N* outputs in *x*, M < N, because of the sparsity, i.e. only *K* non-zeros in *x*, K < M, and incoherence of *A* (e.g. [7]).



Figure 3. Principle of a Bayesian-CS algorithm (e.g. FL from [1]) shown on a test case by initial iterations (starting with MF) and the final result. The test case is generated as given in III, with 20dB output SNR and -10dB *h* from (1).

Bayesian SSR iteratively assesses x from p(x|y) based on a multi-layer hierarchical prior $p(\mathbf{x})$. Such a fast SSR algorithm ([1]) was adapted to radar signals, and called complex Fast Laplace (cFL, [19]). In cFL, the prior $p(\mathbf{x}|\lambda)$ is built from a complex Gaussian prior for x and a Γ hyper-prior for the variance of x. cFL starts actually with x_{MF} , i.e. all contributions $A^{H}y$ in the grid, and refines x_{MF} in a number of iterations by selecting significant elements as long as the threshold h and convergence criteria hold. This selection is based on increase in the assessed posterior in each element, as shown in Figure 3. In this stochastic manner, not only the non-zeros but also their estimation errors are provided. cFL is also fast because of a greedy implementation based on optimization separable for each n, n = 1, N. Other algorithms can also be fast but they are not stochastic (e.g. [22]). Moreover, in cFL the output SNR and the processing gain remain the MF equivalents.



Figure 4. Gaussian likelihood (black) and a resulting posterior (blue) with a norm- l_1 sparsity prior (solid green lines) of a non-zero (solid) and a zero (dotted) at different noise variances γ : a) 0.1 and b) 0.01. Comparable P_{fa} implies different sparsity parameters λ as well as thresholds *h* from (1).

There is also another class of Bayesian approaches in CS that involves priors on random sensing matrices, such as e.g. belief propagation and fast iterative message passing (see e.g. [21]). This Bayesian work is also heavily based on the Gaussian assumptions as needed for the closed-form solutions.

B. Bayesian SSR Using Monte-Carlo Methods

MAP or MMSE estimates can be numerically approximated by using a large number of Monte-Carlo (MC) realizations from an estimated posterior (e.g. [20]). Advances in MC techniques together with increasing computational power encouraged the development of feasible MC solutions, as e.g. with sequential MC methods, also known as particle filtering (PF) (e.g. [13] and [15]). The MC approach is also investigated in CS theory (e.g. [3], [4], [8], and [16]). Here the approach is more practical as aimed for the flexibility needed in a realistic radar system.

When translating the MAP estimation from (1) into an MC version, the goal remains the same: to identify significant non-zeros $\{x_{n(k)}\}$ in x satisfying (1) where n(k) indicates a column $a_{n(k)}$ of A, and $\{n(k)\}$ is a resulting support set K, k = 1, K. A non-zero can be sought randomly, but it converges faster when the choice is carefully tailored, as in cFL, in Figure 3.

In this MC work for radar, sparsity is related to detection at a probability of false alarms P_{fa} . Optimal detection strategy: via likelihood or posterior is also studied (e.g. [18]), as drafted for one x_n in FL in Figure 4. The value h in (1) equals $\gamma\lambda$ with variance γ from the likelihood and sparsity λ in the prior, but also e.g. $\sqrt{\gamma \ln(1/P_{fa})}$ in the generalized LRT (GLRT, [11]). Thus, both γ and λ can be known. The optimization in (1) can also be greedy when split to individual posteriors (e.g. [20]).

Accordingly, an approximation $p(n | \mathbf{y}, \gamma, \lambda)$ of the posterior $p(x_n | \mathbf{y}, \mathbf{x}_{-n})$ (assuming the rest \mathbf{x}_{-n} is known or zero) is studied to build an importance density of an individual atom *n*, as:

$$p(n \mid \mathbf{y}, \gamma, \lambda) \propto p(n \mid \mathbf{y}, \gamma) p(n \mid \lambda)$$
(2)

where $p(n|\mathbf{y},\gamma)$ encourages selecting a good candidate *n* based on a detection test. An individual prior $p(n|\lambda)$ can be any prior but it serves the sparsity of \mathbf{x} here via the detection threshold *h*. In a cFL case, such a detection test can be built with the MF statistics from the GLRT: $p(\mathbf{y}|x_n, \mathbf{x}_{-n})/p(\mathbf{y}|0, \mathbf{x}_{-n})$ at \mathbf{x}_{MF} , n = 1, N.

An individual *n* with higher probability of being a good candidate will be drawn more often from $p(n | y, \gamma, \lambda)$. A large number *L* of MC realizations (or particles) is used to draw an element n^l with the weight w_n^l , $w_n^l \propto p(n^l / y, \gamma, \lambda)$, l = 1, *L*. In each iteration *k*, a single non-zero n(k) is sought, as in cFL, from the greedy residuals $y_{\text{res, k}}$, initially *y* or the greedy remains $\mathbf{x}_{\text{res, k}}$, initially \mathbf{x}_{MF} . i.e. $\mathbf{x}_{\text{res, l}} = \mathbf{A}^H \mathbf{y}$. The best candidate n(k) with the highest weight assessed by $\{w_n^l\}$ is selected. An estimate $x_{\text{BCS}, n(k)}$ of its amplitude $x_{n(k)}$ is computed from $\mathbf{x}_{\text{res, k}}$. The MC generation repeats from $p(n / \mathbf{y}_{\text{res, k+1}}, \gamma, \lambda)$ updated with the *k*th non-zero model-based contribution: $\mathbf{x}_{\text{res, k+1}} = \mathbf{x}_{\text{res, k}} - \mathbf{x}_{\text{BCS}, n(k)}$ is the *k*th estimate. This selection of significant elements continues as long as the detection threshold *h* and convergence criteria hold.

In a realistic case, the likelihood p(y|x) as well as the prior p(x), can be any distribution, not restricted as in (1) to the Gaussian likelihood or the (Laplace) sparsity prior only. Accordingly, the proposed MC approach shall need no such restrictions. The distribution p(y|x) can represent any likelihood including even empirical distributions being learned from data y. The distribution p(x) can represent any prior information on x, including that of sparsity but moreover, not the sparsity necessarily given only by the Laplace distribution.

C. Future Upgrades

This MC-based Bayesian framework shall employ not only sparsity priors but also other prior knowledge available about radar signals and their acquisition. Any radar noise or clutter and optimal detection strategies are also to be supported.

The freedom in estimation grid (and observation grid) is to be employed as good as possible in CS radar (e.g. [14]). The estimation grid is studied from the Bayesian perspective ([17]), and also within the scope of information geometry (e.g. [5]).

Tracking is also to be embedded in this Bayesian framework. Sparsity and compressive acquisition are expected to be natural and beneficial in PF. Moreover, priors from the sequential estimation can improve Bayesian CS and enable its adaptive acquisition (e.g. [15]). Finally, nonlinear dynamics and non-Gaussian distribution are supported as needed for processing of realistic radar measurements.

III. COMPARISON RESULTS

Existing cFL ([19]) and its MC translation (IIB) are compared in the same test cases with simulated measurements. To keep the tests simple but explanatory enough, the measurements are range only in pulse radar with parameters given in TABLE I. In the basic radar application of the range profile, a linear model: y = Ax + z, contains measurements y over one single pulse repetition time (PRT), an unknown range profile x, receiver noise z, and a known sensing matrix A containing delayed replicas of a transmitted waveform (being a linear chirp of bandwidth close to the reference sampling frequency f_{s} , $f_s = 1$). In order to have an underdetermined linear system without compressive acquisition, M input observations remain while the estimation grid is up-sampled to N outputs, M < N.

 TABLE I.
 MODEL PARAMETERS FOR RANGE ONLY

Parameters	Notation	Value
Number of samples in a pulse (linear chirp)	$N_{\rm p}$	25
Number of input samples (in observation grid)	М	108
Number of range cells in reference grid	$M_{ m R}$	83
Range upsample factor	$F_{\rm R}$	3
Number of range cells (in estimation grid): F_RM_R+1	Ν	250

The target locations are uniformly randomly chosen over all N possible range cells. The true value of the amplitude of every target in x is set to 1. The target SNR is given as an output SNR.



Figure 5. MSE of range profile x estimated by cFL ([19]) and its MC version (IIB) for: a) all elements and b) zeros only, at different SNR from 100 runs.



Figure 6. Profile x estimated by cFL, its MC version and MF in one run.

The normalized mean squared error (MSE) in the estimated \mathbf{x}_{BCS} , MSE(\mathbf{x}_{BCS}) = $|\mathbf{x}_{BCS} \cdot \mathbf{x}|^2 / |\mathbf{x}| |\mathbf{x}_{BCS}|$, is computed for non-zero (targets) and zero elements in the true \mathbf{x} , with the two cFL versions at different output SNR from 100 noise runs, all with 10 targets, and shown in Figure 5. The number L of the MC realizations in the Monte-Carlo version (IIB) is 100.

At lower SNRs the MSE performance is even improved for non-zeros in x, and slightly degraded for zeros as clear from Figure 5. At higher SNRs the MSE performance is comparable. For more clarity in SSR, a single run is also shown with both BCS estimates, together with MF and its detection threshold in the Gaussian case in Figure 6. The model-based BCS produces points only, and not the whole response above the threshold.

IV. CONCLUSIONS

Bayesian CS is improved by using Monte-Carlo methods what encourages work on the CS flexibility needed for non-Gaussian cases in a realistic radar system. This MC algorithm for Bayesian CS is based on the importance density that promotes the sparsity via a detection test. The performance of the MC method is comparable with the original Bayesian CS. In future work, other priors on radar signals and their acquisition, other distributions of noise or clutter, other detection strategies, grid matching and tracking will be embedded in this framework.

REFERENCES

- S.D Babacan, R. Molina, and A. K. Katsaggelos, "Bayesian Compressive Sensing Using Laplace Priors," IEEE Transactions on Image Processing, vol.19/no.1, pp.53-63, 2010.
- [2] R.G. Baraniuk, V. Cevher, and M.B. Wakin, "Low-Dimensional Models for Dimensionality Reduction and Signal Recovery: A Geometric Perspective", Proceedings of the IEEE, vol.98/no.6, pp.959-971, 2010.
- [3] D. Baron and M. F. Duarte, "Universal MAP estimation in compressed sensing," *Allerton Conf. Comm., Control, Computing*, Sep. 2011.
- [4] T. Blumensath and M.E. Davies, "Monte Carlo Methods for Adaptive Sparse Approximations of Time-Series", IEEE Transactions on Signal Processing, vol. 55/no. 9, pp. 4474–4486, September 2007.
- [5] Y.Cheng, X. Wang, T. Caell, X. Li and B. Moran, "On Information Resolution of Radar Systems", IEEE Transactions on AES, vol. 48/no. 4, pp. 1289–1306, October 2012.
- [6] C. E. Cook and M. Bernfeld. Radar signals; an introduction to theory and application. Academic Press, 1967.
- [7] D. Donoho, "Compressed sensing," IEEE Transactions on Information Theory, vol. 52/no. 4, pp. 1289–1306, 2005.
- [8] M. Elad and I. Yavneh, A Plurality of Sparse Representations is Better than the Sparsest One Alone, IEEE Transactions on Information Theory, vol. 55/no. 10, pp. 4701-4714, October 2009.
- [9] J.H.G. Ender, "On Compressive Sensing Applied to Radar," Elsevier Signal Processing 90, pp. 1402-1414, 2010.
- [10] M. Farshchian, and F. Posner (2010). The Pareto distribution for low grazing angle and high resolution X-band sea clutter. US NRL.
- [11] J.J. Fuchs, "The Generalized Likelihood Ratio Test and the Sparse Representations Approach", vol. 6134, pp. 245–253, of ICISP in Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2010.
- [12] M. Herman, and T. Strohmer, "High-resolution Radar via Compressive Sensing," IEEE Transactions on Signal Processing, vol.57/no.6, pp.2275-2284, June 2009.
- [13] Y. Huang, and P. Djurić, "Bayesian Detection of Transient Signals in Colored Noise", in ICASSP 2000, pp.725-728, 2000.
- [14] R. Jagannath, G. Leus, R. Pribić, "Grid Matching for Sparse Signal Recovery in Compressive Sensing," *EuRad 2012*, pp.111-114, 2012.
- [15] I. Kyriakides, "A configurable compressive acquisition matrix for radar tracking using particle filtering," *EuRad* Amsterdam, pp. 103-106, 2012.
- [16] I. Kyriakides and R. Pribić, "Bayesian Compressive Sensing using Monte-Carlo Methods" to appear at EUSIPCO 2013.
- [17] I. Kyriakides, R. Pribić, H. Sur, and N. At, "Grid Matching in Monte-Carlo Bayesian Compressive Sensing" to appear at FUSION 2013.
- [18] S. Paris, D. Mary and A. Ferrari, "Composite Hypothesis Tests for Sparse Parameters", IEEE Workshop SSP, pp.737-740, 2011.
- [19] R. Pribić and H. J. Flisijn, "Back to Bayes-ics in Radar: Advantages for Sparse-Signal Recovery," *CoSeRa 2012*, Bonn, May 2012.
- [20] B. Shen, M.F. Bugallo, and P. Djurić, "Estimation of Multimodal Posterior Distribution Of Chirp Parameters with Population Monte-Carlo Sampling", in ICASSP 2012, pp.3861-3864, 2012..
- [21] J. A. Trop, and S.J. Wright, "Computational Methods for Sparse Solution of Linear Inverse Problems," Proceedings of the IEEE, vol.98/no.6, pp.948-958, June 2010.
- [22] B. Wahlberg, S. Boyd, M. Annergren, and Y. Wang, "An ADMM Algorithm for a Class of Total Variation Regularized Estimation Problems", IFAC 2012.
- [23] P. M. Woodward, Probability and information theory, with applications to radar. Pergamon, 1953.
- [24] L. Zegov, R. Pribić and G.Leus, "Optimal Waveforms for Compressive Sensing Radar", to appear at EUSIPCO 2013.