Compressive Sensing and Fast Simulations

Applications to Radar Detection

Compressive Sensing and Fast Simulations

Applications to Radar Detection

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof.ir. K.C.A.M. Luyben,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen
op maandag 5 november 2012 om 12:30 uur

door

Laura ANITORI

Laurea di Dottore in Ingegneria delle Telecomunicazioni
Università degli Studi di Pisa
geboren te Ronciglione, Italië

Dit proefschrift is goedgekeurd door de promotor:

Prof.ir. P. Hoogeboom

Samenstelling promotiecomissie:

Rector Magnificus,	voorzitter
Prof.ir. P. Hoogeboom,	Technische Universiteit Delft, promotor
Prof.dr.-Ing J. Ender,	University of Siegen/Fraunhofer FHR
Prof.dr. F. Le Chevalier,	Technische Universiteit Delft/Thales
Prof.dr. R. Srinivasan,	Advanced Institute of Mathematics, Statistics and Computer Science/ Cassidian-IW Bangalore
Prof.dr.ir. A. J. van der Veen,	Technische Universiteit Delft
Ir. M. P. G. Otten	TNO

This research was supported by the European Office of Aerospace Research and Development, London, UK, in collaboration with the Air Force Office of Scientific Research under Award FA8655-04-1-3025, and by TNO research programs ‘VP Space Data Utilization’ and ‘V922 Countering Improvised Explosive Devices and Landmines’.

ISBN 978-94-6191-427-9
Compressive Sensing and Fast Simulations: Applications to Radar Detection.
Dissertation at Delft University of Technology.

Ai miei genitori
e Jurriaan

Contents

Chapter 1

Introduction

1.1 The need for new sampling techniques

For most modern high-resolution multi-channel radar systems one of the major problems to deal with is the huge amount of data to be acquired, processed and/or stored. But why do we need all these data? According to the well known Nyquist-Shannon sampling theorem [1–3], natural signals have to be sampled at at least twice the signal bandwidth to prevent ambiguities. Therefore, sampling of very wide bandwidths may require Analog to Digital Converter (ADC) hardware that is unavailable or very expensive. Especially in multi-channel systems, both the cost and power consumption can become critical factors. A similar limit applies also in the spatial domain; for instance, for wide scanning arrays the spatial sampling must be smaller than $\lambda/2$, with λ being the signal wavelength. In some applications, such as interleaving of radar modes in time (pulses) or space (antenna aperture), multi-function operation can lead to conflicting requirements on sampling rates in both the time and spatial domain. So, while on one hand the increased number of degrees of freedom improves the system performance, on the other hand it puts a significant burden on both the off-line analysis and performance evaluation of sophisticated detectors, and on the real-time acquisition and processing. For example, in Space Time Adaptive Processing (STAP) algorithms, evaluating the optimal filter weights is an immense computational load both when simulating such detectors in the design phase and in real time implementation.

In some cases, measurement time may also be a constraint, especially if the application must work in (near) real-time. This is the case for example in 3D radar imaging for airport security inspection of passengers. Acquisition of full 3D high resolution image data may require a measurement time that is unacceptable in this situation.

Since the issue of processing and storing huge amounts of data is a problem shared by most modern digital systems, this subject has attracted much attention in the last decade. Particularly, it was observed that in many systems while the data are sampled at very high rates, often a compression stage follows the acquisition to reduce the data dimensionality. This is based on the fact that most data sets contain significantly less information than the number of measured samples. For example, in search radars the observation space is divided in thousands of range, Doppler and azimuth bins, to search for the presence of targets. However, while the volume contains many voxels, the number of targets that might be present in the scene at a given time is orders of magnitude less than the numbers of voxels in the search cone. Based on similar observations, in mid 2006, Candès, Romberg and Tao [4] proposed a new, nonlinear sampling theorem, stating that *sparse* signals can be recovered from measurements acquired at a rate much lower than the one dictated by the Nyquist-Shannon theorem. This innovative theorem served as the backbone for the development of a new field called Compressive Sensing or Compressed Sensing or Compressive Sampling (CS). The fundamental goal of CS is to perform simultaneously data acquisition and compression, thus reducing significantly both the sampling requirements and amount of data.

Although CS can efficiently deal with issues related to real-time data acquisition, the design and analysis of sophisticated, multi-dimensional radar detectors is also an issue that the radar designer has to deal with. In fact, when designing multi-dimensional, efficient detection schemes, whose performances are difficult or impossible to characterize analytically, we need to resort to simulations to evaluate the detector performance. Even if the computational capacity of modern computers has drastically increased in the last decades, generating and processing millions of (multi-dimensional) random variables is still a time consuming task. This is the case for example in STAP detectors. Such detectors are widely used in multi-channel systems for improving detection capability of targets which are embedded in clutter, noise and jammers. In typical STAP detectors, an angle-Doppler map of the background disturbance surrounding the cell under test (CUT) is estimated from a set of target-free training data and used to compute the optimal two-dimensional beamformer to suppress disturbance and interference and enhance the signal of interest. STAP detectors are notoriously intensive from a computational point of view, the main processing burden being the inversion of the sample covariance matrix. Furthermore, the more advanced (and robust) STAP detectors are also analytically difficult to examine, [5]. In such cases, the detector design must be performed by simulations. A critical aspect of design is the false alarm probability, which should be very low. In conducting rare-event simulations of systems that involve signal processing operations that are mathematically complex, there are two principal issues that contribute to simulation time. These have to be dealt with effectively. The first issue concerns the rare event itself whose probability is being sought. The second concerns the computational intensity that accompanies the signal processing. It is a daunting task to conduct conventional Monte Carlo (MC) simulations that involve several millions of trials to estimate low false alarm probabilities (FAP), with as many matrix inversions.

To alleviate the simulation burden of classical MC methods, forced Monte Carlo or Importance Sampling (IS) simulations can be implemented. By efficiently sampling the underlying probability density function (pdf), IS provides a very fast alternative to more conventional MC methods. If one is able to design powerful biasing schemes, the simulation time required to estimate even very low FAPs might be several orders of magnitude smaller than the one of a MC estimator. Furthermore, in some cases inverse IS can be used to design the detection thresholds of complex detectors whose performance cannot be analytically evaluated [6].

1.2 Background and Research Objectives

In this thesis we investigate sampling methods that can deal with the problems of processing complexity as well as efficient performance evaluation by reducing the required amount of samples. By cleverly using properties of the signals or random variables involved, the considered techniques, namely CS and IS, both alleviate the burden related to data handling in complex radar systems. These methods, although very different in nature, provide an alternative to classical sampling and performance analysis techniques. In the next sections, we will briefly review the theories of CS and IS, which will help to put our research objectives and results into perspective.

1.2.1 Compressive Sensing

The first method investigated, CS, is a revolutionary acquisition and processing theory that enables reconstruction of sparse signals from a set of non adaptive measurements sampled at a much lower rate than required by the Nyquist-Shannon theorem. This results in both shorter acquisition times and reduced amounts of data. At the core of the CS theory is the notion of signals sparseness. In fact, the information contained in many natural signals can be represented more concisely if the signals are looked at in a proper transform domain. If this is the case, then the (complex) signal of interest $\mathbf{x} \in \mathbb{C}^N$ can be recovered from an undersampled set of linear projections $\mathbf{y} \in \mathbb{C}^n$, with $n \ll N$, via ℓ_1-norm minimization.[1] ℓ_1-norm based algorithms impose sparsity on the estimated signal by looking for the solution with minimum ℓ_1-norm which is in agreement with the measurements [4, 7–11].

---

[1] Throughout this thesis we consider the case of signals and measurements in the complex domain, since it is more appropriate for radar applications. Equivalent results can be obtained for the case of real signals.

The sensing model can be mathematically represented as

$$\mathbf{y} = \mathbf{A}\mathbf{x}, \tag{1.1}$$

where the $n \times N$ matrix $\mathbf{A}$ is called the *sensing matrix*.

The CS theory is based upon three fundamental concepts: sparsity, sensing matrix properties and ℓ_1-norm minimization. In the next paragraphs we briefly review these concepts.

Sparsity

Consider a discrete time (or space) signal having N samples. This signal can be represented as a complex N-length vector $\mathbf{x} \in \mathbb{C}^N$. The signal is said to be k-sparse if a basis[2] $\boldsymbol{\Psi} = [\psi_1|\psi_2|\cdots|\psi_N] \in \mathbb{C}^{N\times N}$ exists such that $\mathbf{x} = \boldsymbol{\Psi}\boldsymbol{\alpha}$ and only k coefficients in $\boldsymbol{\alpha}$ are non-zero. In this case we can write

$$\mathbf{x}_\Omega = \boldsymbol{\Psi}_\Omega \boldsymbol{\alpha}_\Omega, \quad \Omega = \{i, i = 1 \cdots, N : \alpha_i \neq 0\},$$

where $\boldsymbol{\Psi}_\Omega$ is the $N \times k$ matrix containing only the basis vectors $\{\psi_i\}_{i=1}^k$ corresponding to the k non-zero coefficients in $\boldsymbol{\alpha}$, and $\boldsymbol{\alpha}_\Omega$ is the k-length vector of non-zero coefficients. If instead $\boldsymbol{\alpha}$ has k large coefficients (with indices in Ω) and $(N-k)$ small coefficients, then the signal $\mathbf{x}$ is not exactly sparse but it is said to be compressible, and the best k-terms approximation can be determined [12].

Signal sparsity or compressibility plays a central role in transform coding, where the N-length signal $\mathbf{x}$ is approximated by preserving only the k largest transform coefficients α_i. This is for example the case in JPEG compression, where only the largest wavelet transform coefficients and their locations are stored [13]. Clearly, if a signal can be represented in a domain where it is sparse, then, in this domain, the signal information is contained in only k of the N coefficients.

In many radar applications the signals of interest are often sparse in range, Doppler and/or azimuth. Consider for example the range dimension of an air surveillance radar, as shown in Figure 1.1. The system scans several kilometers in range searching for targets. Most of the time, however, only a few, one or no targets will be present at any given time. In this case, the target response (or scene) can be considered as being sparse in the range (time) domain, and therefore the matrix $\boldsymbol{\Psi}$ is the canonical basis, i.e., $\boldsymbol{\Psi} = \mathbf{I_N}$, and $\mathbf{x} = \boldsymbol{\alpha}$.

---

[2] Throughout this thesis, we consider the sparsifying basis $\boldsymbol{\Psi}$ to be an orthonormal basis (thus preserving vectors distances). Examples of orthobasis are discrete Fourier transform and wavelet transform, which are widely used in image compression.

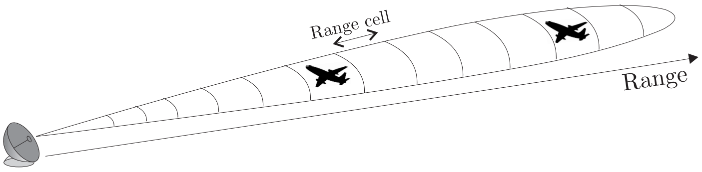


Figure 1.1: Example of air surveillance radar.

Example: *Range Pulse Compression*

Consider a pulsed radar transmitting a chirp signal of duration $T = 30\mu$s and bandwidth $B = 3$MHz (corresponding to 50m range resolution). There are five targets present at ranges of 10.4Km, 10.6Km, 23Km, 26Km, and 30.5Km, respectively. The received real signal, which is shown in Figure 1.2(a), is sampled at Nyquist rate ($f_s = 6$MHz). The target scene can be reconstructed using conventional pulse compression, without undersampling, as shown in Figure 1.2(b) in dashed line. Alternatively, because of the sparsity in the range domain, we can use the CS approach. In this example, the received signal is randomly subsampled in time by a factor of 5, and the true target positions are still correctly found, without ambiguous responses. ☐

The sparse approximation also applies to high resolution imaging radars, when the target reflectivity can be approximated as consisting of the sum of a few main (independent) scatterers distributed over the whole observed range-azimuth plane.

Sensing Matrix Properties

The sensing mechanism of acquiring n measurements via linear projection through the operator $\mathbf{A}$ is mathematically represented as in (1.1). Using the sparse representation, we can write

$$\mathbf{y} = \mathbf{A}\mathbf{x} = \mathbf{A}\boldsymbol{\Psi}\boldsymbol{\alpha} = \boldsymbol{\Phi}\boldsymbol{\alpha}, \tag{1.2}$$

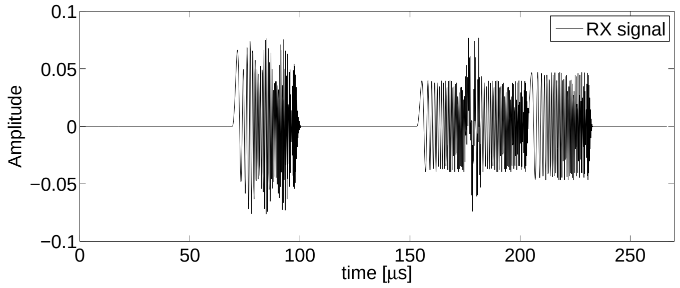
where $\boldsymbol{\Phi} = \mathbf{A}\boldsymbol{\Psi}$ is an $n \times N$ matrix.

However, the problem of recovering the length N vector $\mathbf{x}$ from the measurements $\mathbf{y}$ is ill posed. In fact, there exist infinite many solutions $\mathbf{x}'$ such that $\mathbf{A}\mathbf{x}' = \mathbf{y}$. Even so, if the sensing matrix is designed properly, it has been proved in [4, 14] that the signal $\mathbf{x}$ can be recovered from the measurements $\mathbf{y}$. Candès, Romberg and Tao introduced the concept of Restricted Isometry Property (RIP), which is defined as follows.[3]

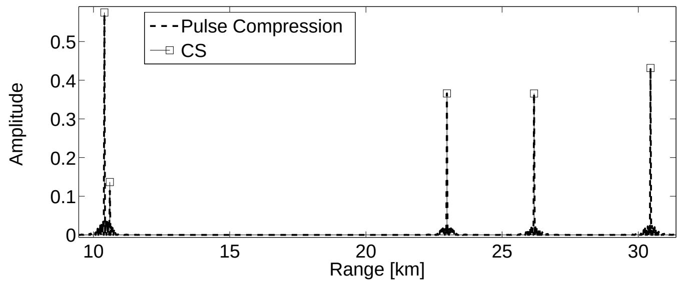
Definition 1.2.1. [4, 14] *A matrix* $\boldsymbol{\Phi}$ *satisfies the RIP of order* k *if there exists*

---

[3] Throughout this thesis we consider matrices $\boldsymbol{\Phi}$ with unit column norm.



(a) Received time signal



(b) Estimated target range response using pulse compression with all measurements (dashed line) and CS with 20% of the original time samples (squares)

Figure 1.2: Example of received signal from range sparse targets and reconstructed scene using pulse compression and CS.

$\delta_k \in (0, 1)$ *such that*

$$(1 - \delta_k) \leq \frac{\|\mathbf{\Phi z}\|_2^2}{\|\mathbf{z}\|_2^2} \leq (1 + \delta_k) \tag{1.3}$$

holds for any k*-sparse vector* $\mathbf{z}$.

In words, the RIP condition ensures that any subset k of columns of $\mathbf{\Phi}$ consists of vectors which are almost orthogonal. If the matrix satisfies the RIP of order $2k$, then it implies that two different k-sparse vectors in $\mathbb{R}^N$ will not map to the same vector in $\mathbb{R}^n$ and their distance is approximately preserved.

Although many algorithms in CS literature rely on the RIP concept to provide recovery guarantees, it is in general very difficult to verify whether or not a matrix satisfies (1.3). To this end, one should check that all possible combinations of any k columns of the sensing matrix satisfy the RIP, i.e., a combinatorial search is required. For the special class of matrices with Gaussian independent and identically distributed (i.i.d.) random entries, it has been demonstrated that the RIP condition is satisfied with high probability for $n > ck \log(N/k)$, with c being a small constant [4]. Since random matrices have nice statistical properties, much attention has focussed on designing CS systems using random entries sensing matrices. However, as we will see later in Chapter 2, for some specific applications, such as range compression, more efficient sampling strategies exist.

Since the RIP is practically impossible to compute for any realistic values of N and k, a metric that is often used for the design of good CS matrices is the *incoherence*, introduced in [15]. The coherence $\mu(\mathbf{B}, \mathbf{C})$ between any two matrices $\mathbf{B} = [b_1|b_2|\cdots|b_N]$ and $\mathbf{C} = [c_1|c_2|\cdots|c_N]$ is defined as [9]

$$\mu(\mathbf{B}, \mathbf{C}) = \sqrt{N} \max_{1\leq i,j\leq N} |b_i^T c_j|, \tag{1.4}$$

i.e., the coherence is the maximum correlation between any two columns of the matrices $\mathbf{B}$ and $\mathbf{C}$.

As we will see in the next paragraph, the concept of coherence plays an important role in determining how many CS measurements are needed to recover the signal of interest $\mathbf{x}$. Define the matrix $\widetilde{\mathbf{A}} \in \mathbb{C}^{N\times N}$ as the orthogonal matrix from which a random subset of rows is selected by an operator $\Theta \in \mathbb{R}^{n\times N}$ to obtain the sensing matrix $\mathbf{A}$ in (1.1). Then, the lower the coherence $\mu(\widetilde{\mathbf{A}}, \mathbf{\Psi})$ between the sensing and sparsity domains, the smaller the number of CS measurements required. An example of maximally incoherent matrices is the Discrete Fourier Transform (DFT) matrix and the canonical basis $\mathbf{I}$.

Recovery via ℓ_1-norm minimization

So far, we have seen that if the signal of interest $\mathbf{x}$ is sparse, then it is possible to recover it from a number of measurements $n \ll N$. The question is now how we can actually recover $\mathbf{x}$ (or an estimate of it) from $\mathbf{y}$, the problem being ill-posed.

A common method, which is used for overdetermined systems of linear equations, is the Least Square (LS) approach, which is based on minimizing the residual energy. For an underdetermined system of linear equations, as is the case with CS, a solution to (1.1) can be found using a similar approach by solving the problem:

$$\widehat{\boldsymbol{\alpha}} = \arg\min \|\boldsymbol{\alpha}'\|_2 \quad \text{subject to} \quad \mathbf{\Phi}\boldsymbol{\alpha}' = \mathbf{y}, \tag{1.5}$$

where the p-norm of a vector $\mathbf{z} \in \mathbb{C}^N$ is defined as $\|\mathbf{z}\|_p := \left(\sum_{i=1}^N |z_i|^p\right)^{1/p}$ for $1 \leq p < \infty$, $\|\mathbf{z}\|_0 := \sum_{i=0}^N \mathbf{1}(z_i \neq 0)$ for $p = 0$, with $\mathbf{1}$ being the indicator function, and $\|\mathbf{z}\|_\infty := \max(|z_1|, \cdots, |z_N|)$.

Using the Lagrange multiplier, it can be shown that the solution to (1.5), for $p = 2$, is given by $\widehat{\boldsymbol{\alpha}} = \mathbf{\Phi}^H(\mathbf{\Phi}\mathbf{\Phi}^H)^{-1}\mathbf{y}$, where $\mathbf{\Phi}^H(\mathbf{\Phi}\mathbf{\Phi}^H)^{-1}$ is the (right) pseudo-inverse of the matrix $\mathbf{\Phi}$. However, using ℓ_2-norm minimization does not provide a sparse solution.

To enforce the a priori knowledge about signal sparsity in the recovery algorithm, one should search for a solution with minimum ℓ_0-norm. Since the ℓ_0-norm counts the number of non-zero elements in a vector, minimizing it is equivalent to looking for the sparsest solution $\widehat{\boldsymbol{\alpha}}$ which is in agreement with the measurements $\mathbf{y}$.

Unfortunately, this problem not only does not have a closed form solution, but it also is NP-hard to solve (combinatorial complexity). However, if we replace the ℓ_0 with the ℓ_1-norm, then the problem is convex and can be solved using standard convex optimization routines. The use of ℓ_1-norm to enforce sparsity dates back to the '80s, when it was first used in geophysical applications to recover sparse signals from missing data [16–18].

In the ideal case of noise free data, using the measurements model in (1.2), the optimization problem can be written as

$$\widehat{\boldsymbol{\alpha}} = \arg\min \|\boldsymbol{\alpha}'\|_1 \quad \text{s.t.} \quad \boldsymbol{\Phi}\boldsymbol{\alpha}' = \mathbf{y}, \tag{1.6}$$

where $\|\boldsymbol{\alpha}\|_1 = \sum_{i=1}^{N} |\alpha_i|$. Once we have access to an estimate $\widehat{\boldsymbol{\alpha}}$, then the signal of interest can be easily obtained as $\widehat{\mathbf{x}} = \boldsymbol{\Psi}\widehat{\boldsymbol{\alpha}}$.

Furthermore, it was demonstrated by Candès and Romberg in [19] that, if $\mathbf{x}$ is k-sparse in the domain $\boldsymbol{\Psi}$, and the matrix $\mathbf{A}$ is an operator that selects a random subset n of rows of $\boldsymbol{\Psi}$, then the solution of (1.6) is exact with very high probability if $n \geq C\mu^2(\widetilde{\mathbf{A}}, \boldsymbol{\Psi})k \log n$. It is now clear that, if the coherence between the measurement domain and the sparsity domain is low, then the number of measurements needed is just of the order $k \log N$ instead of N. This fundamental theorem essentially justifies the use of ℓ_1-norm minimization for recovering sparse signals from highly undersampled measurements.

So far, we have only considered ideal settings. In practice however, the measurements are always corrupted by errors and the model in (1.2) must be modified to[4]

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}, \tag{1.7}$$

where $\mathbf{n}$ might be a vector of white Gaussian noise, or represent the quantization error, or any other disturbance that is added to the measurements [20–22]. In this thesis, we focus on the case of measurements noise with i.i.d. Gaussian entries, i.e., $\mathbf{n} \sim \mathcal{CN}_n(0, \sigma^2\mathbf{I})$, where $\mathcal{CN}$ stands for complex Gaussian distribution.[5]

---

[4] For easiness of notation, in the remainder of this thesis we will assume that the signal of interest $\mathbf{x}$ is sparse in the canonical basis, i.e., $\boldsymbol{\Psi} = \mathbf{I_N}$ and $\mathbf{x} = \boldsymbol{\alpha}$. The results can be easily generalized to the case $\boldsymbol{\Psi} \neq \mathbf{I_N}$ and will not change the derivation of our results.

[5] It was shown in [23–25] that, if white Gaussian noise is added to the signal $\mathbf{x}$ before sampling with the matrix $\mathbf{A}$, i.e.,

$$\mathbf{y} = \mathbf{A}(\mathbf{x} + \mathbf{n}),$$

then this results in a noise folding, which essentially can be treated exactly as (1.7) except for a scaling by a factor N/n in the noise variance.

In this scenario, the following convex optimization problem recovers a close approximation of $\mathbf{x}$ [7, 10]:

$$\widehat{\mathbf{x}} = \operatorname*{argmin}_{\mathbf{x}'} \|\mathbf{x}'\|_1 \quad \text{s.t.} \quad \|\mathbf{A}\mathbf{x}' - \mathbf{y}\|_2^2 \le \varepsilon, \tag{1.8}$$

where ε is a threshold proportional to the noise variance. Alternatively one can solve the unconstrained problem, referred in literature as the Least Absolute Shrinkage and Selection Operator (LASSO) [26] or Basis Pursuit Denoising (BPDN) [27], given by

$$\widehat{\mathbf{x}} = \min_{\mathbf{x}'} \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}'\|_2^2 + \lambda \|\mathbf{x}'\|_1, \tag{1.9}$$

where λ is a regularization parameter that controls the trade off between the sparsity of the solution and the ℓ_2-norm of the residual. Note that the relation between λ and ε that makes the two problems equivalent is not known and data dependent. Since the cost function in (1.8) and (1.9) is convex, these problems can be solved using standard techniques, such as gradient based or interior point methods. A variety of algorithms to solve (1.8) and (1.9) can be found in [28–40] and references therein.

In [11], an important theorem was derived which upper bounds the reconstruction error.

Theorem 1.2.2. [9,11] *Define* $\mathbf{x}_S$ *as a vector containing the* S *largest (in amplitude) coefficients of the vector* $\mathbf{x}$ *and zeros elsewhere. If the sensing matrix* $\mathbf{A}$ *obeys the RIP with constant* $\delta_{2k} \le \sqrt{2} - 1$*, then the solution to* (1.8) *obeys*

$$\|\widehat{\mathbf{x}} - \mathbf{x}\|_2 \le \frac{C_0}{\sqrt{k}}\|\mathbf{x} - \mathbf{x}_S\|_1 + C_1\varepsilon.$$

In words, Theorem 1.2.2 states that the reconstruction error has an upper bound that equals the sum of the error caused by the sparse approximation and an error proportional to the measurement noise.

The class of greedy algorithms, which is not investigated here, is also widely used to recover sparse approximation. These algorithms include Matching Pursuit (MP) and Iterative Hard Thresholding (IHT) based approaches [41–52].

Research Objectives and Results

Since the development of the CS theory, many papers have appeared demonstrating successfully that CS represents a valid alternative to conventional sampling and processing schemes [53–77].

However, while most of the work so far has served to establish the usefulness of CS in many radar applications, several practical issues have not been addressed yet. For example, if we need to perform target detection, how can the detection probability be optimized against false alarm rate? How can the false alarm rate be controlled adaptively against unknown noise and clutter? What is the relation between the BPDN threshold (λ or ε) and the detection and false alarm probabilities? Furthermore, how can a CS-radar be designed, in particular, what amount of undersampling is acceptable, and at what cost, in terms of power?

One of the main issues to be addressed to answer our questions is how to find the 'optimal' value of the threshold ε or regularization parameter λ to be used in the LASSO. These parameters are related to both the underlying noise power and the number of non-zero coefficients k. Hence, in a practical scenario, where neither the variance of the disturbance, nor the number of targets are known a priori, tuning of these thresholds is a very difficult task. Furthermore, even if we know the disturbance variance and the sparsity level, tuning the parameters for a given performance is still an open problem, since most of the theoretical works in CS provide pessimistic sufficient conditions that cannot be used in practice [78].

In most operational radars, Constant False Alarm Rate (CFAR) processors are used for adaptive target detection. Several CFAR schemes have been designed to attain good performance in the presence of different types of clutter and target scenarios [79–89]. The modeling and prediction of False Alarm Probability (FAP) is essential for the design of CFAR schemes. This in turn requires some level of knowledge of the underlying noise (or clutter) distribution which is input to the detector. Designing CFAR schemes seems to be out of reach for CS radar systems, due to the so far unknown relations between FAP/noise statistics and the parameters involved in the ℓ_1-norm reconstruction. This is due to both the non linearity of the reconstruction algorithms and to the fact that the estimated signal has only a limited number of non-zero coefficients, making it difficult to estimate any statistics.

The objectives of this thesis concerning CS are to i) design adaptive detection schemes with CFAR properties for CS-based radars that can be implemented in operational systems and ii) quantify their performance in terms of compression factor, sparsity and Signal to Noise Ratio (SNR). To this end, we will address and solve a number of issues, including: characterizing the input/output relations of the nonlinear LASSO reconstruction; adaptively estimating the recovery threshold parameter that provides the optimal reconstruction SNR; analyzing the target signal and noise distribution after ℓ_1-norm recovery; establishing the dependence of the recovery SNR on the amount of compression and number of targets present in the signal and designing suitable adaptive detection architectures with CFAR property that can be used without any prior knowledge of the original signal or noise.

1.2.2 Importance Sampling

The second technique investigated in this thesis, IS, has roots in statistical physics and represents a fast and effective method for the design and analysis of detectors whose performance has to be evaluated by simulations. Using fast simulations, the number of random variables needed to estimate rare events can be significantly reduced, thus resulting in much shorter simulation times [6, 90–93].

Fast simulations rely on techniques such as biasing and the g-method estimator [6, 94, 95], and can be used to not only predict the performance of detection schemes but also to design detector thresholds, using inverse IS techniques. In the following sections, we briefly review some fundamental concepts of IS that will be used throughout the second part of this thesis, based on [93].

Fast simulations

Importance Sampling (IS) is an efficient technique for estimation of events which occur with very low probabilities, i.e., rare events. Consider for example the problem of estimating the probability p_t that a random variable (RV) X with pdf $f(x)$ is above the threshold t, i.e.,

$$p_t = P(X > t) = E\big(\mathbf{1}(X > t)\big) \tag{1.10}$$

where $\mathbf{1}$ is the indicator function. If t is large enough, then p_t represents a rare event. Using classical MC simulations, an estimate of the probability p_t is obtained by generating K i.i.d. RVs X_i, $i = 1, \cdots, K$ from the same distribution as the original RV X and counting the number of times k_t that the realizations are above the threshold, i.e.,

$$\widehat{p}_t = \frac{k_t}{K} = \frac{1}{K}\sum_{i=1}^{K} \mathbf{1}(X_i > t). \tag{1.11}$$

Here, $\widehat{p}_t$ is an unbiased estimate of p_t and its variance is given by

$$\operatorname{var} \widehat{p}_t = \frac{1}{K}(p_t - p_t^2). \tag{1.12}$$

Using the Central Limit Theorem (CLT) it can be shown that the number of trials K required to estimate p_t with an error no greater than 20% with 95% confidence is about $K \approx 100/p_t$, [91]. Clearly, in certain applications such as radar and communications, where one needs to estimate false alarm probabilities or Bit Error Rates of the order of 10^{-6}, using classical MC methods becomes prohibitive, since it requires generating millions of RVs to achieve a desired accuracy. This translates directly into extremely long simulation times.

As an alternative to the classical MC method, IS works by biasing the original probability distributions in ways that accelerate the occurrences of rare events, conducting simulations with these new distributions, and then compensating the obtained results for the changes made. The principal consequence of this procedure is that unbiased probability estimates with low variances are obtained quickly.

Define f_* as the biased density function, i.e., f_* is the pdf of some transformation of the original RV X. Then equation (1.10) can be rewritten as

$$p_t = E\big(\mathbf{1}(X > t)\big) = \int \mathbf{1}(x > t)\frac{f(x)}{f_*(x)} f_*(x)dx = E_*\big(\mathbf{1}(X > t)W(X)\big), \qquad (1.13)$$

where $W(\cdot) = f(\cdot)/f_*(\cdot)$ is the weighting function and E_* denotes expectation with respect to the density function f_*.

Using the last equality in (1.13), the IS estimator of p_t can be derived, and it is given by [93]

$$\widehat{p}_t = \frac{1}{K}\sum_{i=1}^{K} \mathbf{1}(X_i > t)W(X_i), \quad X_i \sim f_*. \qquad (1.14)$$

In equation (1.14), $X_i \sim f_*$ indicates that the RVs X_i are drawn from the distribution f_*. This estimator is also unbiased and its variance is equal to

$$\operatorname{var} \widehat{p}_t = \frac{1}{K}[E_*\big(\mathbf{1}^2(X > t)W^2(X)\big) - p_t^2] = \frac{1}{K}[E\big(\mathbf{1}(X > t)W(X)\big) - p_t^2]. \qquad (1.15)$$

The expectation on the right hand side is called the I-function, and can be used to characterize the performance of the designed IS estimator. While running simulations, an estimate of the I-function can be evaluated as

$$\widehat{I} = \frac{1}{K}\sum_{1}^{K} \mathbf{1}(X_i > t)\, W^2(X_i), \quad X_i \sim f_*. \qquad (1.16)$$

If an optimal biasing distribution can be determined, it has been proved that, for estimating the same rare event, the variance of an IS estimator is always smaller than or equal to the one of the MC estimator [93]. As can be seen from (1.15), a sufficient condition so that the variance of the IS estimator is smaller than the variance of the MC estimator is that $W(x) < 1$ for $x > t$.

Biasing

One of the main tasks in IS is determination of good simulation distributions f_* such that the variance of the IS estimator is smaller than the variance of the MC estimator. Several types of biasing techniques, such as scaling, translation and exponential twisting, exist. An example of biasing by scaling is shown in the following example. A more detailed description about the properties of the optimal biasing distribution and the different biasing strategies can be found in [92,93]. These simple yet effective schemes can be used to bias one or more of the random variables involved in the estimation of the rare event. The amount of biasing is then controlled by a scalar θ or a vector $\boldsymbol{\theta}$ of biasing parameters, which, in turn, for a given f_* must be optimized to achieve the minimum variance IS estimator.

Example: *biasing by scaling* [93]

Suppose that we are interested in estimating the probability of false alarm $\widehat{p}_t$ that a random variable X with Rayleigh distribution $f(x) = \frac{x}{\sigma^2} e^{\frac{-x^2}{2\sigma^2}}$ and parameter $\sigma = 1$ is above the threshold $t = 5.2565$. It is known in closed form that the probability $p_t = e^{-t^2/2}$. Therefore, setting $t = 5.2565$ leads to $p_t = 10^{-6}$. Using Monte Carlo simulations, it would require to generate at least 10 million realizations of the RV X for estimating the rare event. However, using the theory of IS, we can think of scaling the original RV X by an amount $\theta > 1$, so that the rare event will occur more frequently. The bias introduced by scaling the RV will be taken care of by applying the proper weighting function during estimation of the false alarm, as shown in (1.14). The distribution f_* of the scaled RV θX can be obtained from the original one f using

$$f_*(x) = \frac{1}{\theta} f\Big(\frac{x}{\theta}\Big),$$

and the weighting function is given by

$$W(x) = \frac{f(x)}{f_*(x)} = \theta \frac{f(x)}{f(x/\theta)} = \theta^2 e^{\frac{-x^2}{2\sigma^2}(1-\frac{1}{\theta^2})}.$$

An example of the original and biased distributions is shown in Figure 1.3. ☐

It is clear that by drawing the RVs from the distribution f_*, the occurrence of the event $X > t$ takes place more frequently. In other words, we are sampling now in the 'important' region, i.e., the region where the event of interest is taking place. Simple schemes such as scaling or translation can be used to bias one or more of the random variables involved in the estimation of the rare event. The effectiveness of the IS estimator depends both on the biasing method applied and on the RVs involved in the estimations. Note that the value of θ that minimizes the IS estimator variance has also to be optimized. The procedure to adaptively estimate the optimal biasing parameter for each desired value p_t is called *Adaptive IS*.

Adaptive IS

After determining the biasing method and corresponding distribution, one has to find the value of θ such that the variance of the IS estimator is minimized.[6] From

---

[6] For simplicity, we consider here the case of a scalar biasing parameter.

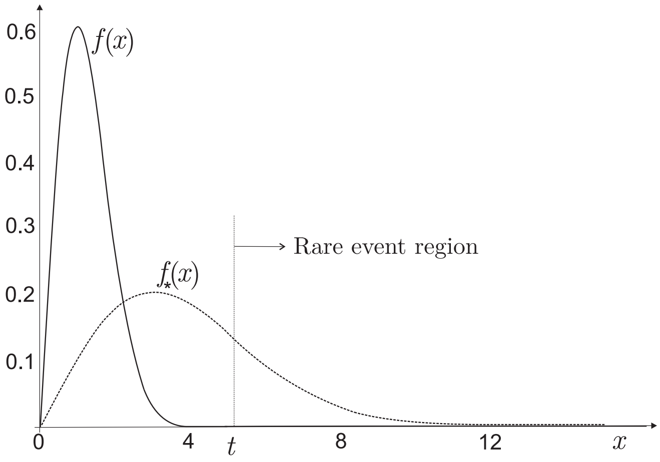


Figure 1.3: Example of biasing by scaling for Rayleigh distribution.

(1.15), it is clear that minimizing var($\widehat{p}_t$) is equivalent to minimizing the I-function (or an estimate of it), which depends, through the weighting function, on the biasing parameter θ. Therefore the optimum θ can be found by solving $\frac{\partial I(\theta)}{\partial \theta} = 0$. An estimate of the optimum biasing parameter can be adaptively computed, simultaneously with the false alarm probability $\widehat{p}_t$ and $\widehat{I}$, using, e.g., the Newton recursion

$$\theta_{m+1} = \theta_m - \delta \frac{\widehat{I'}(\theta_m)}{\widehat{I''}(\theta_m)}. \tag{1.17}$$

In (1.17) δ is a step-size parameter used to control convergence, m is the recursion index, $\widehat{I'}$ and $\widehat{I''}$ indicate the first and second partial derivatives of the estimated I-function with respect to θ, respectively. It is straightforward to show that the I-functions derivatives can also be estimated using the equations

$$\widehat{I'}(\theta_m) = \frac{1}{K} \sum_{1}^{K} \mathbf{1}(X_i > t)\, W(X_i, \theta_m) W'(X_i, \theta_m), \quad X_i \sim f_*,$$

$$\widehat{I''}(\theta_m) = \frac{1}{K} \sum_{1}^{K} \mathbf{1}(X_i > t)\, W(X_i, \theta_m) W''(X_i, \theta_m), \quad X_i \sim f_*,$$

where the partial derivatives of the weighting function, $W' \equiv \partial W / \partial \theta$ and $W'' \equiv \partial^2 W / \partial \theta^2$, can also be evaluated numerically during the recursions.

When we use more than one biasing parameter, the Newton recursion can be easily adapted to the multi dimensional case by replacing the estimates of the first and second derivatives of the I-function with estimates of the gradient and Hessian of the I-function [96].

Simulation Gain

Simulations performed using biased distributions can provide enormous speed-ups if they are chosen with due care and mathematical precision. Indeed, if applied successfully, simulation lengths needed to estimate very low probabilities become (only) weakly dependent on the actual probabilities. It is thus possible to evaluate any probability in reasonable amounts of simulation time.

A measure to characterize the effectiveness of the chosen IS scheme is the simulation gain Γ. The simulation gain of an IS estimator compared to a conventional MC estimator is defined as the ratio of the simulation lengths required by the MC and IS estimators in order to achieve the same error variance, i.e.,

$$\Gamma = \frac{K_{MC}}{K} = \frac{p_t - p_t^2}{I(\boldsymbol{\theta}) - p_t^2},$$

where K_{MC} and K are respectively the simulation lengths of the MC and IS estimators, and $\boldsymbol{\theta}$ is the vector containing the biasing parameters. In practice, since the quantities involved in the computation of Γ are unknown, the true gain is replaced by an estimate, which is given by

$$\widehat{\Gamma} = \frac{\widehat{p}_t - \widehat{p}_t^2}{\widehat{I}(\boldsymbol{\theta}) - \widehat{p}_t^2}. \tag{1.18}$$

The g-method estimator and Inverse IS

The g-method, first presented in [6] and later extended in [94, 95], is a conditional estimator which exploits knowledge of underlying distributions more effectively, thus yielding a more powerful estimator. Moreover, the g-method estimator can be used with or without IS, in both cases achieving a lower variance than the standard MC or IS ones. An additional advantage is that the *inverse IS* problem (for threshold optimization or selection) can be easily solved. Consider for example a generic CFAR processor, where the test statistic is given by

$$X \underset{H_0}{\overset{H_1}{\gtrless}} \eta Y,$$

where X is, e.g., the envelope or square law value of the CUT and Y is some function of the cells in the CFAR window, e.g., the arithmetic average in the well known Cell Averaging (CA) CFAR processor [79, 97] or geometric mean of the surrounding cells such as in the well known LOG/CFAR detector [85, 87, 88]. In the absence of target in the CUT, the FAP can be written as

$$\alpha_g \equiv P(X > \eta Y | H_0) = E\{P(X > \eta Y | Y, H_0)\} = E\{g(\eta Y)\},$$

where $g(x) = 1 - F(x)$ and F is the cumulative distribution function of X conditioned on Y and H_0. An estimator of the FAP using the g-method can be set up as

$$\widehat{\alpha}_g = \frac{1}{K} \sum_{i=1}^{K} g(\eta Y) W(Y, \theta), \quad Y \sim f_*. \tag{1.19}$$

If the g-method estimator is used without IS, $W(Y, \theta) = 1$ and $Y \sim f$.

The variance of the estimator in (1.19) is given by

$$\mathrm{var}\ \widehat{\alpha}_g = \frac{1}{K} [E_* \big(g^2(\eta Y) W^2(Y, \theta) \big) - p_t^2], \tag{1.20}$$

where the expectation on the RHS is called the I_g-function. Since the function g denotes a probability, it can be easily proved [6] that the variance of the g-method estimator is lower that the variance of a MC or IS estimator using the indicator function instead of the g-function.

Using the g-method estimator the inverse problem, namely that of finding by fast simulation the value of detector threshold η_o satisfying a prescribed FAP α_o, can be readily solved [6]. This is done by minimizing the stochastic objective function

$$J(\eta) = [\widehat{\alpha}_g(\eta) - \alpha_o]^2,$$

An example of the function $J(\eta)$ is shown in Figure 1.4. All detection algorithms that involve a threshold crossing will possess objective functions that have the general behavior shown in Figure 1.4,[7] assuming that the FAP estimate is a decreasing function of its argument η. Minimization of J with respect to η is carried out by the algorithm

$$\eta_{m+1} = \eta_m + \delta_\eta \frac{p_t - \widehat{\alpha}_g(\eta_m)}{\widehat{\alpha_g'}(\eta_m)}, \quad m = 1, 2, \ldots, \tag{1.21}$$

where δ_η is a step-size parameter and the derivative estimator in the denominator is with respect to the threshold η. At convergence, the recursion in (1.21) provides an estimate $\widehat{\eta}_o$ of the sought threshold η_o.

One of the powerful features of adaptive and inverse IS is that the same set of random variables can be repeatedly used for parameters optimization for different threshold settings. Assume that we have a biasing scheme that promises to be effective once the parameters of the biasing distributions have been optimized. If system performance can be characterized in terms of certain random ‘metrics’ (we use the word with a slight abuse of terminology), then these metrics can be pre-computed for a given set of input variables, and used repeatedly (which, in complex systems such as STAP detectors eases the computational burden) in adaptive biasing optimization

---

[7] This can be proven by evaluating the first and second derivatives of $J(\eta)$ at $\eta = \eta_0$ [6].

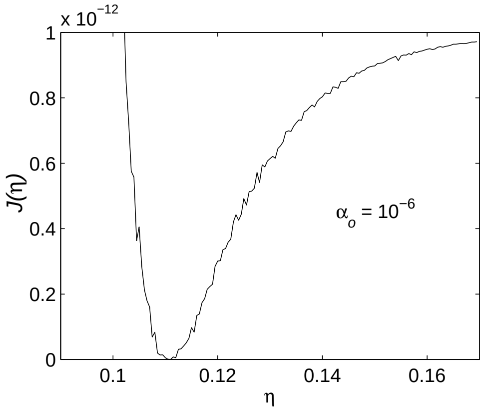


Figure 1.4: Example of IS objective function.

algorithms. These latter algorithms themselves usually require no more than 100 iterations and can be extremely fast. Resulting IS simulation gains can be simultaneously estimated and these tell us whether we need more or less pre-generated variables to achieve certain accuracies. Adjusting this latter number, biasing and system parameter optimization (inverse IS) algorithms can be run, once. Thus there is an initial stage of at most a few steps during which gains are estimated based on pre-computed metrics and the number of these metrics is adjusted.

Research Objectives and Results

The principal objectives in this thesis concerning IS are to examine the viability of using importance sampling methods for STAP detection, develop these methods into powerful analysis and design algorithms, and use them for synthesizing novel detection structures.

Fast simulation using IS methods have been notably successful in the study of conventional constant false alarm rate (CFAR) radar detectors [6, 90, 98–100], and in several other applications. The first paper to appear on application of IS to STAP algorithms was [101]; it laid the groundwork for developing powerful estimation algorithms, based on the work in fast simulation carried out in [93]. In this work also some new detection algorithms that are intended to be robust under various conditions were presented.

Building upon the work in [96, 101] for the space time Adaptive Matched Filter (AMF), we have investigated several STAP detectors from the standpoint of applying importance sampling to characterize their performances. Various biasing techniques have also been devised and implemented, such as biasing by rotation and two dimensional biasing, resulting in significant speed-ups in performance evaluation compared to conventional MC methods. The important problem of detector threshold determination has been addressed and solved by fast simulation. Robust variants such as the envelope-law and geometric-mean detectors for STAP processing have been suggested, their CFAR property established, and performance thoroughly evaluated using IS techniques. It is shown that their detection performances are decidedly better than those of their conventional square-law counterparts when training data are contaminated by interferers, while maintaining almost equal detection performances under homogeneous conditions.

We have also investigated the class of low rank STAP detectors. Being computationally efficient, the low rank adaptive detector is a candidate for future implementation in STAP radar detectors. Subject to an approximation for the disturbance covariance matrix in a clutter dominated scenario, the FAP of the Low Rank NAMF (LRNMF) detector is known via a simple formula, [102] and [103]. However, using fast simulations techniques we have provided an analytical derivation of the exact FAP of the LRNMF detector for data possessing an arbitrary covariance matrix.

The work reported here paves the way to development of more advanced estimation techniques that can facilitate design of powerful and robust detection algorithms designed to counter hostile and heterogeneous clutter environments.

1.3 Outline of the Thesis

The remainder of the thesis is organized as follows:

Chapter 2 introduces the topic of CS and its application to radar. We derive the CS radar signal model and describe some efficient ways of using CS for pulse compression. We also emphasize some of the main issues related to ℓ_1-norm based CS recovery algorithms and the effects of this non linear reconstruction on CS radar detectors design.

Chapter 3 presents two architectures for radar detection using CS. These architectures are designed and analyzed using the Complex Approximate Message Passing (CAMP) algorithm, which is a CS recovery algorithm that was developed in cooperation with the Digital Signal Processing group at Rice University. The CAMP algorithm enjoys several useful statistical properties and allows a complete description of the input/output relations. Using the properties of CAMP, we devise two novel CS CAMP based radar detectors and propose the first, fully adaptive CFAR CS radar detector. The proposed schemes are analyzed and compared by means of ROC curves using both theoretical and simulated results.

Chapter 4 describes the results of a measurement campaign performed at Fraunhofer FHR, in Germany, using Stepped Frequency (SF) radar waveforms for applying CS in range. The performance obtained with the CS adaptive detection schemes presented in Chapter 3 are analyzed and compared. Some insight can be gained about practical aspects of CS for radar detection by observing the experimental results.

Chapter 5 reports on new IS biasing schemes for characterizing the Normalized Matched Filter (NMF) and its adaptive version, the Normalized Adaptive Matched Filter (NAMF). Furthermore, two novel STAP detectors, the Envelope (E) and Geometric Mean (GM) variants of the NAMF detector, are also derived here and their performance analyzed by means of fast simulations. These detectors prove to be robust against interfering targets in the secondary vectors.

Chapter 6 deals with theoretical investigations into the performance of low-rank (LR) STAP detectors. Using the $g-$method estimator, we have been able to characterize the false alarm probability performance of LR STAP detectors in terms of detection thresholds and disturbance backgrounds. FAP simulation results are compared to the theoretical ones. The high simulation gains obtained using the proposed estimator are also shown.

Chapter 7 summarizes the main results of the study which have led to this thesis and also offers recommendations for future work in the areas of CS and IS.

1.4 Chronology

The work on Importance Sampling was carried out in 2005-2006 at the Telecommunication Group, University of Twente, Enschede, The Netherlands. The work was supported by the European Office of Aerospace Research and Development (EOARD), under Award No. FA8655-04-1-3025 and it was carried out under the supervision of Prof. Dr. R. Srinivasan.

The research on Compressive Sensing was performed in 2010-2012 at TNO, The Hague, The Netherlands, at the department of Radar Technology. This part of the research was funded by TNO research programs ‘VP Space Data Utilization’ and ‘V922 Countering Improvised Explosive Devices and Landmines’, under the supervision of Prof.ir P. Hoogeboom and Ir. M. Otten from TNO, and in cooperation with Prof. Dr.-Ing. J. Ender from Fraunhofer FHR, Germany, and Prof. Dr. R. Baraniuk and Dr. A. Maleki from Rice University, Houston.

Chapter 2

Compressive Sensing Applied to Radar

In this chapter we review some concepts from classical radar detection theory and formulate the problem of compressive sensing radar detection. We derive the discrete linear model that is commonly used in the CS framework, and describe the effects of the non-linear ℓ_1-norm minimization recovery on the estimated signal. Based on a statistical analysis of the target and noise characteristics obtained by solving the Basis Pursuit Denoising (BPDN) problem, we can identify several challenges that need to be addressed for the design and analysis of operational CS radar detectors. We devise two possible CS radar detection architectures and perform a preliminary comparison of their performance using Receiver Operating Characteristic curves. [1]

2.1 Radar Signal Model

Consider the case of a one dimensional (1D) radar, which is concerned with the exploitation of the target echoes as a function of range (or time delay) $x(t)$ over an observation interval $t \in (T_{min}, T_{max})$.[2] The radar transmits a Radio Frequency (RF) signal modulated by a waveform $a(t)$, and, in absence of other sources of noise and interference, the received and demodulated signal, which represents the interaction of the transmitted pulse with the observed scene $x(t)$, can be mathematically represented

---

[1] This chapter is based on articles [**C5**], [**C6**], [**C7**] and [**C11**] (a list of the author's publications is included at the end of this dissertation, p. 183.)

[2] For the results in this chapter, the simple 1D model suffices to illustrate our arguments. The extension to azimuth and Doppler domains can be easily obtained in a similar way, see for example [54, 61, 75].

by the convolution

$$y(t) = C \int a(t-\tau)x(\tau)\delta\tau \tag{2.1}$$

where C is a constant including the target Radar Cross Section (RCS), phase terms and propagation effects. The time delay variable t maps to range via the equation $t = \frac{2r}{c}$, where c is the speed of light and r is range (or distance). If there are k point targets located at ranges $r_i, i = 1, \cdots, k$, corresponding to time delays τ_i, the target reflectivity distribution can be expressed as $x(t) = \sum_{i=1}^{k} c_i \delta(t - \tau_i)$, where c_i is the i-th target RCS [59,63,104,105]. Hence, the (complex) baseband, received signal can be rewritten as

$$y(t) = \sum_{i=1}^{k} x_i a(t - \tau_i) \tag{2.2}$$

where $x_i, i = 1, \cdots, k$ is a (complex) amplitude proportional to, amongst others, the target RCS, target distance and transmitted power [106]. Hence, in the remainder of this thesis we consider $|x_i|^2$ as the power received from a target at position i.

As described in Section 1.1 of Chapter 1, in the framework of CS it is most convenient to represent signals in vector form, therefore resorting to a discrete time (and/or space) model. To this end, let the vector $\mathbf{x}$ represent the target response (or scene) at discretized range bins,[3] i.e., $r = [r_1, \cdots, r_N]$, with $r_1 = \frac{cT_{min}}{2}$, $r_N = \frac{cT_{max}}{2} = r_1 + N\Delta R$, where ΔR is the range bin size. Furthermore, assume that targets can only be present at locations corresponding exactly to discrete grid points.[4] Using the Nyquist sampling theorem, the received signal $y(t)$ is sampled at a rate $f_s \geq B$, where B is the bandwidth of the transmitted signal. Then, the sampled received signal $y(t_l), l = 1, \cdots, L$ in (2.2) can be rewritten in vector form as

$$\mathbf{y} = \mathbf{A}\mathbf{x} = \sum_{i=1}^{k} x_i \mathbf{a}_i \tag{2.3}$$

where each column $\mathbf{a}_i, i = 1, \ldots, N$ of the matrix $\mathbf{A}$ is a time delayed version of the sampled transmitted waveform corresponding to the received signal from a target at range bin i,[5] and $\mathbf{x}$ is a length N vector with amplitude x_i at indices i corresponding to target located at ranges r_i and zero elsewhere. Taking into account noise and clutter, which are added to the signal received from the targets, we obtain

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n} \tag{2.4}$$

where we consider $\mathbf{n} \sim \mathcal{CN}(0, \sigma^2\mathbf{I})$.

---

[3] Because of the relation between time delay and range, we will use the two equivalently.

[4] In Section 2.4.2 we will briefly examine the effect of discretization errors (or model mismatch).

[5] For the 1D range case, the matrix $\mathbf{A}$ is determined by the transmitted waveform ($\mathbf{a}$), the range sampling interval ΔR and the received signal sampling frequency f_s. In more general scenarios, using also Doppler and angular information, the sensing matrix $\mathbf{A}$ will depend both on the transmitted waveform and on the sensing geometry.

2.2 Classical Radar Detection

In the most basic operation mode, a radar is concerned with the detection and localization of targets in the observed scene. Over the years, radars have evolved into very complex systems that are capable of not only detecting targets, but also estimate their range, velocity and Direction or Angle Of Arrival (DOA or AOA) with very high accuracy. Furthermore, radars are also used to obtain high resolution images of areas or objects when operating in Synthetic Aperture Radar (SAR) or Inverse SAR (ISAR) modes. In all cases, the final objective of a radar is to learn as much as possible about the observed scene $\mathbf{x}$ via the available measurements $\mathbf{y}$. Many classical books classify radars, based on their tasks, as either imaging, search or tracking radars [107]. Although imaging radars typically have very high resolution, produce a multi dimensional image, and may leave the detection to a human operator for image interpretation, it is sometimes argued in the radar community that, on a more fundamental level, every radar is an imaging radar [105]. In fact, in all applications, a 1, 2 or 3D image estimate $\widehat{\mathbf{x}}$ of the observed scene is obtained before further processing. Whatever the classification, one of the questions to be answered by almost any radar is how many targets are present and where are they located [108]. Therefore, radars are mainly concerned with a joint detection and estimation problem [108–110].

In radar signal processing the detection and estimation of target parameters uses concepts and tools from statistical decision theory, which are similar to the ones used also in other applications, such as communication and image signal processing. However, there are also some major differences that make the radar problem unique. For example, it is hardly enough for a radar to simply declare the presence of a target (using e.g., an energy detector). It is just as important to know how far the target is and at what speed it is moving towards or away from the radar, see ([109], p. 245) and ([108], p.8). Another significant difference is that radar is a two-way active sensor. Whereas in communication systems the message is added at the transmitter side and travels one way, in radar the ‘message’ is added by the targets present in the surrounding environment, and the signal is known. Considering that the free space loss grows very rapidly with range as R^4, the received power is scarce and mostly radars operate at very low (input) SNR. Since the performance, both in terms of detection capability and estimation accuracy of target parameters, are directly dependent on the SNR, it is necessary to use all received power and to design both the transmitter and the receiver to maximize the performance for the given radar task and available resources.

2.2.1 The Matched Filter

It is well known that the Matched Filter (MF) is the filter that optimizes the SNR of a known signal in white Gaussian noise, see e.g., [107]. The impulse response of the MF is a time reversed copy of the (known) transmitted signal, and the range response of a point target after MF is given by the autocorrelation function of the transmitted waveform. The range resolution at the output of the MF is determined by the transmitted pulse length T. For a radar transmitting an unmodulated pulse, the range resolution is given by $\delta R = \frac{cT}{2}$. Hence, improving the range resolution would require shortening the pulse duration, thus resulting in reduced transmitted energy (for a limited fixed peak power). A common way to improve resolution without reducing the pulse length, but rather increasing it, is to use frequency or phase modulated pulses such as Linear Frequency Modulation (LFM) or chirp waveform, Barker codes and pseudorandom noise (PN) sequences. In this case, the output of the MF is a compressed pulse with resolution $\delta R = \frac{c}{2B}$, where B is the bandwidth of the transmitted pulse. Because of this property, the operation of matched filtering is mostly referred to as pulse compression (PC). The SNR gain of the MF after pulse compression is given by the time bandwidth product BT. Although pulse compression by matched filtering result in significant range resolution improvements compared to unmodulated pulses, one of the issues that needs to be addressed is the sidelobes level. In fact, since usually radar systems have very large dynamic range, it is necessary to suppress sidelobes of strong targets in order to detect weaker ones. For example, the autocorrelation function of a LFM pulse exhibits large sidelobes (about -13 dB) with respect to the mainlobe [111]. As we will see in the next section, large sidelobes may result in masking of weaker targets in a multiple targets scenario and in a severe increase of the false alarm rate.

2.2.2 Target Detection

For the detection of a single target (with known parameters) embedded in white Gaussian noise with known variance, the use of statistical decision theory shows that the optimum (Neyman-Pearson) receiver consists of an MF followed by a fixed threshold detector [108, 110, 112]. Using the Neyman-Pearson theorem, it is possible to design the detector threshold to achieve a false alarm probability not exceeding a pre-determined value of, say, α. Observe that the optimum detector, based on MF, is derived for the ideal case of one known signal in white Gaussian noise. This is hardly the case in practical operations, where mostly the target amplitude, phase, time delay (range) and Doppler frequency are unknown.[6] When the target parameters are unknown, a common approach is to set up a Generalized Likelihood Ratio Test (GLRT) for each discrete time delay τ_i. Assuming for the initial target phase a

---

[6] For simplicity of exposition, we limit our discussion to the case of stationary targets, and therefore ignore Doppler frequency shift.

uniform distribution (and averaging it out), we obtain the test statistic [110]

$$|\int_0^T a^*(t-\tau_i)y(t)dt| \underset{H_0}{\overset{H_1}{\gtrless}} \gamma, \tag{2.5}$$

where T is the received signal length.

Thus, the GLRT test consists in computing the envelope (or power) of the MF output at all discrete time delays and comparing it to a threshold to determine the presence (declare hypothesis H_1) or absence (declare hypothesis H_0) of a target at time delay τ_i. Using the discrete linear model in (2.4), we can rewrite the test statistic in (2.5) as

$$|\widehat{x}_i| = |\mathbf{a_i}^*\mathbf{y}|, \quad i = 1, \cdots, N \tag{2.6}$$

where * indicates the conjugate transpose of a vector. In words, MF consists in computing the cross-correlation of the received signal with time delayed versions of the transmitted waveform, for each time delay that we want to test. Combining the MF outputs for all time delays, we can write the MF discrete output signal in vector form as $\widehat{\mathbf{x}}_{MF} = [\widehat{x}_1, \cdots, \widehat{x}_N]^T$, where

$$\widehat{\mathbf{x}}_{MF} = \mathbf{A}^H\mathbf{y}. \tag{2.7}$$

Note that the test statistic $\widehat{\mathbf{x}}_{MF}$ is a noisy estimate of the targets range response.

The envelope (or power) of each component of the vector $\widehat{\mathbf{x}}_{MF}$ is compared to the threshold γ to decide upon the presence of a target. For the envelope detector, the threshold γ should be set to $\gamma = \sqrt{-\sigma^2 \ln \alpha}$ to achieve a FAP equal to α [107].

Detection of Multiple Targets

Mostly, although the number of targets might be significantly smaller than the number of resolution cells, more than one target is present in the received window. As explained above, the task of the radar is to determine how many targets are present and to estimate their locations.

If an MF is used at the receiver, expanding (2.7) we obtain

$$\widehat{\mathbf{x}}_{MF} = \mathbf{A}^H\mathbf{y} = \mathbf{A}^H(\mathbf{A}\mathbf{x}+\mathbf{n}) = \mathbf{x} + (\mathbf{A}^H\mathbf{A} - \mathbf{I})\mathbf{x} + \mathbf{A}^H\mathbf{n}. \tag{2.8}$$

From the above equation it can be seen that each entry $\widehat{x}_i$ of the MF output signal is the sum of the true target response at location i (x_i), plus the interference caused by the presence of other possible targets at locations $j \neq i$ ($\sum_{j=1,j\neq i}^N x_j \mathbf{a}_i^*\mathbf{a}_j$), plus Gaussian noise. Clearly, the interference is proportional to the cross-correlation between time delayed version of the transmitted waveform [113].

In the ideal case that time delayed versions of the transmitted waveform are orthogonal and have unit norm columns, i.e., $\mathbf{A}^H\mathbf{A} = \mathbf{I}$, we obtain

$$\widehat{\mathbf{x}}_{MF} = \mathbf{x} + \mathbf{z}. \tag{2.9}$$

where $\mathbf{z} \sim \mathcal{CN}(0, \sigma^2\mathbf{I})$.

Therefore, in the orthogonal case the components of the vector $\widehat{\mathbf{x}}_{MF}$ are independent of one another and each range bin can be treated independently. Applying a binary hypothesis test at each time delay will produce an estimate of how many targets are present and their time delays.[7]

However, for practical frequency or phase modulated waveforms the orthogonality condition is never met. Furthermore, as targets are never exactly on a grid point, there will also be discretization errors. These errors are often referred in the radar literature as range gate straddling loss or scalloping error [107, 114]. Hence, the presence of a target interferes, through the sidelobes, with the detection of another. A reduction of the sidelobes level (SLL) can be accomplished by applying a weighting function during matched filtering. However, the weighted MF output is no longer matched to the transmitted signal, and therefore, while reducing the SLL it also reduces output SNR. Alternatively, one can design waveforms with low sidelobes levels [115–117] or use different pulse compression filters, such as Adaptive Pulse Compression (APC) [104, 105] or mismatched filters [118–122]. A review of mismatched filters can be found in [113]. Such filters can be obtained using standard convex optimization algorithms and can be based on the minimization of different ℓ_p-norms of the sidelobe energies vector [123]. In [123] it is shown that, the optimum filter that minimizes the Integrated Sidelobe Level (ISL) is obtained by minimizing the ℓ_1-norm of the vector whose elements are the sidelobe energies. Note that the design of mismatched filters is based on iterative algorithms and the optimum filter weights have to be estimated separately for each range bin. Therefore, such an approach is more computationally complex than the MF one.

Furthermore, if the sidelobes are not sufficiently suppressed by the receive filter, performing a binary GLRT at each range bin independently of the others does not take into account the interaction between targets. In this case, the theoretically optimum detection/estimation strategy is the multiple hypothesis test, considering all possible combinations of number of targets and locations. This however, would require to compute 2^N probabilities, one per each hypothesis. Clearly, this would become prohibitive as the size of the problem N increases. A similar problem is encountered in asynchronous communication channels, and is referred herein as multiple users interference. In a paper by Verdú [124], it is shown that the optimum detector for this scenario is the multiple hypothesis test. However, this is NP-hard to compute. Thus, although not optimal, in practical applications a weighted MF followed by a detector is still the most popular approach.

---

[7] Note that, as described in the introduction, the GLRT acts as both a detector and an estimator.

2.2.3 Constant False Alarm Rate detectors

The fixed threshold γ in (2.5), which fixes the FAP, depends upon the noise power σ^2. In practice, however, the noise power is varying and not known in advance. Hence, in classical radar detectors a CFAR processor is employed. In CFAR schemes the Cell Under Test (CUT) $\hat{x}_i$ (which corresponds to the output of the receive filter at time delay τ_i), is tested for the presence of a target against a threshold that is derived from an estimated clutter plus noise power. The $2N_w$ cells (CFAR window) surrounding the CUT are used to derive an estimate of the local background and they are assumed to be target free. Commonly, $2N_G$ guard cells immediately adjacent to the CUT are excluded from the CFAR window. The great advantage of CFAR schemes is that they are able to maintain a constant false alarm rate via adaptation of the threshold to a changing environment.

The general form of a CFAR test is

$$X \underset{H_0}{\overset{H_1}{\gtrless}} \beta Y, \tag{2.10}$$

where the random variable X represents some function (generally envelope or power) of the CUT $\hat{x}_i$, β is a threshold multiplier that controls the false alarm rate, and Y is also a random variable function of the cells in the CFAR window $[\hat{x}_{i-N_w-N_G}, \cdots, \hat{x}_{i-N_G-1}, \hat{x}_{i+N_G+1}, \cdots, \hat{x}_{i+N_w+N_G}]$. In the well known Cell Averaging CFAR (CA-CFAR) detector, Y is the average of the cells in the CFAR window [79, 84, 97, 125–127]. The CA-CFAR detector is optimum in the presence of homogeneous i.i.d. Gaussian noise. However, in situations in which the clutter changes rapidly or in the presence of interfering targets in the CFAR window, or when the clutter and noise distribution are not Gaussian, the CA-CFAR detector performance degrades severely. For this reason many alternative CFAR schemes have been devised, such as Greatest Of (GO), Smallest Of (SO), Trimmed Mean (TM), Logarithmic (LOG), and Order Statistic (OS) CFAR processors [79–83, 85, 86, 88, 89, 98]. Each of these CFAR schemes has some optimality property in a specific clutter and interference scenario.

In most CFAR detectors some assumptions are made about the distribution of the noise[8], which is assumed to be known to within some shape and/or scale parameters that can be estimated from the data. Depending on the characteristics of the expected noise and interference scenario, the most appropriate CFAR scheme can be selected or designed. Clearly, one has to know the relation between the CFAR threshold multiplier and P_{fa}, so that β can be adjusted to maintain P_{fa} constant during the observation time. Hence, the noise distribution should be known to design a CFAR scheme.

---

[8] From this point on, the term noise will be used to signify noise and/or clutter with white Gaussian distribution.

The design of CFAR-like detectors for CS radars is one of the challenges tackled in this work. As we will see in the next section, due to the characteristic of the signal recovered from ℓ_1-norm minimization algorithms, a new approach to adaptive CFAR detection for CS radar seems to be necessary.

2.3 CS applied to radar

As explained in Section 1.2.1, compressive sensing is a novel technique for data acquisition and processing that allows reconstruction of sparse signals from a number of measurements much smaller than the one dictated by the Shannon-Nyquist sampling theorem. CS exploits the sparsity or compressibility of a signal to reduce both the sampling rate (while keeping the resolution fixed) and the amount of data generated. The prior information about sparsity is used in the reconstruction phase, which is based on constrained ℓ_1-norm minimization as opposed to more conventional least squares [4, 7, 11, 27].

The possibility of using CS as a means to reduce the acquisition time or else to achieve better performance with a reduced set of measurements has attracted the interest of many researchers in several fields. Recent publications have demonstrated the applicability of CS to optical and magnetic resonance imaging, remote sensing, communications and radar, see [128] for an extensive list of publications.

In many radars the increasing demands in terms of resolution (bandwidth), and trends towards multi-channel systems with less analog and more digital processing, keep increasing the required sampling rates and amounts of data to be handled. As the number of targets is typically much smaller than the number of resolution cells in the illuminated area or volume,[9] the prerequisite on the signal sparsity is often met in many radar applications. Based on the target sparsity in the range-, angle-, and/or Doppler domain, CS represents a valid alternative to conventional acquisition/processing schemes with the advantage of relaxing some system requirements. An overview of radar modes with corresponding CS applicability was presented in [129], and it is summarized here in Table 2.1. This table shows that CS is most suitable for radar applications where the observed scene is sparsely populated, such as air surveillance and SAR tomography.

Much of the work on applications of CS to radar has focussed on demonstrating the capability of recovering the target scene from a set of CS measurements using different algorithms, sensing matrices, and sparsity domains [53–75, 131–133]. Particularly, the work of [54] has established the fundamental limit of achievable resolution when using CS to improve upon the resolution that can be achieved with classical imaging techniques, such as MF. In [60] the achievable resolution of CS recovery was also

---

[9] An exception to this assumption is the case of SAR imaging radars, as described in Table 2.1.

related to the grid discretization. We will see with an example in Section 2.4.2 and with the experimental measurements in Chapter 4, that the super resolution achieved using CS breaks down if targets are not located exactly on the discrete grid points that are used to set up the sensing matrix.

In the work presented in this thesis, we focus on the design and analysis of CS radar detectors rather than on CS high resolution radar imaging. Particularly, we are interested in determining a strategy for optimal detection of targets from CS measurements, designing an adaptive CFAR detector to achieve a desired pair (P_d, P_{fa}) in a detection problem, and understanding the trade-offs between amount of compression and SNR/detection capabilities. Moreover, the way to perform adaptive detection in a CS framework is a problem that ultimately has to be addressed also for some imaging radars where the estimated scene is not the final result (as it might be in some SAR or ISAR applications) but it is just an intermediate step before automatic target detection is performed. To design an adaptive detector in the CS framework, one has to deal with a number of issues that are related mostly to the non-linearity of the ℓ_1-recovery. These problems will become clear as we proceed.

Throughout the remainder of this thesis we concentrate our attention on the case of a 1D radar operating in the range domain, when observing a single transmitted pulse (or equivalently a burst for stepped frequency radar). We consider mainly two models

Table 2.1: Radar operations and corresponding CS applicability [129].

Radar Task	**Scene Sparsity**	**CS Domain**	**Applicability**
Air surveillance [61]	Few targets	Thinned array	+++
MTI [61, 70, 77]	Few targets	Reduced slow time sampling	++
Range profile [59, 130] (isolated vehicles)	Few dominant scatterers	Reduced TX frequencies	+
ISAR (isolated vehicles) [53–55, 63, 75, 76]	Few dominant scatterers	Reduced TX frequencies/angular positions	++
SAR imaging [56,57, 66]	Not very sparse	-	-
SAR tomography [68, 69, 71, 72]	Few elevation angles with reflections	Across track array/number of tracks	+++
MIMO imaging [58, 60, 65, 67]	Few dominant scatterers	Reduced TX/RX units	++

for the sensing matrix $\mathbf{A}$. Specifically, we will consider $\mathbf{A}$ to be either a matrix with random i.i.d. entries (either Gaussian or Bernoulli) or a partial DFT.[10] Because of their well-established properties, such as the RIP, these two models have been widely used in the CS radar literature. From a system point of view, a random matrix $\mathbf{A}$ is obtained when the transmitted waveform is a PN sequence and the received signal is digitized at a sampling rate lower than the Nyquist rate [53,63,73]. This approach is referred to as the 'random filter' in [53]. The partial DFT matrix instead is obtained when transmitting a stepped frequency waveform in which the transmitted frequencies are non uniformly spaced over the total transmitted bandwidth B [61,75,130,134]. In Section 2.3.1 we make an explicit connection between the partial Fourier matrix and the received signal from a SF radar.

For multidimensional radars, CS can be applied to just one or more domains, as long as the signal permits a sparse representation in the domains where CS is applied. Furthermore, depending on the transmitted waveform, there might be cases where the application of CS in a given domain results in a significant performance gain in another domain. An example of this is shown in [77], where it is demonstrated that applying CS in range can results in either an increased Doppler unambiguous range or increased Doppler resolution.

Notation

Let $|\alpha|$, $\measuredangle\alpha$ and α^* denote the amplitude, phase, and conjugate of $\alpha \in \mathbb{C}$ respectively. Furthermore, for the matrix $\mathbf{A} \in \mathbb{C}^{n\times N}$, $\mathbf{A}^H$, A_i, A_{ij} denote the Hermitian, i^{th} column and ij^{th} element of matrix $\mathbf{A}$, respectively. We are interested in detecting the number and locations of targets present in the k-sparse signal $\mathbf{x} \in \mathbb{C}^N$ from an undersampled set of n noisy linear measurements $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}$, with $n \ll N$. $\mathbf{A} \in \mathbb{C}^{n\times N}$ has either i.i.d. random elements (with independent real and imaginary parts) from a given distribution that satisfies $\mathbb{E}(A_{ij}) = 0$ and $\mathbb{E}(|A_{ij}|^2) = \frac{1}{n}$, or is a partial DFT matrix with unit norm columns. $\mathbf{n} \in \mathbb{C}^n$ is the measurement noise vector with i.i.d. components distributed as $\mathcal{CN}(0, \sigma^2)$. Furthermore, we define the compression factor $\delta = n/N$ and the relative signal sparsity $\rho = k/n$. Also, we assume that the elements of $\mathbf{x}$ are i.i.d. distributed as $x_i \sim (1-\rho\delta)\delta(x_i) + \rho\delta G(x_i)$, where G denotes the probability distribution of the non-zero coefficients and $\delta(x_i)$ is the Dirac delta function.

---

[10] An $n \times N$ partial Fourier matrix is obtained from an $N \times N$ discrete Fourier transform matrix by preserving only a random subset n of the original N matrix rows.

2.3.1 Stepped Frequency waveform for CS in range

As described in Section 2.2.1, to achieve high range resolution, the transmitted waveform must be a chirp, SF or phase coded signal. One approach to applying CS in range could be to subsample the received signal $y(t)$ in the time domain by, for example, choosing only a subset of the Nyquist sampling instants t_l or by using the Random Demodulator of Tropp *et al.* [135]. However, subsampling the received signal in the time domain will necessarily lead to an SNR loss, [59,135]. The loss of SNR, although it can very well be tolerated in certain applications, is unacceptable in most radar systems, which are already working at critical SNRs [107].

For example, it is known that at the output of the MF the SNR gain is equal to the number of integrated samples, assuming the noise bandwidth is sampled at Nyquist rate. While oversampling above the Nyquist rate does not bring any advantage, due to the noise correlation, undersampling reduces the MF gain, because less samples are integrated. In terms of final SNR, throwing away time samples is equivalent to throwing away transmitted power, and therefore does not appear to be a good strategy. The method proposed in [59] suggests that a good strategy to perform CS range pulse compression is to transmit only a subset of frequencies and to sample only the transmitted frequencies at the receiver, so that also the noise is sampled only on the selected TX frequencies and not over the whole bandwidth, as opposed to undersampling in time. Using the methods proposed in [59], i.e., sequential or parallel transmission of a reduced set of frequencies,[11] also the TX waveform is designed to perform CS, as opposed to time undersampling using the random demodulator. In fact, although we cannot assign power to transmitted time samples, we can concentrate all the transmitted power on the subset of transmitted frequencies, so that all transmitted power is equally divided amongst the measured (frequency) samples.

In SF radar (without CS), the transmitted waveform consists of a burst of N pulses, each pulse being a rectangular pulse of duration τ modulated at the carrier frequency $f_n = f_1 + n\Delta f$, $n = 1, \cdots, N$, where Δf is the separation between sequential transmitted frequencies. The bandwidth of a single pulse is equal to $B_f = 1/\tau$. An example of the TX waveform is shown in Figure 2.1.

After reception and demodulation, each range bin maps to N phases proportional to the N transmitted frequencies [61, 136]. The N samples y_n of the measurement vector $\mathbf{y}$ are given by

$$y_n = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} e^{-j2\pi f_n 2r_i/c} x_i + n, \tag{2.11}$$

---

[11] The model used in our simulations is the same whether the transmission of the selected frequencies is performed sequentially or simultaneously. Therefore, in the remainder of this thesis, we refer in general to subsampled SF waveform.

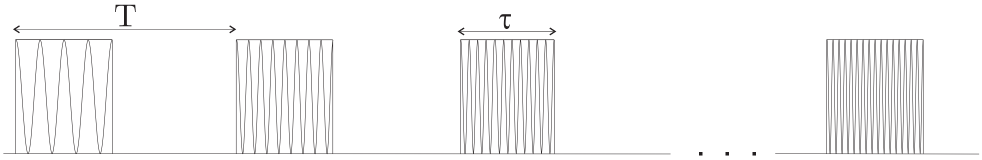


Figure 2.1: Stepped frequency waveform.

where $r_i = r_1 + i\,\Delta R$, $i = 1, \dots, N$, is the range bin index and ΔR is the range bin size. Finally, $\{x_i\}_{i=1}^{N}$ are the components of the sparse target vector $\mathbf{x}$ containing the k complex target responses at indices corresponding to target range bins and zeros elsewhere. Therefore, we can rewrite (2.11) in matrix form as in (2.3), where the sensing matrix $\widetilde{\mathbf{A}}$ (without CS) is the normalized $N \times N$ DFT matrix. This implies that, classical MF processing is equivalent to an Inverse DFT.[12]

Note that, contrary to classical pulsed radars, in SF radar one high resolution measurement is obtained by processing together all the pulses in a burst (or, equivalently, each pulse in the burst provides a single frequency measurement). This results in an equivalent total burst bandwidth $B = N\Delta f$. The range resolution is therefore given by $\delta_R = c/2B$. Differently from pulsed/chirped radar, where the maximum unambiguous range is determined by the PRF, in SF radar the frequency step size (Δf) translates into a maximum unambiguous range of $R_{un} = c/2\Delta f$ [107].

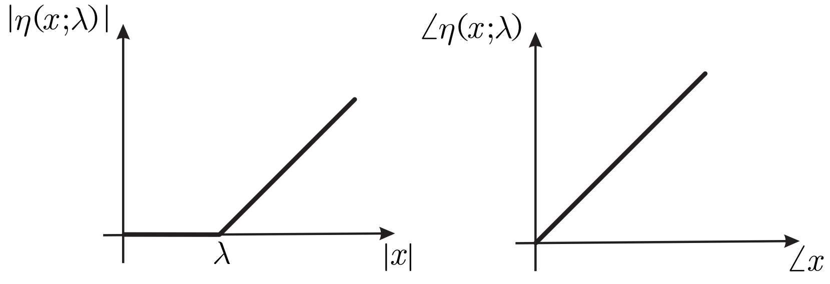
To apply CS, the number of discrete frequencies which are transmitted is reduced from N (which represents the Nyquist rate for unambiguous mapping of ranges to phases over the whole bandwidth) to n, with $n \ll N$. Consequently, if we keep the pulse length (τ) and the intra-pulse spacing (T) unchanged, the duration of the total burst is also reduced of a factor δ. In the remainder of this thesis, when using CS the subset of transmitted frequencies is selected uniformly at random from the original set of Nyquist frequencies. Therefore, the CS sensing matrix $\mathbf{A}$ consists of a random subset of rows of $\widetilde{\mathbf{A}}$.

2.3.2 ℓ_1-norm recovery

As described in Section 1.2.1, a tractable approach to finding the sparsest solution to the underdetermined system of linear equations in (1.7) is to solve the Basis Pursuit for the noise free measurement case and the LASSO or BPDN problem [138] for the the case of measurements corrupted by additive noise.

---

[12] Note that, formulating the relation in (2.11) without the use of complex numbers would be extremely unwieldy. Therefore, we consider only algorithms that can deal with signals in the complex domain. To this end, in Chapter 3 we will introduce the CAMP algorithm, which is an extension of the AMP algorithm originally presented by Donoho *et al.* [137] to the case of signals and measurements in the complex domain.



(a) Amplitude (b) Phase

Figure 2.2: Complex soft thresholding function.

The cost function in (1.9) is convex and can be solved by standard techniques such as interior point or homotopy methods [29, 30, 139]. However, these approaches are computationally expensive, and therefore researchers have considered several iterative algorithms with inexpensive per-iteration computations. For more information on these algorithms, see [37, 140] and the references therein. These methods rely on the fact that the optimization problem $\arg\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{u}-\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1$ has the closed form solution $\mathbf{x} = \eta(\mathbf{u};\lambda) \triangleq (|\mathbf{u}| - \lambda)e^{j\measuredangle\mathbf{u}}\ \mathbf{1}(|\mathbf{u}| > \lambda)$, where $\eta(\cdot;\lambda)$ is called the complex soft thresholding function and $\mathbf{1}$ is the indicator function. The complex soft thresholding function acts (componentwise) on the amplitudes of the input vector $\mathbf{u}$ and produces a sparse signal by shrinking to zero all the elements of $\mathbf{u}$ whose amplitude is below the threshold, or regularization parameter, λ, thus enforcing sparsity on the solution. The components of $\mathbf{u}$ that are above the threshold will be biased towards zero by an amount equal to the threshold λ, and their phase is unchanged by the soft thresholding. The complex soft thresholding function is shown in Figure 2.2.

In the most general case, algorithms to solve (1.9) that are based on Iterative Soft Thresholding (IST) use the following iterations. Starting with $\widehat{\mathbf{x}}^0 = 0$, at each iteration t the estimate $\widehat{\mathbf{x}}$ of the vector $\mathbf{x}$ is updated using

$$\widehat{\mathbf{x}}^{t+1} = \eta\left(\widehat{\mathbf{x}}^t + \mathbf{A}^H(\mathbf{y} - \mathbf{A}\widehat{\mathbf{x}}^t);\lambda\right). \tag{2.12}$$

Therefore, at each iteration, the current residual is projected along the waveforms, and added to the previous solution. In other words, at each iteration of (2.12), the algorithm moves in the negative gradient direction (of the objective function) $-\mathbf{A}^H(\mathbf{y} - \mathbf{A}\widehat{\mathbf{x}}^t)$, (where the terms between brackets is known in the signal processing literature as the residual), and it applies the soft thresholding function to enforce sparsity on the solution. Therefore, at convergence, the estimated sparse signal $\widehat{\mathbf{x}}$ will contain many zero components, and a few non-zeros, depending on the threshold parameter λ, the SNR, δ and ρ. Furthermore, if we fix the values of δ, ρ, and SNR, the solution of (2.12) will depend on the free parameter λ. It is clear that, as λ

increases, the solution to (2.12) will be more sparse. Furthermore, if $\lambda > \|\mathbf{A}^H\mathbf{y}\|_\infty$, the only feasible solution is $\widehat{\mathbf{x}} = 0$.

Let $\widehat{\mathbf{x}}_\lambda$ be the solution of (2.12) with parameter λ. The Mean Square Error (MSE) is defined as

$$\mathrm{MSE}_\lambda = \frac{\|\widehat{\mathbf{x}}_\lambda - \mathbf{x}\|_2^2}{N} \tag{2.13}$$

Since the quality of the solution to (2.12) (in terms of MSE) varies significantly with the (free) parameter λ, its value should be selected very carefully. However, the relation between λ and the MSE of the solution is not known and data dependent. One of the major issues to be addressed for the detection problem is how to set this free parameter adaptively and according to some criterium of optimality.

In [27] it is suggested that setting the parameter $\lambda = \sigma\sqrt{2\log n}$, where n is the number of CS measurements, produces a solution that has some optimality properties in the MSE sense. However, there is no closed form relation between this value and the MSE. In [141, 142] the authors proposed three different methods for adaptive estimation of the threshold or hyper-parameter λ for SAR imaging. The first and second method, respectively the Stein's unbiased risk estimator (SURE) and Generalized Cross Validation[13] (GCV), are based on iterative algorithms that estimate and minimize the predictive risk $\|\mathbf{A}\widehat{\mathbf{x}}_\lambda - \mathbf{A}\mathbf{x}\|_2^2$. The third method, is based on estimating the L-curve, and locating its corner. All three algorithms perform relatively well at high SNRs, however, since the minimum MSE is a better measure of the reconstructed image quality, in situations when the two measures differ significantly the proposed algorithms will fail to estimate the optimal parameter. Furthermore, these algorithms are computationally very expensive, as they require estimation and inversion of potentially large matrices at each iteration. To alleviate the computational burden of such methods, in [144] a method is proposed that is based on the estimation of the kurtosis of the restored image. The proposed method is both computationally less costly than the methods proposed in [141, 142], and does not require any prior knowledge of the input noise variance, as it is the case for the SURE estimator. Both the method proposed in [144] and the methods proposed in [141, 142] are based on optimizing some function that measure the quality of the recovered image. Still, *the relation between the threshold and the detection or false alarm probabilities of the recovered signal is not known*, and it is not understood how to set the parameter for achieving a design performance or maintaining a constant false alarm rate.

Another popular algorithm that is used to solve (1.8) is the Spectral Projected Gradient ℓ_1 (SPGℓ_1) method [28]. This algorithm is intuitively easier to use in terms of threshold selection. In fact, from the formulation in (1.8) it is clear that the threshold parameter ε is related to the noise variance σ. Particularly, ε should be set to some constant C times the noise std. In the following section we will see how

---

[13] A similar method based on cross validation was also previously proposed in [143].

the multiplicative constant C plays an important role in terms of detection and false alarm performance of a CS radar detector. The SPGℓ_1 algorithm (available online at `http://www.cs.ubc.ca/labs/scl/spgl1/`) will be used to perform some of the simulations presented in the remainder of this chapter.

Properties of the recovered signal

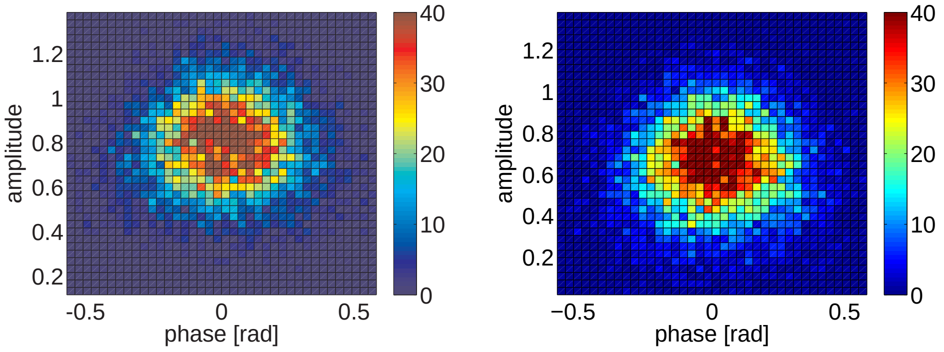
For the design of possible CS radar detectors, it is essential to determine the (statistical) properties of the signal obtained from ℓ_1-norm minimization algorithms. To this end, we report here some examples of the 2 dimensional distribution of the reconstructed signal amplitude and phase. In the following figures we will use a single target with amplitude 1 and phase equal to zero. The SNR here is defined as $\frac{|\hat{x}_i|^2}{\sigma^2}$, which is equal to the SNR that we would obtain after a MF with full Nyquist measurements. The input (per sample) SNR of the CS measurements can be derived as $\text{SNR}_{in} = \frac{\text{SNR}}{n}$. The number of range bins is fixed to $N = 200$, and the number of CS measurements varies. The noise is Gaussian with i.i.d. entries with variance σ^2. The results are obtained using 10000 Monte Carlo realizations.

Figure 2.3 shows an example of the 2 dimensional histogram of the reconstructed target amplitude and phase using 1) SPGℓ_1 with $\varepsilon = \sigma$ (the input noise variance is assumed to be known in the following simulations), 2) IST with parameter $\lambda = 0.25$.

From Figure 2.3 we can make the following observations.[14] First, the estimated amplitude and phase are uncorrelated. Second, the soft thresholding function biases the estimated target amplitude by an amount proportional to the threshold ε or λ, depending on the algorithm used. For certain applications, such as imaging or clustering, it may be important to reduce the bias on the estimated amplitudes of detections. A possible way to do this is by applying least square estimation on the reconstructed non-zero samples of $\widehat{\mathbf{x}}$. This operation is referred to in literature as de-biasing [63,145]. Finally, we observe from Figure 2.3 that the (mean) target phase is preserved, as the complex soft thresholding function does not act on the input signal phase. This aspect, although not relevant for detection (as only the envelope of the signal is used), is very important if any coherent processing has to take place after CS recovery. In fact, it is common in radar to have at least two stages of coherent processing, usually range and Doppler. Preferably, the coherent summation should be performed before CS recovery, as this would maximizes the processing gain. However, if range compression must be performed before Doppler compression, for example to compensate range migration, then it becomes important that the signal phase is preserved.

---

[14] We conducted many simulations for different values of the recovery thresholds, target amplitudes, SNR, δ and ρ and obtained consistent results.



(a) SPGℓ_1 (b) IST

Figure 2.3: 2 dimensional histogram of CS estimated target amplitude and phase. SNR=12dB, $N = 200$ and $\delta = 0.33$. True target amplitude is equal to 1 and true target phase is equal to 0.

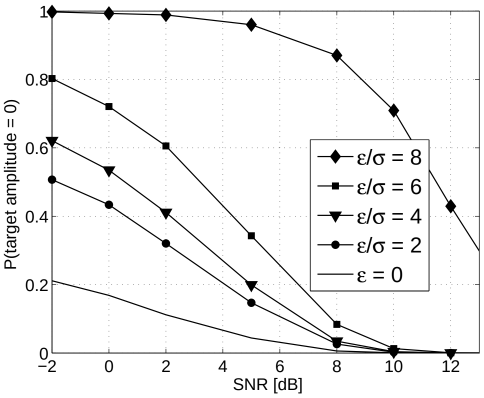
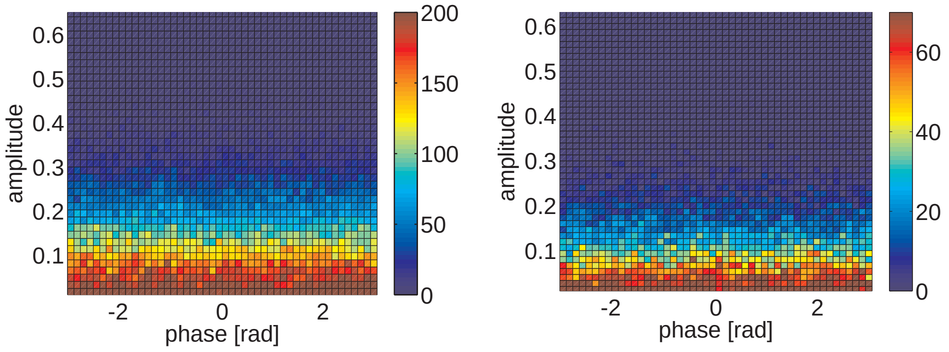


Figure 2.4: Probability of having zero amplitude at target position using SPGℓ_1 versus SNR for $\delta = 0.25$ and different values of the parameter ϵ.

Furthermore, although it can not be seen from Figure 2.3, another important phenomenon associated with the non-linear soft thresholding operation is that, especially as the SNR decreases, the recovered amplitude at the target location might be equal to zero. If this happens, there is no way to recover the target, not even with subsequent coherent processing, as the target is simply lost. Figure 2.4 shows the probability of having a reconstructed target amplitude equal to zero versus SNR for different values of ε and $\delta = 0.25$.

Figure 2.5 shows the 2 dimensional histogram of the recovered non-zero samples at



(a) SPGℓ_1 (b) IST

Figure 2.5: 2 dimensional histogram of CS estimated amplitude and phase at the target free range bins. $N = 200$ and $\delta = 0.33$. The input noise is i.i.d. Gaussian with variance $\sigma^2 = 0.063$. The thresholds are $\lambda = 0.25$ for IST and $\varepsilon = \sigma$ for SPGℓ_1.

locations not containing the targets, i.e., the recovered noise samples. For the recovery, we use the same parameters as in Figure 2.3. We observe that, most of the noise samples are reconstructed with zero amplitude. For the non-zero samples, the phase is uniformly distributed between $[-\pi, \pi]$, as for a standard Gaussian distribution. However, some of the noise samples are reconstructed with non-zero amplitude, and therefore, if no further processing is applied, they will produce false targets in the recovered image or range profile. Increasing the threshold will reduce the number of non-zero noise recovered samples, but it may also result in suppressing the targets, especially as the SNR decreases. Note that the difference (in scale) in the histograms obtained with IST and SPGℓ_1 is due to the threshold selection. As described in Section 1.2.1, the relation between λ and ε that makes the 2 solutions equal is not known and data dependent. We manually selected 2 values that provided comparable results.

2.4 CS Radar Detection

So far, the main focus of the CS literature has been on the reconstruction quality in terms of MSE. Furthermore, most of the theoretical results provide only sufficient conditions on the MSE with unknown/large constants that are not useful for practical applications [4, 78, 145, 146]. In radar instead, the primary performance parameters are the detection and false alarm probabilities. As explained in the previous sections, classical radar architectures (without CS) use well-established signal processing algorithms and detection schemes, such as MF and CFAR processors. CS instead relies on the nonlinear ℓ_1-minimization for reconstructing the target scene. In the existing CS recovery algorithms the threshold or regularization parameter (ε or λ depending on the algorithm) affects the quality of the reconstruction result and for the best performance it is essential to optimally tune this parameter [140]. Furthermore, the recovery threshold depends on the input noise level as $\varepsilon = C\sigma$. Therefore, it is necessary to not only determine the appropriate constant as a function of the desired output performance but also to accurately estimate the noise level from the CS measurements.

The problem of noise estimation is also encountered in classical radar detectors, where the detection threshold adapts to the environment using a CFAR scheme. *The design of CFAR-like schemes for CS radar systems has not been addressed so far in the literature, as the relations between the false alarm probability/noise statistics and the parameters of the recovery algorithm are not known.* In fact, even if we would know how to set the constant C to achieve a design pair (P_d, P_{fa}), a way to adapt the detector to the unknown environment is still to be found and far from obvious, considering the non-linearity of the recovery algorithm and the fact that the reconstructed signal has only few non zero samples. This makes it very difficult to estimate any statistics from it. Clearly, if a CS based radar detector should be designed, the way to set the reconstruction and detection parameters while maintaining CFAR must be established. Our objective is to design a CFAR CS radar detection scheme whose performance can be predicted and that shows optimality in some sense.

Several papers in literature have investigated the problem of detecting not the presence of targets but rather the presence of a signal (which may consists of the sum of independent signals received from multiple targets) from a set of CS measurements [147–153]. In the papers by Davenport *et al* [149, 150] the authors refer to CS detection as the problem of deciding upon the presence of a known signal $\mathbf{Ax}$ from CS measurements. Following standard statistical decision theory, the binary hypothesis test is formulated as

$$\begin{aligned} \mathrm{H}_0 &: \mathbf{y} = \mathbf{n} \\ \mathrm{H}_1 &: \mathbf{y} = \mathbf{Ax} + \mathbf{n}. \end{aligned} \tag{2.14}$$

Defining the signal $\mathbf{s} = \mathbf{Ax} \in \mathbb{C}^n$, (2.14) can be rewritten in the same formulation that is used in classical detection theory. It is well known that the optimum detector consists of a MF followed by a fixed threshold, where the filter is now matched to the signal $\mathbf{Ax}$. The authors refer to this approach as the Compressive Matched Filter (CMF). The test statistic is given by

$$T = \mathbf{y}^* \mathbf{s} \tag{2.15}$$

and its absolute value is compared to a threshold γ to decide which of the two hypothesis is in force. Clearly, such an approach is not applicable to radar, as we do not know the number and the locations of possible targets in advance, i.e., $\mathbf{x}$ is not available.

A similar derivation is proposed in [148]. Here the authors also consider two options for the more practical scenario when $\mathbf{x}$ is unknown. The first option is to use an energy detector to decide between the two hypothesis. This consists in comparing $\|\mathbf{y}\|_2^2$ to a threshold to decide if a signal $\mathbf{s}$ is present in the measurements or not. However, this approach does not allow to estimate the targets locations, as it performs a single binary test over the whole received signal. Otherwise, if an estimate $\widehat{\mathbf{x}}$ of the signal is available through 'future knowledge', this estimate can be used in the statistic T in (2.15).

In [151, 152] the authors propose a subspace compressive detector, again to decide between the two hypothesis in (2.14). In the first scenario it is assumed that the k non-zero components of the signal $\mathbf{x}$ are known and they are used to design a sensing matrix matched to the signal subspace. Later, for the case of unknown signal $\mathbf{x}$, the authors propose to solve the LASSO problem to derive an estimate of the signal subspace.

In [147] the authors propose the Incoherent Detection and Estimation Algorithm (IDEA) for detecting the presence of a signal embedded in white Gaussian noise. This algorithm uses a modified version of the greedy MP algorithm to estimate the strongest sparse signal coefficients. Then, the highest coefficient is compared to a threshold.

An important difference between our work and the work of [147–152] is that we are not only interested in estimating the presence of the (unknown) signal $\mathbf{x}$ but, as it is always the case in radar, we want to detect the number of non zero coefficients in $\mathbf{x}$, i.e., the number of targets, and estimate their locations, i.e., the position in the vector $\mathbf{x}$. A way to achieve our goal is to obtain an estimate of the target range response $\widehat{\mathbf{x}}$ that can be further used as input to a conventional binary hypothesis test. Since the targets are resolved, using each range bin estimate in a binary test produces an estimate of the number of targets present in the received signal and their locations.

A similar approach is also presented in [153], although it is assumed that a single target is present. Here the authors compare three different ways for detecting the presence of one target with known time delay (that fixes the range bin index i), but unknown amplitude and phase. The results presented in [153] are based on transmitting a fully sampled chirp signal, and using a random demodulator at the receiver to obtain the sub-sampled CS measurements. The first method is the MF as given in (2.6) with the full set of Nyquist measurements. The second approach, that is also referred to as the CMF, computes the statistic

$$\widehat{x}_{i\mathrm{CMF}} = \mathbf{a_i}^{*}\mathbf{y}. \tag{2.16}$$

Hence, $\widehat{x}_{i\mathrm{CMF}}$ is an estimate of the target response at bin i obtained by matching the received signal to the time delayed version of the undersampled transmitted waveforms. For the case of detection from a single pulse (or burst), as e.g., in a CFAR processor, the false alarm rate at range bins other than the target is significantly increased compared to the MF case. Furthermore, if multiple targets are present, this approach will suffer form severe targets sidelobes, as time delayed versions of the transmitted waveform are not orthogonal in the CS case, i.e., $\mathbf{A}^H\mathbf{A} \neq \mathbf{I}$. This will result in an increase of false alarms and in possible masking of weaker targets.

The third approach consists in estimating the sparse vector $\widehat{\mathbf{x}}$ via ℓ_1-norm minimization. The sparse estimate is then multiplied by the matrix containing time delayed versions of the transmitted waveform to obtain a non-sparse estimate of the Nyquist sampled received signal. This estimate is then further used for detection.

As also described in [147], it is clear that the one-shot CMF is not optimal due to the targets sidelobe interference. Particularly, the CMF is the estimate of the target response at the first iteration of most ℓ_1-minimization (and greedy) algorithms, before soft thresholding is applied. The difference between the CMF and ℓ_1-minimization is that, at each iteration, iterative algorithms produce an estimate of the active components in the vector $\mathbf{x}$, and use this information to suppress the interference (sidelobes). This will improve the detection of weaker components at the next iteration. Iterative algorithms are very similar to mismatched filters and they will, in general, outperform the CMF.

Furthermore, we argue that the constrained ℓ_0-norm minimization (of which ℓ_1-norm is a computationally feasible approximation), which can be formulated as

$$\text{minimize } \|\mathbf{y} - \mathbf{A}\mathbf{x}'\|_2^2 \ \text{ s.t. } \ \|\mathbf{x}'\|_0 \leq k, \tag{2.17}$$

represents the optimum, NP-hard, solution to the multi-hypothesis test for detecting multiple targets with unknown amplitudes and time delays, as described in [124], under the sparsity constraint of having less than k targets.

From the arguments above, for CS detection we propose to use an approach that is similar to the classical radar approach, with the difference that pulse compression is now performed via ℓ_1-norm minimization instead of using matched or mismatched filters. The pulse compressed and sparse signal estimate obtained from ℓ_1-norm minimization can be either considered as a vector of detections (where the soft thresholding at the last iteration acts as the detector) or can be used as input to a second detection stage. An example of this architecture for CS radar detection is shown in Figure 2.6. One of the questions to be answered is whether it is possible to optimize (P_d, P_{fa})

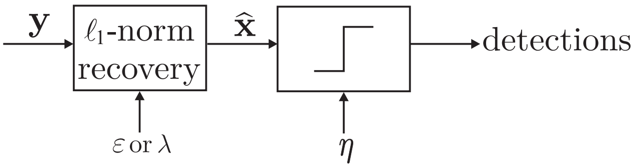


Figure 2.6: Example of CS radar detection architecture. The case $\eta = 0$ corresponds to considering the CS non-zero reconstructed samples as detections.

by properly choosing ε and setting $\eta = 0$, or if the performance can be improved by using a separate detection block (i.e., $\eta \neq 0$).

In addition, a prime parameter in almost every radar design is transmitted power, which may have to be reconsidered when undersampling schemes are introduced. This implies that undersampling based on CS principles will have a major impact on radar design, both in terms of performance and design methodology. Therefore, in order to make CS applicable to practical radar operation and design, the following questions must be addressed for CS-radar: how to optimized the detection probability against the false alarm rate; how to adaptively control the false alarm rate against unknown noise and clutter (CFAR), and how to design a CS-radar, in particular, what amount of undersampling is acceptable, and at what cost, in terms of power.

2.4.1 Receiver Operating Characteristic Curves

To understand the behavior of the CS detector in Figure 2.6, we estimated via simulations the ROC curves for several values of the parameters ε, SNR, δ, and ρ. For each set of parameters 10000 Monte Carlo simulations were performed and results are compared on the basis of an equal MF output SNR, as defined on page 35. For the case of a single target ($k = 1$), we set $N = 200$ and $n = 10$, 20, 30, 40, 50, and 66, corresponding to $\delta = [0.05, 0.1, 0.15, 0.2, 0.25, 0.33]$ and $\rho = [0.1, 0.05, 0.033, 0.025, 0.02, 0.0152]$. Furthermore, since IST and SPGℓ_1 would provide similar results under proper tuning of the parameters ε and λ, in the following figures we will report only the results obtained using the SPGℓ_1 algorithm.

In the following figures, the FAP is computed over all output samples except the target sample with 2 guard cells margin. The range of values of SNR is between -2 and 15 dB. Values outside this range were simulated as well [154], but not reported here, as they are of less interest, since detection is either impossible (low SNR) or almost certain (high SNR). The simulated target is a point target with constant amplitude. For each SNR, ROC curves were evaluated for set of values of ε proportional to σ and, for varying n, the ratio ε/σ was kept constant and equal to 0, 2, 4, 6 and 8.[15] In Fig-

---

[15] If the parameter ε is set above a certain maximum value, then the only feasible solution is the

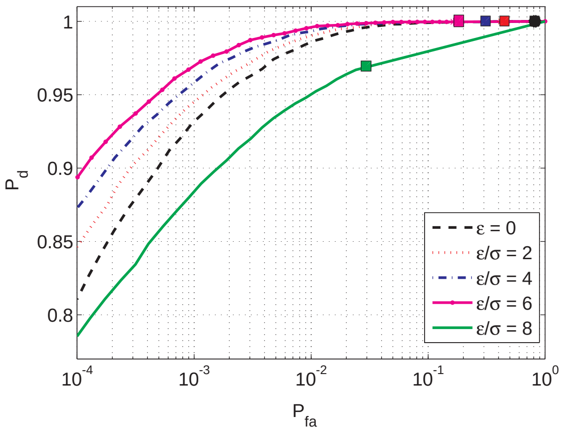


Figure 2.7: P_d versus P_{fa} (ROC) for SNR= 12dB and a single target. $\delta = 0.33$.

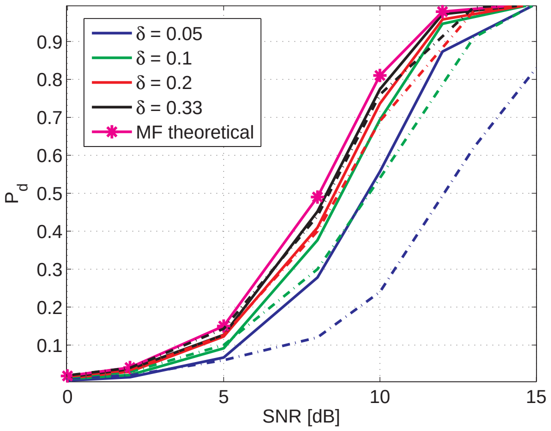


Figure 2.8: P_d versus SNR. FAP $= 10^{-3}$. Solid lines: CS plus detector. Dotted lines: CMF plus detector.

ure 2.7 an example of ROC curves is shown for $\delta = 0.33$ and SNR $=$ 12dB. To obtain the ROC curve for a fixed ε, a second detector is used after CS reconstruction, with varying detection threshold η. The case in which the CS reconstructed signal samples are considered as detections is a special case of the model in Figure 2.6 (i.e., $\eta = 0$).

---

trivial solution $\widehat{\mathbf{x}} = 0$, as the threshold used in the soft thresholding function is too high. In our experiments, we obtained almost always the zero solution when using $\varepsilon/\sigma = 10$.

We will refer to this detector as the ℓ_1-detector. For the ℓ_1-detector, for each value of ε a single pair (P_d, P_{fa}) is obtained. The performance of the ℓ_1-detector is shown in Figure 2.7 with a square, the color of which corresponds to a given ε, as shown in the legend. Thus, the ROC curve of the ℓ_1-detector can be obtained by joining the points obtained for different values of ε. From the simulated ROC plots it appears that using CS followed by a separate detector one can achieve better performance than in the $\eta = 0$ case, especially since tuning the second threshold η allows to reduce FAP while maintaining P_d relatively high. Moreover, notice that for a suboptimum value of the threshold ε, the performance degrade gracefully when a separate detector is used. This behavior suggests that, in an adaptive scheme, estimation errors on the noise level could be well tolerated using this type of architecture.

A comparison of CS with both the MF and the CMF (all followed by a separate detector) is shown in Figure 2.8. Here P_d is plotted versus SNR for a fixed $P_{fa} = 10^{-3}$ for several values of δ. From Figure 2.8 we observe that, for the single target case CS performs very close to the optimum, fully sampled MF. On the other hand, CS performs significantly better than CMF for small values of δ, because the effect of decimation in the CMF creates high target sidelobes, thus generating more false alarms at a given P_d, which becomes more apparent at high SNRs.

2.4.2 Grid discretization errors

The simulations reported in Figures 2.7 and 2.8 were performed under ideal conditions, in the sense that no sources of error other than noise were present. In practice, there is always an additional source of error, which is caused by the discretization of the grid. This error is due to the fact that, in reality, the target position never falls exactly on one of the discrete grid points. However, it can be shown by analysis and simulation that the effect of the grid mismatch is similar to additional noise. Furthermore, just as in conventional matched filtering, the detected peak will be reduced by the fact that the target falls between grid samples. It is common practice in any radar system design to include a straddling loss, and in fact such a loss occurs in CS as well as in conventional MF. The grid/target mismatch error has the effect of reducing the effective output SNR, therefore resulting in a reduced P_d for a given FAP. This phenomenon can be observed by comparing the ROC curve obtained from a target that falls on or off a grid point, in both cases with equal SNR. Such an example is shown in Figure 2.9, both for MF and for CS. In Figure 2.9, we set $\delta = 0.2$ and the grid mismatch error to $\Delta R/2$, which corresponds to the worst case scenario in which the target is exactly in between two grid points. In this figure, the target amplitude and noise variance are such that the MF output SNR is equal to 15 dB. However, the grid error results in a loss in detection probability, as the target received power is split between two grid points. The loss in detection performance can be converted into an

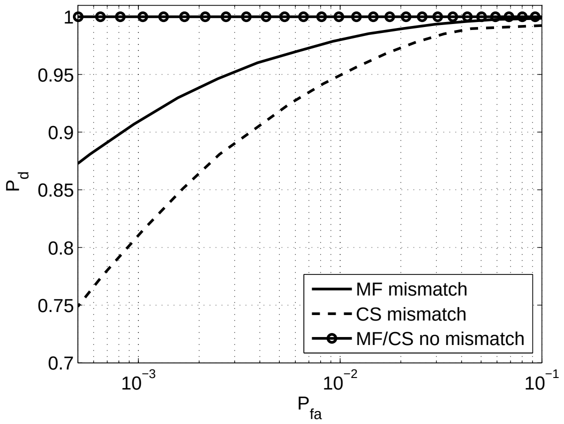


Figure 2.9: ROC with $\Delta R/2$ grid mismatch error. Ideal MF output SNR = 15 dB (without grid error). $\delta = 0.2$.

equivalent loss in output SNR, which amounts for this example to approximately 3 dB. A practical way to compute the straddling loss is to average the ROC curves obtained for different amounts of grid mismatch. An extensive discussion of grid mismatch is outside the scope of this work. A more detailed analysis of the discretization error can be found in [155] and references therein. In [59] and [60] the authors propose to oversample the reconstruction grid to reduce 'grid' noise and straddling losses. In [156], the authors propose to use a regularized Total Least Square (TLS) algorithm to improve the target localization.

2.5 Conclusions

In this chapter we introduced the radar detection problem in the context of CS and analyzed the properties of the signal recovered using ℓ_1-norm minimization. We devise two possible architectures for CS detection in a multiple, k-sparse target scenario. From simulation results (with known noise level) it appears that detection after CS reconstruction can closely approach classical detection performance, with fewer measurements. However selection of the threshold or regularization parameter that is needed for the ℓ_1-norm based recovery plays a significant role, and it appears that better performances in terms of ROC are achieved when the CS reconstruction is followed by a separate detector. Moreover, ideally CS does not produce target sidelobes in the way that CMF does, which improves the FAP of the CS detector with respect to CMF for a fixed threshold.

For the design and analysis of adaptive CS CFAR radar detectors we have identified a number of problems that need to be dealt with:

- unknown relation between the recovered signal and the regularization parameter of ℓ_1-norm recovery;
- unknown statistical properties of the recovered signal as a function of the CS parameters δ and ρ, and the input SNR;
- how to accurately and efficiently estimate the noise level from the CS measurements;
- how to set adaptively the recovery threshold;
- if a separate detector is used after CS recovery:
  - how to set the detector threshold as a function of the unknown and sparse recovery noise;
  - how to make the detector adaptive;
  - how to obtain CFAR property.

To deal with these issues, in the next chapter we will introduce the Complex Approximate Message Passing (CAMP) algorithm for ℓ_1-norm recovery. This algorithm, which is an extension of the original AMP algorithm [137, 157–160] to the case of signals in the complex domain, allows to statistically characterize the recovered signal. Using the properties of CAMP, we are able to derive closed form expressions for the detection and false alarm probabilities of the CAMP based CS detector. Also, using the statistical properties of the estimated signal, we propose an adaptive detector that can be combined with conventional CFAR processing. Our results and the novel architecture based on CAMP allow to use classical radar design tools even in the CS framework.

Chapter 3

Compressive Sensing Radar Detectors

In this chapter we focus on the problem introduced in Chapter 2 of target detection from a set of CS radar measurements corrupted by additive white Gaussian noise. We propose two novel architectures based on Complex Approximate Message Passing (CAMP) and compare their performance by means of receiver operating characteristic curves. Using asymptotic arguments and the properties of the CAMP algorithm, we characterize the statistics of the ℓ_1-norm reconstruction error and derive closed form expressions for both the detection and false alarm probabilities of both schemes. Although the theoretical results on CAMP assume sensing matrices with i.i.d. random entries, our simulations show that the general trend of our findings also applies to partial Fourier sensing matrices. Furthermore, we demonstrate that, of the two architectures, the best one consists of a reconstruction stage based on CAMP followed by a detector. This architecture, which outperforms the ℓ_1-detector in the ideal case of known background noise, can also be made fully adaptive by combining it with a conventional CFAR processor. Finally, using the state evolution framework of CAMP, we also derive SNR plots in which the reconstruction SNR, which is input to the detection stage, is plotted against the undersampling factor (δ) and the relative signal sparsity (ρ). Given a maximum number of expected targets in the scene, such graphs can be used to evaluate how transmitted power can be traded for undersampling and thus serve as a guide for the CS radar designer.[1]

---

[1] This chapter is based on articles [**J1**], and [**J2**] (a list of the author's publications is included at the end of this dissertation, p. 183.)

3.1 Complex Approximate Message Passing (CAMP) algorithm

Complex Approximate Message Passing [37] is an iterative algorithm for solving (1.9). CAMP is an extension of the original Approximate Message Passing (AMP) algorithm, first proposed in [137] for real signals, to the case of signals and measurements in the complex domain. The AMP algorithm and its properties have been thoroughly investigated in [157–162] for the case of both noise-free and noisy CS measurements. However, as we already mentioned in the previous chapter, in radar it is most common to work with signals in the complex domain. Therefore, in the remainder of this chapter we will concentrate exclusively on CAMP and, in this section, we briefly review some of its main features that will be used in the derivation of our novel CS radar detection schemes.[2] As we will see in the next sections, these properties of CAMP enable us to achieve the following objectives:

- characterize the distribution of the noise after ℓ_1-norm minimization;
- establish the relation between the regularization parameter λ of ℓ_1-norm recovery and the quality, in terms of SNR, of the recovered solution;
- adaptively set the regularization parameter λ in a way that optimizes the recovery SNR;
- design a fully adaptive CS radar detector that can be combined with classical CFAR processing;
- determine the tradeoffs between amount of undersampling, signal sparsity and detection performance.

We first review the iterations of the CAMP algorithm, that is given in Algorithm I. In Algorithm I, $\langle \cdot \rangle$ denotes the average of a vector, η^I and η^R are the imaginary and real parts of the complex soft thresholding function, $\frac{\partial \eta^R}{\partial x_R}$ is the partial derivative of η^R with respect to the real part of the input, $\frac{\partial \eta^I}{\partial x_I}$ is the partial derivative of η^I with respect to the imaginary part of the input, and maxiter is the (user specified) maximum number of iterations. Furthermore, notice that in CAMP the soft thresholding function is applied with threshold parameter $\tau\sigma_t$, where σ_t is the current noise standard deviation and τ is a (fixed) user specified threshold. We will see in the next subsection how the parameter τ of CAMP relates to the regularization parameter λ in (1.9). Also, recall that the undersampling factor is defined as $\delta = \frac{n}{N}$, i.e., the

---

[2] The derivation of the CAMP algorithm from message passing algorithms can be found in [37]. In this chapter, we will mainly review the properties of CAMP that are necessary to solve the detection problem.

Algorithm I: Ideal CAMP Algorithm

```
Input: $\mathbf{y}$, $\mathbf{A}$, $\tau$, $\mathbf{x}$
Initialization $\widehat{\mathbf{x}}^0 = 0$, $\mathbf{z}^0 = \mathbf{y}$
for $t = 1 : \text{maxiter}$
    $\widetilde{\mathbf{x}}^t = \mathbf{A}^\dagger \mathbf{z}^{t-1} + \widehat{\mathbf{x}}^{t-1}$
    $\sigma_t = \text{std}(\widetilde{\mathbf{x}}^t - \mathbf{x})$
    $\mathbf{z}^t = \mathbf{y} - \mathbf{A}\widehat{\mathbf{x}}^{t-1} + \mathbf{z}^{t-1} \frac{1}{2\delta} \big( \langle \frac{\partial \eta^R}{\partial x_R}(\widetilde{\mathbf{x}}^t; \tau\sigma_t) \rangle + \langle \frac{\partial \eta^I}{\partial x_I}(\widetilde{\mathbf{x}}^t; \tau\sigma_t) \rangle \big)$
    $\widehat{\mathbf{x}}^t = \eta(\widetilde{\mathbf{x}}^t; \tau\sigma_t)$
end
Output: $\widetilde{\mathbf{x}}, \widehat{\mathbf{x}}, \sigma_*$
```

number of CS measurements divided by the number of samples to be recovered, and the relative signal sparsity is defined as $\rho = \frac{k}{n}$, i.e., the number of non-zero coefficients in $\mathbf{x}$ divided by the number of CS measurements.

We now explain each variable in the CAMP algorithm:

(i) $\widehat{\mathbf{x}}^t$ is an estimate of $\mathbf{x}$ at iteration t. If the parameter τ is tuned properly, $\widehat{\mathbf{x}}^t \to \widehat{\mathbf{x}}(\lambda)$ as $t \to \infty$. The tuning of τ in terms of λ is given later in (3.3) and will be explained in detail in Section 3.1.1.

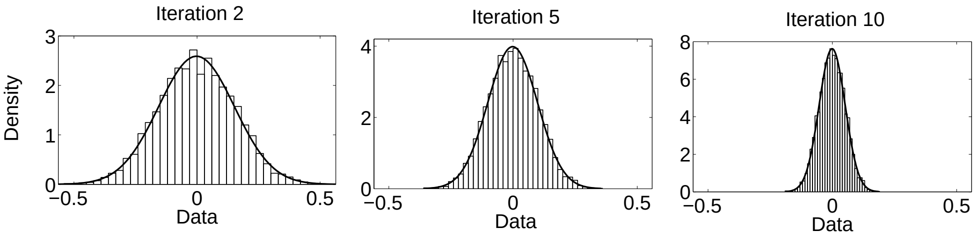
(ii) $\widetilde{\mathbf{x}}^t$ is a non-sparse, noisy estimate of $\mathbf{x}$. Define the 'noise' vector $\mathbf{w}^t = \widetilde{\mathbf{x}}^t - \mathbf{x}$ at iteration t of CAMP. The histogram of $\mathbf{w}^t$, which is shown in Figure 3.1, suggests that the empirical distribution of $\mathbf{w}^t$ is 'close' to a zero mean Gaussian probability density function. We will more rigorously discuss the Gaussianity of $\mathbf{w}^t$ in Section 3.1.1.

(iii) σ_t is the standard deviation of $\mathbf{w}^t$. Furthermore, we use the notation $\sigma_* \triangleq \lim_{t\to\infty} \sigma_t$.

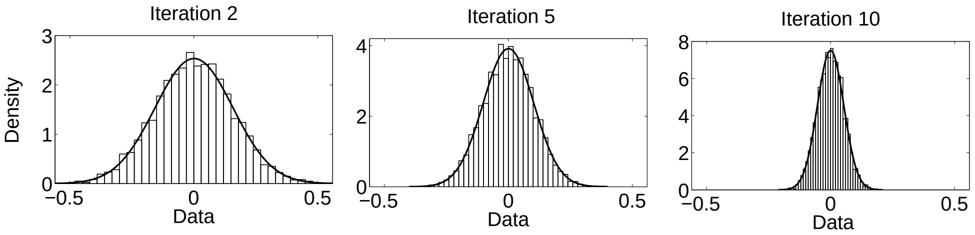
In words, CAMP first finds a noisy estimate of the signal, that we call $\widetilde{\mathbf{x}}^t$. Since this estimate is not sparse, the soft thresholding function is applied to obtain a sparse estimate $\widehat{\mathbf{x}}^t$. Here for the clarity of exposition we have assumed that the algorithm uses the sought $\mathbf{x}$ for computing the noise standard deviation σ_t. Hence, we refer to this algorithm as *Ideal CAMP*. In Section 3.3 we propose a more practical scheme where σ_t is replaced with an estimate.

3.1.1 State evolution: A framework for the analysis of CAMP

There are three important questions that we need to answer before we proceed to show how CAMP can be used in the analysis and design of CS radar detectors. (i) Under



(a) Histogram of the real part of $\mathbf{w}^t$.



(b) Histogram of the imaginary part of $\mathbf{w}^t$.

Figure 3.1: Histograms (bars) of the real and imaginary parts of the CAMP residual noise signal $\mathbf{w}^t$ at different iterations of Ideal CAMP. A Gaussian probability density function is fitted to the histograms (black, solid line). In these plots, $N = 4000$, $\delta = 0.6$, and $\sigma = 10^{-3}$.

what conditions are the above heuristics accurate? (ii) Can we use the properties of CAMP to predict its performance theoretically? (iii) What is the formal connection between CAMP and the LASSO problem defined in (1.9)?

The first question is accurately discussed in [37, 159, 161, 162]. It has been proved that in the asymptotic setting $N \to \infty$, while δ and ρ are fixed, the above heuristics are correct. Consider the following definition from [162].

Definition 3.1.1. *For a given $(\delta, \rho) \in [0, 1]^2$, a sequence of instances $\{\mathbf{x}(N), \mathbf{A}(N), \mathbf{n}(N)\}$ is called a converging sequence if the following conditions hold:*

- *The empirical distribution of $\mathbf{x}(N) \in \mathbb{R}^N$ converges weakly to a probability measure p_X with bounded second moment as $N \to \infty$.*
- *The empirical distribution of $\mathbf{n} \in \mathbb{R}^n$ $(n = \delta N)$ converges weakly to a probability measure p_n with bounded second moment as $N \to \infty$.*
- *The elements of $\mathbf{A}(N) \in \mathbb{R}^{n \times N}$ are i.i.d. drawn from a Gaussian distribution.*

Theorem 3.1.2. [162] *Let $\{\mathbf{x}(N), \mathbf{A}(N), \mathbf{n}(N)\}$ be a converging sequence, and let $\widetilde{\mathbf{x}}^t(N)$ be the estimate provided by the CAMP algorithm. The empirical law of $\mathbf{w}^t(N) =$*

$\widetilde{\mathbf{x}}^t(N) - \mathbf{x}(N)$ *converges to a zero-mean Gaussian distribution almost surely as* $N \to \infty$.

While the above result has been proved for Gaussian measurement matrices in the asymptotic setting, simulation results confirm that it is still accurate even for medium problem sizes with $N \sim 200$ and different classes of sensing matrices.[3] We explore this claim for the partial Fourier matrices that are of particular interest in radar applications using stepped frequency waveforms in Section 3.5.1.

Theorem 3.1.2 enables us to answer the second question as well. Using the Gaussianity of the noise in the asymptotic regime, one can predict the performance of CAMP through what is called the "state evolution" (SE). SE tracks the evolution of the standard deviation of the noise σ_t across iterations. Let the marginal distribution of $\mathbf{x}$ converge to p_X, and σ_t denotes the standard deviation of $\mathbf{w}^t$. In the asymptotic setting, the value of the standard deviation at time $t+1$ is calculated from σ_t according to the following equation:

$$\sigma_{t+1}^2 = \Psi(\sigma_t), \tag{3.1}$$

where

$$\Psi(\sigma_t) = \sigma^2 + \frac{1}{\delta}\mathbb{E}\left(|\eta(X + \sigma_t Z; \tau\sigma_t) - X|^2\right), \tag{3.2}$$

$Z \sim \mathcal{CN}(0,1)$, σ^2 is the input noise variance (as defined on page 22) and the expectation is with respect to the two independent random variables $X \sim p_X$ and Z. In fact, σ_t being the standard deviation of the noise at iteration t, $\mathbb{E}\left(|\eta(X + \sigma_t Z; \tau\sigma_t) - X|^2\right)$ equals the MSE of the estimate $\widehat{\mathbf{x}}^t$ after applying the soft thresholding function. For this reason, the expectation in (3.2) is also referred in literature as the risk of the soft thresholding function. It has been proved in [37] that the function Ψ is concave, and therefore the iteration (3.1), (3.2) has at most one stable fixed point σ_*^2. Also, CAMP converges to this fixed point exponentially fast (linear convergence according to optimization literature). Appendix A provides an example of how to evaluate the Ψ function for a given distribution p_X. The following two consequences of (3.1), (3.2) are particularly useful for the radar application:

(i) The SE framework establishes the input/output relation for CAMP. Particularly, the output noise power σ_*^2 is the sum of the actual system noise (σ^2) plus a noise-like component caused by the reconstruction itself ($\frac{1}{\delta}$ MSE). Consequently, for a given σ^2, minimizing the reconstruction error also minimizes σ_*^2, therefore maximizing the CAMP reconstruction SNR.[4]

---

[3] Empirical studies have already confirmed that this theoretical prediction holds for other sensing matrices with i.i.d. elements other than Gaussian, [137, 159, 162].

[4] For a given target received power a^2, the CAMP input and output SNR are defined here re-

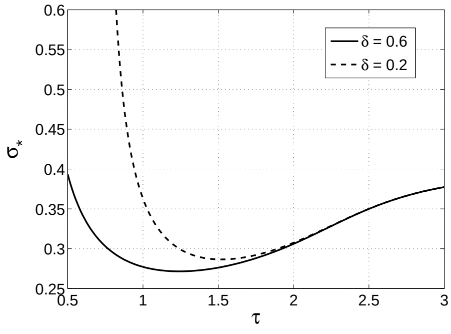


Figure 3.2: Fixed point σ_* versus threshold τ for Ideal CAMP with $\sigma = 0.23$, compression factors $\delta = 0.2$ (dashed line) and $\delta = 0.6$ (solid line). $\rho = 0.1$ and the non zero entries of the vector $\mathbf{x}$ have all amplitudes equal to 1 and phase uniformly distributed between $-\pi$ and π. The sensing matrix $\mathbf{A}$ has i.i.d. Gaussian entries. These curves are obtained using the formulae derived in Appendix A.

(ii) The fixed point σ_* depends on τ as well as on δ, p_X, and σ. Figure 3.2 exhibits the dependence of σ_* on τ for two distinct values of δ and for a fixed problem instance, i.e., fixed p_X and σ. As it is clear from the figure, there is a value of τ, say τ_o, for which σ_* is minimized. Moreover, as the number of measurements decreases, both the optimal threshold τ_o and the corresponding output noise standard deviation increase.

Finally, to answer the third question we observe that there is a nice connection between the CAMP and LASSO algorithms. According to [37], if τ is chosen to satisfy

$$\lambda \triangleq \tau\sigma_* \left(1 - \frac{1}{2\delta}\mathbb{E}\left(\frac{\partial \eta^R}{\partial x_R}(X + \sigma_* Z; \tau\sigma_*) + \frac{\partial \eta^I}{\partial x_I}(X + \sigma_* Z; \tau\sigma_*)\right)\right), \tag{3.3}$$

then in the asymptotic setting CAMP with threshold τ solves the LASSO problem (1.9) with parameter λ.

3.2 CS target detection using CAMP

Using the properties of CAMP explained previously, we now propose two CS detection schemes and analyze their performance as measured by their ROC curves. Let k be the number of targets, i.e., the number of non-zero coefficients in $\mathbf{x}$, and define G as

---

spectively as $\text{SNR}_{in} = a^2/(n\sigma^2)$ and $\text{SNR}_{out} = a^2/\sigma_*^2$. Note that the output (reconstruction) SNR depends on, amongst other quantities, the CAMP threshold τ via σ_*. A more rigorous definition of SNR is given in Section 3.4.

the distribution of the non-zero elements of $\mathbf{x}$. In this section we assume that k, G, and σ are known. In Section 3.3 we will investigate the more realistic scenario where these parameters are not known a priori and describe how to implement the proposed architectures in this case.

The two architectures we consider are displayed in Figure 3.3. Since any sparse recovery algorithm is intrinsically a detection scheme, Architecture 1 seems a natural choice for a CS radar detector. In this architecture, the measurements $\mathbf{y}$ are input to a recovery algorithm (here CAMP or equivalently LASSO). This algorithm returns a sparse vector $\widehat{\mathbf{x}}$ with the non-zero values being detections. Clearly, the threshold parameter τ in CAMP (or equivalently the regularization parameter λ in LASSO) controls the false alarm probability α (or P_{fa}) and detection probability (P_d) of the algorithm.

Proposition 3.2.1. *Consider the CAMP iteration with threshold* $\tau_\alpha = \sqrt{-\ln \alpha}$. *If* $\mathbf{A}(N), \mathbf{x}(N), \mathbf{w}(N)$ *is a converging sequence and* $\widehat{x}(N)$ *is the fixed point of CAMP, then*

$$\lim_{N\to\infty} \frac{1}{N-k} \sum_{i=1}^{N} 1_{\{\widehat{x}_i(N)\neq 0, x_i(N)=0\}} = \alpha. \tag{3.4}$$

almost surely. Also, τ_α *is the only value of* τ *for which* (3.4) *holds.*

Proof. Define $z \sim \mathcal{CN}(0,1)$. According to [161] we know that

$$\lim_{N\to\infty} \frac{1}{N-k} \sum_{i=1}^{N} 1_{\{\widehat{x}_i(N)\neq 0, x_i(N)=0\}} = \mathbb{P}(|\sigma_* z| > \tau_\alpha \sigma_*) = \mathrm{e}^{-\tau_\alpha^2}.$$

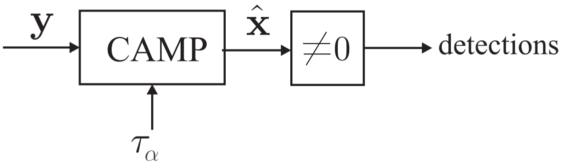
□

Note that the connection between τ and the detection probability is not clear yet. We will discuss this issue later.

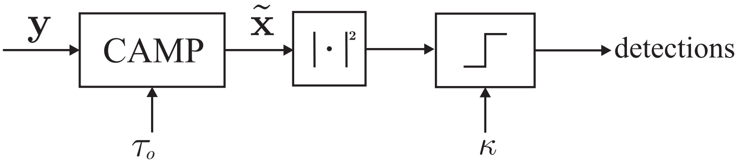
The second architecture is inspired by the standard, non-CS radar approach. Most commonly, radar detectors comprise two stages: an estimation stage, where a noisy estimate of the signal is computed, followed by a detection stage. Usually, a matched filter is used to obtain a noisy estimate of the signal with optimum SNR.[5] Then this noisy estimate is fed to a detection block that controls the False Alarm Probability (P_{fa} or FAP). Inspired by this philosophy and by the properties of CAMP, we introduce Architecture 2. We first use CAMP to obtain a noisy, non-sparse estimate

---

[5] As explained in Chapter 2, the MF is optimal (in terms of SNR) only for the case of a single target in white Gaussian noise. For the case of multiple targets, the optimality is only satisfied if the matrix $\mathbf{A}$ is orthogonal. Nevertheless, even for non orthogonal $\mathbf{A}$ and in a multiple targets scenario, MF (with apodization) is one of the most popular approaches.



(a) Architecture 1



(b) Architecture 2

Figure 3.3: Block diagrams of the proposed architectures for CS-radar detection.

of the signal $\widetilde{\mathbf{x}} = \mathbf{x} + \mathbf{w}$. Similar to classical estimation procedures, the goal of the first stage is to minimize σ_* by choosing the optimal CAMP threshold $\tau = \tau_o$. Once σ_* is minimized, or equivalently the output SNR is maximized, the noisy signal $\widetilde{\mathbf{x}}$ can be fed to a detection block with fixed threshold $\kappa = \sigma_* \sqrt{-\ln(\alpha)}$, which is used to control the false alarm rate. From the Gaussianity of $\mathbf{w}$, it is clear that in the asymptotic setting this choice of κ results in the false alarm probability α as derived in Proposition 3.2.1.

As will be clarified later, Architecture 2 is much more appropriate for practical radar applications, since all of the parameters can be optimized and estimated efficiently even without prior knowledge of k, G, and σ. Furthermore, from a detection perspective, Architecture 2 outperforms Architecture 1. Suppose that the Gaussianity of $\mathbf{w}^t$ holds. Let τ_o be the optimal value of τ that leads to the minimum σ_* and that this optimal value is unique and can be computed using (3.2). Under these assumptions, which provably hold under the conditions specified in Theorem 3.1.2, the following theorem can be derived.

Theorem 3.2.2. *Set the probability of false alarm to α for both Architecture 1 that uses τ_α and Architecture 2 that uses τ_o in CAMP. If $P_{d,1}$ and $P_{d,2}$ are the detection probabilities of the two schemes, then*

$$P_{d,1} \leq P_{d,2}.$$

Furthermore the equality is satisfied at only one specific value $\alpha^\star = e^{-\tau_o^2}$.

The above theorem is proved in Appendix B by comparing the detection and false alarm probabilities of the two architectures. However, since in Architecture 2

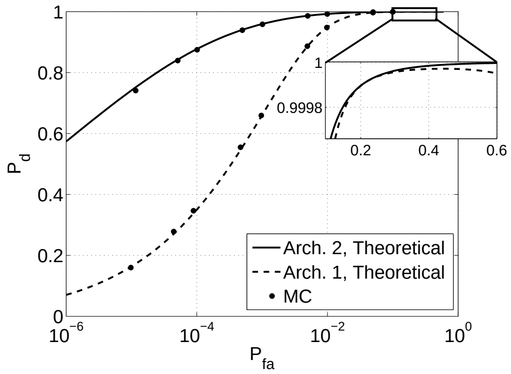


Figure 3.4: Comparison of the ROC curves for Architectures 1 and 2 at $\delta = 0.6$, $\rho = 0.1$ and $\sigma^2 = 0.05$. The distribution G of the non-zero coefficients of $\mathbf{x}$ is chosen such that all non zero components have the same amplitude (equal to 1) and phase uniformly distributed between $-\pi$ and π. The solid and dashed lines represent the theoretical predictions based on the SE equation. The dots are the results of MC simulations using Ideal CAMP. The sensing matrix for MC simulations had i.i.d. Gaussian entries.

the CAMP threshold is designed to minimize the output noise variance, intuitively it is clear that, for the same P_{fa}, any other choice of τ will lead to a higher σ_*, i.e., a lower SNR, and therefore a lower P_d. Example ROC curves for Architectures 1 and 2 are shown in Figure 3.4. The solid and dashed lines are obtained using the analytical equation derived in Appendix A. The theoretical ROC curves are verified by Monte Carlo simulations (dots). In the simulations we run the Ideal CAMP algorithm given in Algorithm I for several values of P_{fa} ranging from 10^{-1} to 10^{-5}. In Figure 3.4 we can also observe an interesting characteristic of Architecture 1 in the region around $P_{fa} = 0.4$ (the zoomed area); the probability of detection, $P_{d,1}$, decreases as P_{fa} increases above this value. To understand why, recall that in Architecture 1 the CAMP threshold τ_α varies with P_{fa}. Therefore σ_* (and hence the CAMP reconstruction SNR) also changes with P_{fa}, and it is not constant along the ROC curve. This explains why $P_{d,1}$ reaches its maximum at around $P_{fa} = 0.4$ and then decreases again as the P_{fa} goes to 1. Instead, in Architecture 2, σ_* is fixed to its minimum along the ROC curve, and therefore $P_{d,2}$ increases with increasing FAP. From this plot it is also possible to observe that there exists only one value of FAP ($\alpha^\star = 0.22$) where $P_{d,1} = P_{d,2}$.[6]

---

[6] The difference in performance between the two architectures depends significantly on the values of (δ, ρ, σ) and on the sensing matrix in use, as explained later in Section 3.5.

3.3 Adaptive CAMP Algorithm

So far, we have assumed to know exactly ρ, σ, and G (to derive the theoretical fixed point) and $\mathbf{x}$ (to run Ideal CAMP). However, in practical radar systems such information is not available, and both the CAMP and detector parameters must be estimated from the CS measurements $\mathbf{y}$. In this section we demonstrate how these issues can be handled in practice and propose a fully adaptive scheme, i.e., one that does not require any prior information and that adapts to changes in noise level. There are three main issues that have to be settled. (i) How to estimate σ_t without knowing $\mathbf{x}$. (ii) How to compute the optimal value τ_o for Architecture 2, efficiently and accurately, without the SE. (iii) How to replace the fixed threshold κ for the detector in Architecture 2 with an adaptive threshold to maintain a CFAR.

The first question can be answered in several different ways. For instance, in the result presented here we use the median to estimate the standard deviation via

$$\widehat{\sigma}_t = \sqrt{\frac{1}{\ln 2}}\,\mathrm{median}(|\widetilde{\mathbf{x}}^t|). \tag{3.5}$$

This estimator is unbiased if $\mathbf{x} = 0$. However, in the presence of targets, i.e., when $\mathbf{x} \neq 0$, (3.5) is a biased estimator. The main advantage of this scheme is its robustness to high SNRs. To see this, we consider the performance of the median estimation in the asymptotic setting. As mentioned above, in the asymptotic setting at each iteration of CAMP we have $\widetilde{\mathbf{x}}^t = \mathbf{x} + \mathbf{w}^t$. Assume that the elements of $\mathbf{x}$ are distributed as $x_i \overset{\text{i.i.d.}}{\sim} \epsilon G(x) + (1-\epsilon)\delta(x)$ with $\epsilon = \delta\rho \ll 1$, $\delta(x)$ is the Dirac delta function and $w_i^t \sim \mathcal{CN}(0, \sigma_*^2)$. The goal is to estimate the median μ^* of $|\mathbf{w}^t|$. However, since we only have access to $\widetilde{\mathbf{x}}$, we estimate the median of $|\mathbf{w}^t|$ as the $\widehat{\mu}$ that satisfies $\mathbb{P}(|\widetilde{x}_i^t| > \widehat{\mu}) = \frac{1}{2}$. The following theorem provides an upper bound on the deviation of the estimated median from the true median.

Proposition 3.3.1. *The error of the estimated median is bounded above by*

$$\frac{|\widehat{\mu} - \mu^*|}{\sigma_*} \leq \frac{|\ln(1-\epsilon)|}{2\sqrt{\ln 2}}. \tag{3.6}$$

Proof. To quantify the error of the median estimator as a function of δ and ρ, consider the random variable x with pdf given by $f_x(x) = (1-\epsilon)\delta(x) + \epsilon\delta(x-a)$ and $Z \sim \mathcal{CN}(0, \sigma^2)$. Define μ^* as the median of the absolute value of the random variable Z, i.e., $P(|Z| \leq \mu^*) = \frac{1}{2}$. Since $|Z| \sim \mathrm{Rayleigh}(\sigma^2/2)$, we obtain $\mu^* = \sigma\sqrt{\log(2)}$.

Also, define μ as the median of the absolute value of the random variable $X + Z$, i.e.,

$$\begin{aligned} P(|X+Z| \leq \mu) &= \epsilon P(|a+Z| \leq \mu) + (1-\epsilon)P(|Z| \leq \mu) \\ &= \epsilon P(|a+Z| \leq \mu) + (1-\epsilon)(1 - e^{-\mu^2/\sigma^2}) = \frac{1}{2}. \end{aligned}$$

Since $0 \leq \mathbb{P}(|a+Z| \leq \mu) \leq 1$, using the upper bound we obtain

$$(1-\epsilon)(1-e^{-\mu^2/\sigma^2}) \leq \frac{1}{2}-\epsilon$$

and therefore

$$\mu \leq \sigma\sqrt{\log\big(2(1-\epsilon)\big)}.$$

Hence the normalized bias introduced by the median estimator in the presence of a target is bounded by

$$\frac{|\mu-\mu^*|}{\sigma} \leq \left|\sqrt{\log\big(2(1-\epsilon)\big)}-\sqrt{\log(2)}\right|.$$

Furthermore, using the inequality $\log\big(2(1-\epsilon)\big) < \log(2)$, the bound in (3.6) follows. $\square$

It is important to note several interesting properties of the upper bound (3.6). First, the distribution G of the non zero components of $\mathbf{x}$ does not play any role. Second, for small values of ϵ, i.e., in very sparse situations, $|\ln(1-\epsilon)| \approx \epsilon$ and, hence, the error is proportional to the sparsity level. Thus, when the noise level and $\mathbf{x}$ are unknown, a practical implementation of Algorithm I is obtained by replacing σ_t with the estimate in (3.5). We will refer to this algorithm simply as CAMP or *Median CAMP*.

The estimate of σ_t provides an approach to answer the second above question as well, namely how to estimate, in Architecture 2, the optimum CAMP threshold $\widehat{\tau}_o$ that minimizes $\widehat{\sigma}_*^2$. Suppose that we know or can estimate $\tau_{\max}$ such that $\tau_o < \tau_{\max}$. Given a step size δ_τ, we define a sequence of thresholds $\tau = \{\tau_\ell\}_{\ell=1}^L$ such that $\tau_1 = \tau_{\max}$ and $\tau_\ell = \tau_{\ell-1} - \delta_\tau$. Starting from $\tau_{\max}$, at each new iteration ℓ CAMP is initialized with $\widehat{\mathbf{x}}_\ell^0 = \widehat{\mathbf{x}}_{\ell-1}$ and $\mathbf{z}_\ell^0 = \mathbf{z}_{\ell-1}$. Using the solution of CAMP at the previous iteration $\ell-1$ as initial value for $\tau = \tau_\ell$, CAMP needs only a few iterations to converge to the solution, and therefore the entire process is very fast. After L iterations, we have a matrix of solutions $\widehat{\mathbf{X}} = [\widehat{\mathbf{x}}_1, \widehat{\mathbf{x}}_2, \cdots, \widehat{\mathbf{x}}_L]$ of size $N \times L$, where each column contains the CAMP solution for a given τ_ℓ. Also, we have L estimates $\{\widehat{\sigma}_*^\ell\}_{\ell=1}^L$. The optimum estimated threshold $\widehat{\tau}_o$ is chosen as the one that minimizes the estimated CAMP output noise variance $\widehat{\sigma}_*^2$. It is clear that δ_τ specifies the trade-off between computational complexity and the accuracy of the algorithm in estimating $\widehat{\tau}_o$. Decreasing δ_τ increases the number of points L needed to span the same τ search region, but it also results in a more accurate estimate of $\widehat{\tau}_o$.

We now explain how to set $\tau_{\max}$. At the first iteration ($\ell = 1$ and $t = 1$) the CAMP algorithm is initialized with $\widehat{\mathbf{x}}^0 = 0$ and $\mathbf{z}^0 = \mathbf{y}$, and we have $\widetilde{\mathbf{x}} = \mathbf{A}^H\mathbf{y}$, where $\mathbf{A}^H$ is the Hermitian of the matrix $\mathbf{A}$. Suppose that $\widehat{\sigma}_0$ is estimated from this vector. Consider now the LASSO problem in (1.9). It is well known that for

Table 3.1: CAMP based algorithms.

Algorithm	**Inputs**	**Outputs**
Ideal CAMP	$\mathbf{A}, \mathbf{y}, \mathbf{x}, \sigma, \tau$	$\widehat{\mathbf{x}}, \widetilde{\mathbf{x}}, \sigma_*$
Median CAMP	$\mathbf{A}, \mathbf{y}, \tau$	$\widehat{\mathbf{x}}, \widetilde{\mathbf{x}}, \widehat{\sigma}_*$
Adaptive CAMP	$\mathbf{A}, \mathbf{y}$	$\widehat{\mathbf{x}}, \widetilde{\mathbf{x}}, \widehat{\sigma}_*, \widehat{\tau}_o$

$\lambda > \lambda_{\max} = \|\mathbf{A}^H\mathbf{y}\|_\infty$ the only solution is the zero solution. Using the calibration equation (3.3) with $\lambda = \lambda_{\max}$ and $\sigma_* = \widehat{\sigma}_0$, we can compute an estimate of $\widehat{\tau}_{\max}$. We will refer to this algorithm as *Adaptive CAMP*, since both the noise variance $\widehat{\sigma}_t$ and the threshold $\widehat{\tau}_o$ are adaptively estimated inside the algorithm itself, and the only input variables are $\mathbf{y}$ and $\mathbf{A}$.

To clarify the differences between Ideal, Median, and Adaptive CAMP, Table 3.1 shows the input and output variables for each of the three algorithms. Please recall that Ideal CAMP is not an algorithm that can be used in practice, as it requires knowledge of the true vector $\mathbf{x}$.

3.3.1 Adaptive CAMP CFAR radar detector

It remains now to establish how to replace the fixed threshold κ with an adaptive one for Architecture 2. In Appendix B we show how to set the fixed threshold κ to achieve the desired FAP when the noise variance σ^2 is known. In practice, however, the noise statistics are not known in advance. As explained in Section 2.2.3 of Chapter 2, in classical radar detectors a CFAR processor is commonly employed to estimate the unknown background noise and clutter level. However, one has to know the relation between the CFAR threshold multiplier and P_{fa} so that β can be adjusted to maintain P_{fa} constant during the observation time. Hence, the noise distribution should be known to design a CFAR scheme. If the CAMP estimate $\widehat{\mathbf{x}}$ were input to a CFAR processor, then estimation of the noise characteristics would be very difficult, since many samples in $\widehat{\mathbf{x}}$ are identically zero. Instead, the signal $\widetilde{\mathbf{x}}$ contains all non-zero samples, and it is modeled as the sum of targets plus Gaussian noise; this can be directly input to a conventional CFAR processor. A block diagram of the Adaptive CAMP CFAR detector based on Architecture 2 is shown in Figure 3.5.

For the properties of CAMP, in Architecture 2 replacing the fixed threshold detector with a CFAR detector provides similar results as in classical CFAR without CS. In other words, once we have access to the signal $\widetilde{\mathbf{x}}$, we can design the detection stage using classical equations of P_d and P_{fa} for the selected detector. All we need to determine is the input/output SNR relations of CAMP, so that we can use the output SNR in the detector equation for the prediction of P_d. As can be seen from the SE equation, for a given input SNR, the output SNR of CAMP will depend on

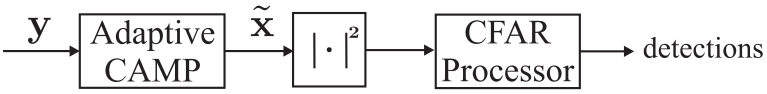


Figure 3.5: Block diagram of the Adaptive CAMP CFAR detector.

the parameters δ, p_X, and τ. Hence, if we know how to compute or estimate the output SNR of CAMP for a given set of parameters, then we can resort to classical radar tools to predict the performance of the CAMP based detector and therefore we can treat the recovery stage and the detection stage independently.

3.4 Defining the SNR for CS CAMP

Before proceeding to the simulation results, we provide a definition of SNR for the CAMP-based CS radar system. This definition will be useful not only for understanding the performance of the proposed detectors, but also to compare the novel CS based architectures to more classical ones for which the performance is known. We observed in (3.2) that the variance σ_*^2 of the noise present in the CAMP estimate depends on the 5-tuple $(\delta, \rho, p_X, \tau, \sigma)$. In fact, even for fixed p_X, σ and for a specific Architecture (that fixes τ), σ_*^2, and therefore the (P_d, P_{fa}) curves, can vary significantly with δ and ρ. This is uncommon in classic radar systems, where the SNR after the MF is uniquely determined for a given input SNR and integration time. Therefore it is important to relate the performance of a given CS scheme to that of MF-based classical systems.

Let $\widetilde{\mathbf{A}} \in \mathbb{C}^{N \times N}$ be the measurement matrix for the case $\delta = 1$, i.e., no undersampling. A conventional radar processor feeds the measurement vector $\mathbf{y} = \widetilde{\mathbf{A}}\mathbf{x} + \sigma\mathbf{z}$, $\mathbf{z} \sim \mathcal{CN}_N(0, \mathbf{I})$, to a MF to obtain a noisy estimate $\widehat{\mathbf{x}}_{MF} = \mathbf{x} + \sigma\mathbf{z}$ of the target received signal with optimum SNR.[7] CAMP enables us to define SNR in a similar way. Once CAMP has converged we have access to both the sparse estimate $\widehat{\mathbf{x}}$ and its noisy version $\widetilde{\mathbf{x}}$. As stated before, even in medium size problems $\widetilde{\mathbf{x}}$ can be accurately approximated by the sum of the true target vector plus white Gaussian noise with variance σ_*^2, i.e. $\widetilde{\mathbf{x}} = \mathbf{x} + \sigma_*\mathbf{z}$. Indicating with a^2 the received power from a target at bin i, we define the SNR at the input (SNR$_{in}$) and output (SNR) of the MF and

---

[7] In a multiple target scenario, the MF SNR is optimum and independent of the number of targets as long as each target is exactly on a grid point and the matrix $\widetilde{\mathbf{A}}$ is orthogonal. We assume these (ideal) conditions are satisfied when computing the MF SNR.

CAMP Architecture 2 respectively as

$$\begin{aligned}\mathrm{SNR}_{in,MF} &= \frac{a^2}{N\sigma^2}, \qquad \mathrm{SNR}_{in,CS} = \frac{a^2}{n\sigma^2},\\ \mathrm{SNR}_{MF} &= \frac{a^2}{\sigma^2}, \qquad \mathrm{SNR}_{CS} = \frac{a^2}{\sigma_*^2}.\end{aligned} \tag{3.7}$$

In the remainder of our discussion we assume that the total transmitted (and received) power is independent of δ.[8] In the case of partial Fourier sensing matrix this is achieved in practice by dividing the total available transmit power over the subset of transmitted frequencies. With this assumption, dividing both sides of (3.2) by a^2, we can derive the CAMP output SNR as a function of the equivalent MF output SNR, i.e., the SNR that we would obtain without compression using a MF. The constraint of keeping the received target power equal for different amounts of undersampling enables us to evaluate the changes in SNR_{CS} due to reconstruction and not due to a reduction in total signal power. In other words, if we normalize (3.2) by a^2, then we keep the first term on the right hand side fixed (which is equal to SNR_{MF}) and observe the change in SNR_{CS} produced by the reconstruction error term ($1/\delta$ MSE). Note that, with the above assumptions made, SNR_{MF} represents an upper bound on the highest SNR that can be obtained from CAMP.

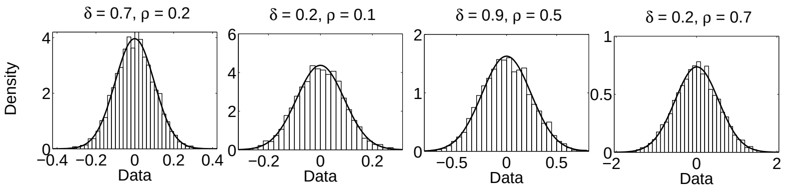
In the remainder of this chapter we will compare the output (or recovery) SNR of CAMP with the output SNR of an (ideal) MF, so that the results will be independent of N or n. However, given the output MF SNR, the input SNR of both MF and CAMP, which depends through N and n on the specific problem being investigated, can be easily derived using (3.7).

3.5 Simulation results

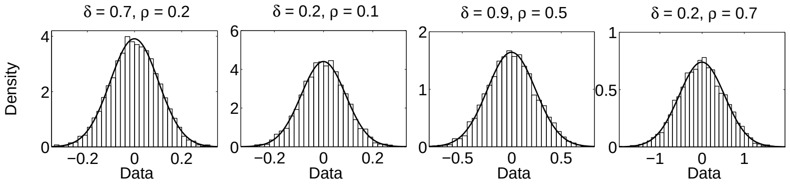
In this section we investigate the performance of median and Adaptive CAMP using MC simulations and compare it with the theoretical results obtained from the SE. We also investigate the performance of the two proposed CS architectures using ROC curves for the cases of fixed threshold and CA-CFAR detectors. Moreover, we consider not only Gaussian sensing matrices, for which SE provably applies, but also partial Fourier matrices, which are of particular interest in radar applications transmitting SF waveforms [59, 61, 136]. Recall that an $n \times N$ partial Fourier matrix is obtained from an $N \times N$ discrete Fourier transform matrix by preserving only a random subset n of the original N matrix rows. The elements of a partial Fourier matrix are given by $a_{i,j} = e^{-j2\pi(g(i)-1)(j-1)/N}$, for $i = 1, \cdots, n$, $j = 1, \cdots, N$ and g is the vector containing the (distinct) indices of the randomly selected rows.

---

[8] This is achieved by using sensing matrices with unit column norms in all simulations.



(a) Histogram of the real part of $\mathbf{w}$.



(b) Histogram of the imaginary part of $\mathbf{w}$.

Figure 3.6: Histograms (bars) of the real and imaginary parts of the noise signal $\mathbf{w}$ for different combinations (δ, ρ) using CAMP with threshold $\tau = 1.8$. A Gaussian pdf is fitted to the histograms (solid line). In these plots $\sigma = 0.1$ and $N = 4000$. The sensing matrix is partial Fourier.

3.5.1 Gaussianity of w using partial Fourier matrices

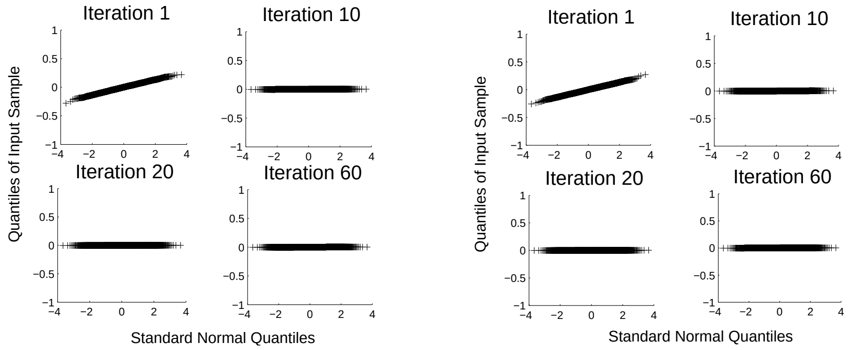
In this section we investigate the Gaussianity of the reconstruction noise vector $\mathbf{w}^t$ for a partial Fourier sensing matrix using MC simulations.

In Figure 3.6 we show a few examples of the empirical distribution of $\mathbf{w}$ at convergence for different combinations of δ and ρ. To obtain the histograms we used CAMP with a fixed (not necessarily optimal) threshold τ and we fixed $\sigma = 0.1$.

In Figures 3.7 and 3.8 we show the quantiles plots (QQplots) of the noise vector along the iterations of CAMP for a few combinations of δ and ρ, for $N = 4000$. In these plots the x-axis indicates the quantiles of the Normal distributions and the y-axis the quantiles of the input samples. The QQplot of a zero mean Gaussian distribution is linear and crosses the origin. Linearity of the QQplots of $\mathbf{w}$ indicate approximate Normality. Also observe that the slope of the line diminishes with iteration number, indicating a decrease in the noise variance.

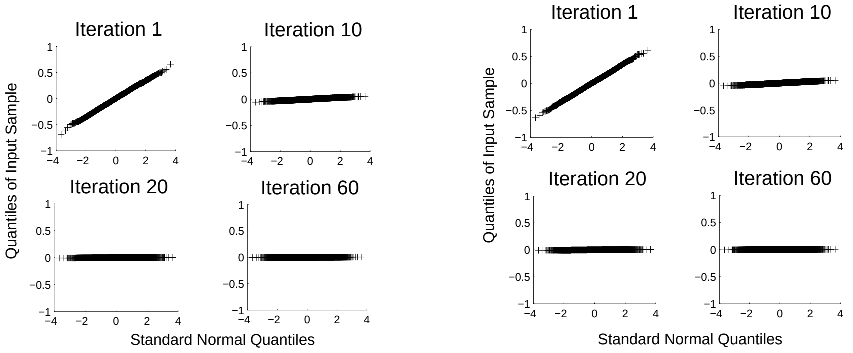
We further investigate the Gaussianity of $\mathbf{w}$ using the Kolmogorov-Smirnov (KS) test [163]. In Tables 3.2 and 3.3 we report the p-values from the KS test for $N = 4000$ and $N = 200$, respectively.

The KS test compares the empirical distribution of the input samples to the Gaussian distribution, and the p-value measures the similarity of the input samples to the reference distribution. In the null hypothesis it is assumed that the samples belong to the reference probability density function. In our simulations, in all cases the null



(a) Real part (b) Imaginary part

Figure 3.7: QQplots of the residual noise vector $\mathbf{w}$ along the iterations of the CAMP algorithm. $\delta = 0.1$, $\rho = 0.01$, $N = 4000$, $\sigma = 10^{-3}$.



(a) Real part (b) Imaginary part

Figure 3.8: QQplots of the residual noise vector $\mathbf{w}$ along the iterations of the CAMP algorithm. $\delta = 0.8$, $\rho = 0.3$, $N = 4000$, $\sigma = 10^{-3}$.

hypothesis was accepted.

We evaluated the histograms, QQplots and KS test also for different combinations of N, δ, and ρ and obtained similar results. Our simulations confirm that the Gaussianity of the noise vector is preserved for partial Fourier matrices as well.

3.5.2 Accuracy of State Evolution

We investigate the accuracy of the SE by comparing the theoretical results obtained from (3.2) with simulation results obtained using the Ideal CAMP algorithm. Moreover, we study the behavior of σ_* for the case of a partial Fourier sensing matrix and

Table 3.2: p-values of the real (R) and imaginary (I) parts of $\mathbf{w}$ for partial Fourier sensing matrix. $N = 4000$.

δ	ρ	p-value (R)	p-value (I)
0.1	0.01	0.7	0.85
0.18	0.5	0.77	0.76
0.2	0.05	0.99	0.84
0.2	0.75	0.83	0.92
0.3	0.6	0.99	0.84
0.4	0.2	0.99	0.98
0.4	0.3	0.46	0.58
0.5	0.15	0.63	0.98
0.6	0.3	0.71	0.8
0.65	0.8	0.62	0.72
0.7	0.6	0.93	0.84
0.8	0.3	0.75	0.78
0.8	0.65	0.86	0.73
0.9	0.1	0.52	0.96
0.9	0.45	0.62	0.36

Table 3.3: p-values of the real (R) and imaginary (I) parts of $\mathbf{w}$ for partial Fourier sensing matrix. $N = 200$.

δ	ρ	p-value (R)	p-value (I)
0.1	0.01	0.84	0.72
0.18	0.5	0.99	0.7
0.2	0.05	0.73	0.98
0.2	0.75	0.86	0.94
0.3	0.6	0.93	0.86
0.4	0.2	0.7	0.84
0.4	0.3	0.7	0.9
0.5	0.15	0.73	0.96
0.6	0.3	0.67	0.82
0.65	0.8	0.89	0.94
0.7	0.6	0.87	0.83
0.8	0.3	0.86	0.97
0.8	0.65	0.93	0.78
0.9	0.1	0.96	0.89
0.9	0.45	0.78	0.9

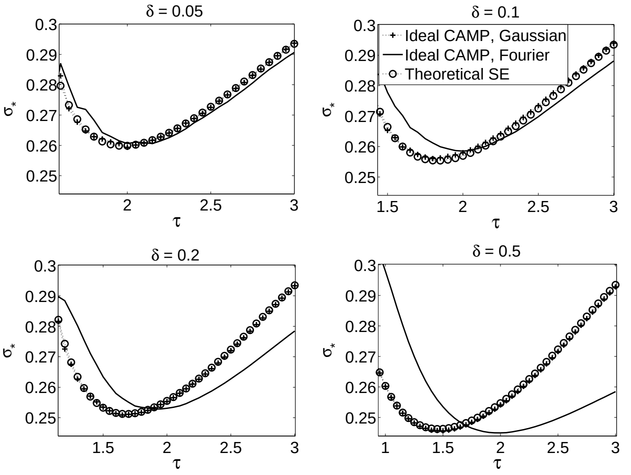


Figure 3.9: σ_* versus τ using Ideal CAMP for both complex Gaussian and partial Fourier sensing matrices. The empirical curves are obtained by averaging over 100 MC samples for $\sigma^2 = 0.05, \rho = 0.05$, $N = 4000$, and different δ. The theoretical SE curve shows the analytical σ_* computed from (A.1) in Appendix A.

investigate how it deviates from the theoretical case of a Gaussian sensing matrix, for which the SE applies. Figure 3.9 compares σ_* obtained from Ideal CAMP for the case of complex Gaussian and partial Fourier sensing matrices with the theoretical one from the SE. The following remarks can be made:

(i) SE correctly predicts the performance of CAMP for the Gaussian sensing matrix.

(ii) SE does not predict the performance of CAMP accurately for partial Fourier matrices, as expected. However, for $\tau = \tau_o$, the value of σ_* for Fourier is not very different from the value for Gaussian.

(iii) As $\delta \to 0$ the predictions of SE become more accurate for the partial Fourier matrix. However, as the number of measurements increases, i.e., $\delta \to 1$, the columns of the partial Fourier matrix become deterministic and orthogonal, and hence the true behavior deviates from the SE that holds for matrices with i.i.d. entries.

(iv) For the partial Fourier matrix, the optimal threshold τ_o seems to be almost the same for different values of δ. Interestingly, although the curves of σ_* are different for different δ's, for a fixed δ and for $\tau > \tau_o$ the variation of the output variance is much smaller in the partial Fourier case than in the Gaussian case. This behavior will have an impact on the difference in performance between

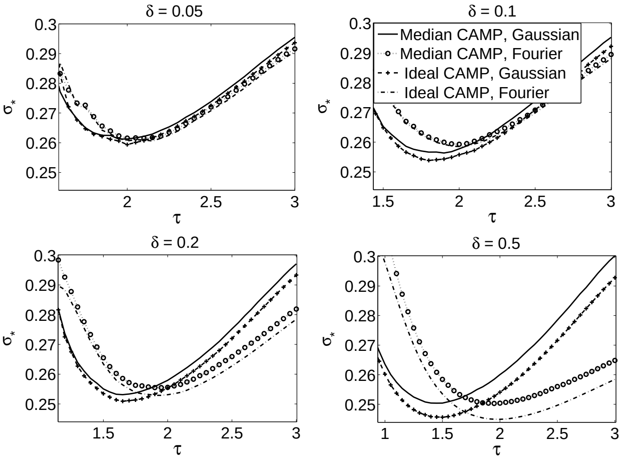


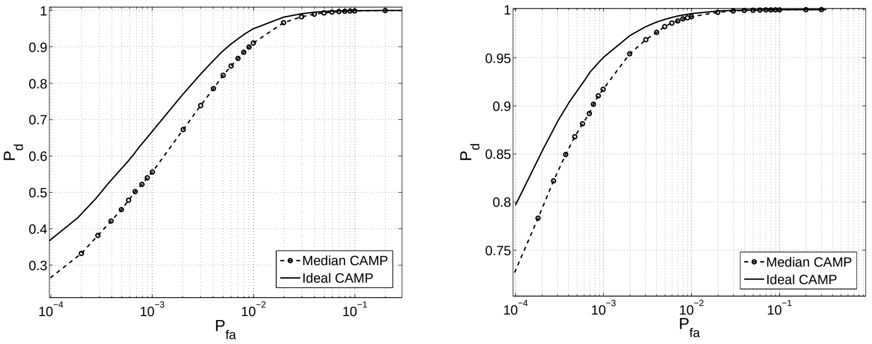
Figure 3.10: Output noise standard deviation versus τ for both Ideal (σ_*) and Median ($\widehat{\sigma}_*$) CAMP for complex Gaussian and partial Fourier sensing matrices. The curves are obtained by averaging over 100 MC realizations for $\sigma^2 = 0.05, \rho = 0.05$, $N = 4000$, and various δ.

Architecture 1 and Architecture 2 for the case of partial Fourier and Gaussian matrices, as the SNR of Architecture 1 varies less along the ROC curves and it is closer to the optimal SNR for the partial Fourier case.

3.5.3 Effects of the median estimator in CAMP

We investigate the effect of replacing the true σ_t with the median based estimate $\widehat{\sigma}_t$ from (3.5) when $\mathbf{x} \neq 0$. In Figure 3.10 the estimated output noise standard deviation is shown for both Ideal and Median CAMP. We observe that the estimate $\widehat{\sigma}_*$ deviates from the Ideal CAMP case because of the bias introduced by the estimator. Furthermore, as is clear from this figure and confirmed by the upper bound provided in (3.6), the deviation diminishes as $\epsilon = \delta\rho$ decreases.

In Architecture 1, overestimating $\widehat{\sigma}_*$ will result in a loss in performance compared to the case of using Ideal CAMP. This is because the soft thresholding function in CAMP uses the parameter $\tau_\alpha \widehat{\sigma}_*$. For a fixed FAP (that fixes τ_α), the bias will have the effect of increasing the overall threshold, therefore resulting in losing detection performance. This is shown in Figure 3.11, where we plot the ROC for Architecture 1 using both Ideal and Median CAMP. From this figure we also observe that, for the value of δ used here, Architecture 1 performs much better when using the partial Fourier sensing matrix as compared to the Gaussian one. This is related to the variation of the noise variance σ_*^2, and therefore the SNR, with the threshold τ_α along the ROC curve. As can be seen from Figure 3.10, in the partial Fourier case as δ



(a) Gaussian sensing matrix

(b) Partial Fourier sensing matrix

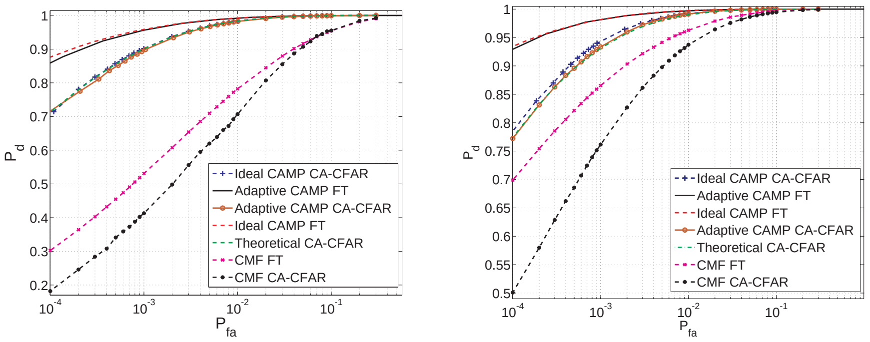
Figure 3.11: ROC curves for Architecture 1 using Ideal and Median CAMP. Here $N = 1000$, $\delta = 0.6$, $\rho = 0.1$, $a^2 = 1$, and $\sigma^2 = 0.05$ (corresponding to a MF SNR = 13dB).

increases the variance curve becomes flatter than in the Gaussian case. This implies that, for decreasing FAPs, in the partial Fourier case the SNR in Architecture 1 deviates much less from the optimum SNR that is achieved for $\tau = \tau_o$.

3.5.4 Adaptive CAMP CFAR detector performance

We investigate now the (P_d, P_{fa}) performance of the Adaptive CAMP CFAR detector. Furthermore, we also compare the performance of Architecture 2 with the Compressive Matched Filter (CMF) [136, 149], which is the filter matched to the subsampled waveform. When using the CA-CFAR processor, this is preceded by a Square Law (SL) detector (recall Figure 3.5) and has a CFAR window of length 20 with 4 guard cells. For all detector cases, the detection probability is estimated for each target separately and by averaging over a number of Monte Carlo simulations. As all targets have equal amplitudes (and therefore equal SNR), we report here the ROC curves for only one of the recovered targets.

It is a well known fact that, if one or more targets are present in the CFAR window, then they cause a rise in the adaptive threshold, thus possibly masking the target in the CUT that has yet to be detected. Therefore, in the basic analysis of CFAR schemes, it is assumed that there are no targets present in the CFAR window. This scenario is referred to as the non-interfering targets scenario. Following a similar approach, we consider the case of multiple but non-interfering targets. For the case of interfering targets, to reduce the interference losses encountered in CA-CFAR, several dedicated CFAR schemes have been proposed in literature, such as OS-CFAR [83]. To keep the discussions concise, we do not pursue these directions and leave them for



(a) Gaussian sensing matrix

(b) Partial Fourier sensing matrix

Figure 3.12: ROC curves for Architecture 2 for different levels of adaptivity. Here $N = 1000$, $\delta = 0.6$, $\rho = 0.1$, $a^2 = 1$, and $\sigma^2 = 0.05$ (corresponding to a MF SNR = 13dB). FT denotes the use of an (ideal) fixed threshold detector.

future research.

Adaptivity imposes extra losses on the system. One loss is due to the use of Adaptive- instead of Ideal CAMP. This means that there is an error in estimating τ_o. A second loss is caused by the CFAR processor and its estimate of the noise standard deviation. In Figure 3.12 we show the ROC curves for Architecture 2 obtained using: (a) Ideal CAMP with ideal (fixed threshold) detector, (b) Ideal CAMP in combination with the CA-CFAR detector, (c) Adaptive CAMP with ideal (fixed threshold) detector, and (d) a fully adaptive scheme consisting of Adaptive CAMP followed by CA-CFAR processor. We also show the theoretical curve of a CA-CFAR processor with the same window length and SNR = 11.55dB for the Gaussian sensing matrix and 11.9dB for the partial Fourier sensing matrix. The SNR can be estimated during simulations, or it can be derived using the procedure described later in Section 3.6. For the Gaussian sensing matrix the optimal threshold for Architecture 2 (using Ideal CAMP) is computed using the SE. For the partial Fourier sensing matrix, the threshold in Ideal CAMP is set to $\tau_o = 1.85$, which is derived from a plot like the ones shown in Figure 3.9 for the case $\delta = 0.6$.

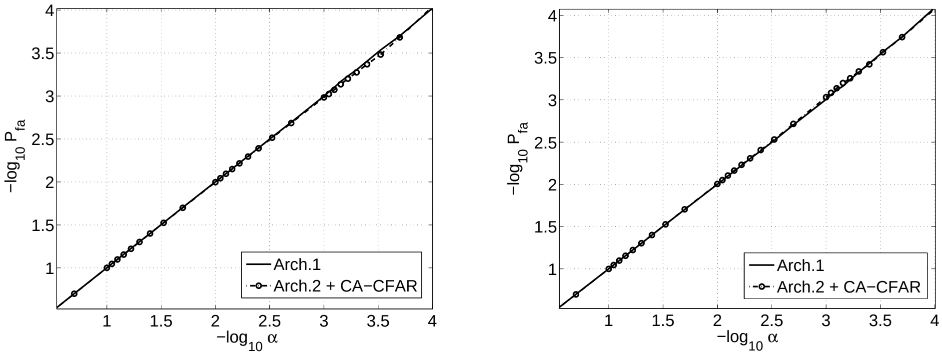
From Figure 3.12 we can make the following observations. First, Adaptive CAMP introduces almost no loss in the detection performance of Architecture 2. This is due to the fact that, although $\widehat{\sigma}_*$ is biased, the value of $\widehat{\tau}_o$ at which the minimum $\widehat{\sigma}_*$ occurs is very close to the true optimal τ_o (see Figure 3.10). The main loss instead is introduced by the adaptive CFAR detector. This is the well-known CFAR loss [79], which can be controlled by changing the CFAR window length. In general, the window length depends on the specific application (e.g., the expected types of target and environment). Furthermore, by comparing the curve of Adaptive CAMP plus CFAR with the theoretical one of a CA-CFAR processor (without CS) with equal parameters, we observe that the CA-CFAR detector performance seems independent of the fact that the input to the detector is obtained by running CAMP instead of a conventional MF.

Second, we observe that when combined with a fixed threshold and a CA-CFAR detector, Architecture 2 significantly outperforms the CMF. This is due to the fact that match filtering with the subsampled waveform produces severe target sidelobes (interference), which result in both an increase of the false alarm rate and the masking of weak targets. We would like to observe that, although most commonly in the radar literature the ROC plots are shown for the case of a single target, in the CS case, and for $N = 1000$, setting $k = 1$ we would observe only a very small region in the (δ, ρ) plane. Instead, in the case of multiple targets we are able to observe both the effects of reconstruction and of the CFAR processor. As it is clear from the SE equation, for the same σ and received target power, changing δ and/or ρ results in a different σ_*, therefore a different CAMP output SNR. This in turns means that even if there are multiple targets, as long as they are not in the CFAR window of one another, the results of the CFAR processor will be independent on the number of actual targets but will depend exclusively on the CAMP output SNR.

By comparing Figures 3.11 and 3.12 it can be seen that, in the fixed threshold case, Architecture 2 always outperforms Architecture 1, as predicted by Theorem 3.2.2. Also in the adaptive case, Architecture 2 followed by a CA-CFAR processor outperforms Architecture 1 using median based CAMP. However, the difference between the two schemes can vary significantly with the system parameters (δ, ρ, σ), sensing matrix type, and CFAR window length. For instance, for the value of δ used in these figures, we observe that Architecture 1 performs much better in the Fourier case than in the Gaussian sensing matrix case. Also, the loss in detection performance is significantly reduced compared to the adaptive detector. This again depends on the behavior of σ_*^2 versus τ. In general, to predict how the two architectures will perform one should observe the behavior of the output noise variance as a function of the threshold τ. If the variation of σ_* versus τ is small, in the ideal detector case the ROC curves of the two architectures will be almost identical, with Architecture 2 always slightly better.

In Figure 3.13 the estimated FAP is shown for both Architecture 1 (which is non adaptive, and uses median based CAMP) and Architecture 2, which uses Adaptive CAMP in combination with a CA-CFAR processor. The desired FAP α, on the x-axis, is used to obtain the threshold multiplier β for the CFAR processor in Architecture 2 and to derive the value of the fixed CAMP threshold τ_α in Architecture 1.

As expected, Figure 3.13 shows that, in homogeneous Gaussian noise, the proposed architectures posses the CFAR property. In simulating the FAP for Figure 3.13,



(a) Complex Gaussian sensing matrix (b) Partial Fourier sensing matrix

Figure 3.13: Estimated FAP versus design FAP α for Architectures 1 and 2. $N = 1000$, and $\delta = 0.6$.

according to hypothesis H_0 (target absent) we generated a measurement vector $\mathbf{y}$ with standard Gaussian distribution and $\mathbf{x} = 0$. However, in practical scenarios, where the noise level may change across range, or in the presence of one or multiple targets located anywhere in the signal $\mathbf{x}$, Architecture 1 can not achieve CFAR. This is because the noise estimate computed by Median CAMP is not performed locally, as in a CFAR processor, but it is based on the whole received signal using the median, which is a biased estimator if $\mathbf{x} \neq 0$. In Architecture 2 instead, separating the reconstruction from the detection gives more flexibility, and, e.g., in a multiple interfering targets scenario, the CA processor can be replaced with a more robust CFAR scheme such as OS-CFAR.

3.6 Design Methodology

Using the tools developed in the previous sections, we propose here a methodology for designing CS radar detectors based on CAMP. Given the detection range, target RCS and system noise, the first step is to compute the transmitted power necessary to reach a given CAMP output SNR. To do this in practice we would use standard radar design methods, e.g., a Blake chart to compute the MF SNR, and then use this value to calculate the CAMP output SNR as a function of δ and ρ (and the associated detection performance). In this section, we focus on the latter part.

For example, if we consider Architecture 2 and we assume that the received power is equal to a^2 for all targets,[9] then using equation (3.2) for a given σ^2 we obtain a

---

[9] This choice of target amplitudes distribution provides a lower bound on the SNR performance, being this the least favorable distribution for the non-zero entries in $\mathbf{x}$ [37].

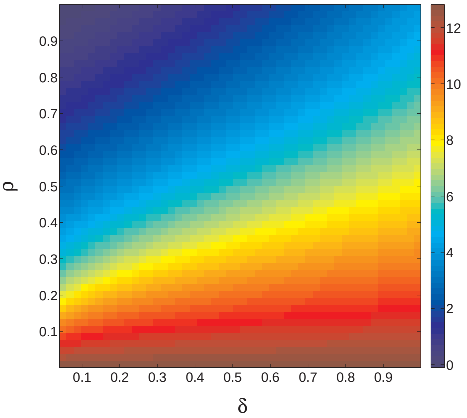


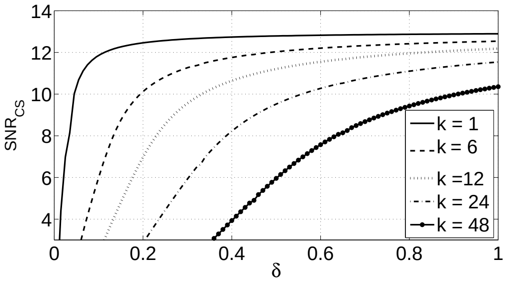
Figure 3.14: CAMP output SNR for Architecture 2 for $a = 1$ and $\sigma^2 = 0.05$, corresponding to $\text{SNR}_{MF} = 13\text{dB}$. The sensing matrix is i.i.d. Gaussian. The colormap is the output SNR of CAMP in a dB scale.

value of σ_*^2 for each couple (δ, ρ). Computing the ratio a^2/σ_*^2 we obtain the CAMP output SNR map, an example of which is shown in Figure 3.14 for a Gaussian sensing matrix.[10] It is clear that CS radar performance depends, besides on the SNR, on δ and ρ. Therefore, in the system design phase, an estimate of the expected number of targets should be made. To design the system for the worst-case scenario, k could be set based on the maximum expected number of targets, i.e., $k = k_{\max}$. Then, for a given number of range (or Doppler) bins N and a given σ, we can vary the number of CS measurements n to obtain several values of $\delta = n/N$ and corresponding $\rho = k_{\max}/n$. By evaluating the SNR at these points, we obtain a curve that shows how power (SNR_{CS}) and undersampling (δ) can be traded against one another.

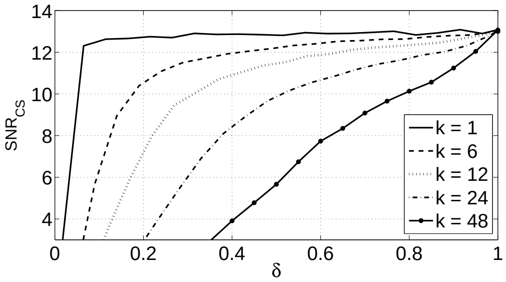
In Figure 3.15 we show an example of such curves for several values of k and $N = 200$. In Figure 3.15(a) the sensing matrix has i.i.d. Gaussian entries and the curves are obtained using the theoretical SE. In Figure 3.15(b) the sensing matrix is partial Fourier and the curves are obtained using MC simulations with Adaptive CAMP. Observe that in both cases the curves are equal for the same k up to approximately $\delta = 0.8$. However, as $\delta \to 1$ the partial Fourier matrix approaches a full Fourier matrix, i.e., it becomes orthogonal, and $\text{SNR}_{CS} \to 13$ dB. On the contrary, even in the limit $n = N$ (i.e., $\delta = 1$) the Gaussian sensing matrix is not orthogonal. Hence, we estimate that if $k > 1$, just as in conventional MF, there will be losses due to

---

[10] For sensing matrices other than Gaussian, the SNR map of CAMP can be obtained via simulations using the Ideal or Adaptive CAMP algorithms, as shown in Figure 3.15(b) for a few sets of points in the map. For Figure 3.14 we used the analytical equations from Appendix A.



(a) i.i.d. Gaussian sensing matrix



(b) partial Fourier sensing matrix

Figure 3.15: CAMP output SNR versus δ for Architecture 2 for different numbers of targets k. Here $N = 200$, $a = 1$, and $\sigma^2 = 0.05$, corresponding to $\text{SNR}_{MF} = 13\text{dB}$.

target interference. In this case the upper bound (SNR= 13dB) cannot be achieved.

Once we have computed the SNR maps or curves, we can use them to predict the performance of our CAMP CFAR detector. Assume, for example, that we would like to choose $\delta = 0.6$ and $k_{\max} = 12$. For this combination (δ, ρ) we derive a value for the CAMP output SNR equal to 11.7dB that can be plugged in the selected CFAR P_d and P_{fa} equations. If we use a CA-CFAR processor and we assume that the targets are not interfering, then for $P_{fa} = 10^{-4}$ we obtain $P_d = 0.7$. This means that, if we want to increase the detection probability for the desired FAP, since $k_{\max}$ is fixed, we either have to increase the number of CS measurements (i.e., increasing δ) or the received power. The latter can be achieved by increasing the transmitted power or any term in the radar equation that increases the received power, such as antenna gain.

We believe the design methodology derived here for the first time represents an important step for understanding and designing CS-based radar systems. In fact, when using Architecture 2, which combines the non-linear ℓ_1-norm minimization with conventional radar detectors, the graphs shown in Figures 3.14 and 3.15 can be used to derived the reconstruction SNR of CAMP that can be plugged in the standard detector equations to predict the performance of the system for different amounts of undersampling and sparsity levels.

3.7 Conclusions

In this chapter we have achieved two main goals. Firstly, we have presented the first architecture for adaptive CS radar detection with CFAR properties. Secondly, we have provided a methodology to predict the performance of the proposed detector, which makes it possible to design practical CS based radar systems.

These goals have been achieved by exploiting CAMP, which features closed-form expressions for the detection and false alarm probabilities in the ideal case of known system parameters. Based on the SE theoretical results, we have demonstrated that, out of two proposed architectures, the combination of a recovery stage based on CAMP with a separate detector achieves the best performance. With a simple modification to CAMP, we have combined conventional CFAR processing with ℓ_1-norm minimization to provide a fully adaptive detection scheme. Our theoretical findings have been supported with evidence from Monte Carlo simulations.

Furthermore, by comparing theoretical and simulated results we have been able to understand the behavior of CAMP for sensing matrices other than i.i.d. Gaussian. In fact, we have shown by means of simulations that our conclusions still hold for the case of partial Fourier sensing matrices, for which, unfortunately, no theoretical claims can yet be made.

We have derived closed form expressions for the CAMP output SNR as a function of system and target parameters. These relations can be used to obtain CS link budget plots that allow the system designer to evaluate the trade off between power and undersampling. Such charts play an important role in determining when and how CS can be applied and at what cost.

In the next chapter we present the results of the first CS radar experimental measurements. The results confirm that our theoretical and simulated predictions still hold even in more realistic settings.

Chapter 4

Experimental Results on Compressive Sensing radar

In this chapter we present some preliminary experimental results to demonstrate that the proposed CAMP detectors can be applied to real CS radar data. The measurements were collected in September 2011 at Fraunhofer FHR, Germany, with the Fraunhofer experimental LabRadOr radar system. The transmitted waveforms were designed to achieve a set of CS SF radar measurements [59,61,130]. The performance of the two CAMP based detection schemes presented in Chapter 3 are compared by means of Receiver Operating Characteristic curves. Our experimental results show that our theoretical conclusions hold even in the non-asymptotic setting and the proposed Adaptive CFAR CAMP detector could potentially be used in operational CS radars.

Furthermore, the performance of CAMP based architectures are also compared to those obtained with the Compressive Matched Filter (CMF) [130,149], previously discussed in Section 2.4. As expected, the CAMP based CFAR radar detector performs significantly better than the CMF.[1]

4.1 Experimental Set-up

In our experiments, we consider the case of a one dimensional radar operating in the range domain. We use as targets five stationary corner reflectors with different RCS. Figure 4.1 shows the five corner reflectors and their relative positions as they were

---

[1] This chapter is based on articles [**J1**], [**C2**], [**C3**] and [**C9**] (a list of the author's publications is included at the end of this dissertation, p. 183.)

Figure 4.1: Five corner reflectors used during the experimental measurement campaign.

placed during the measurements. For each transmitted burst (containing n pulses, see Figure 2.1) 300 measurements with the same set-up were performed. A description of the radar system and the waveforms used in the measurements is given below.

4.1.1 Radar system

The measurements were carried out at Fraunhofer FHR, in Germany, using the LabRadOr experimental radar system, which is shown in Figure 4.2. LabRadOr is a software defined pulsed radar, with maximum transmit power of 32dBm (and an attenuator of 1dB step size) using separate transmit (TX) and receive (RX) reflector antennas, with gain of 31.6dB each. The digital waveform designed by the user is transferred from the control computer to an FPGA, where a Digital-to-Analog Converter (DAC) converts the digital data to an analog signal. The analog waveform is then transferred to an RF front-end for up conversion to the carrier frequency $f_c = 8.9$ GHz. At the receiver, after down conversion, the signal is returned to the FPGA, where an Analog-to-Digital Converter (ADC) samples the received analog signal at 2 GHz sampling rate. The real time samples are then transferred to the control unit where they are stored for further processing. Because of internal FPGA limitations, the maximum number of samples per sweep that can be recorded is 1024, thus limiting the receiver record window length to 512nsec. The start time of the record window can be set by the user within the Pulse Repetition Interval (PRI), which is fixed and equal to 10msec. Since our objective is to perform SF measurements, but the maximum TX pulse length is limited to 512nsec, we transmit one frequency per pulse, and later combine all the n frequencies (thus n pulses) to obtain a single SF measurement. Hence, we assume that the scene is stationary at least to within $n\,$PRI seconds. Also, since both the corners and the radar are fixed, the target amplitudes

Figure 4.2: LabRaDor radar system of Fraunhofer FHR used during the experimental measurement campaign.

can be modeled as Swerling Case 0 [164].[2]

4.1.2 Transmitted waveform

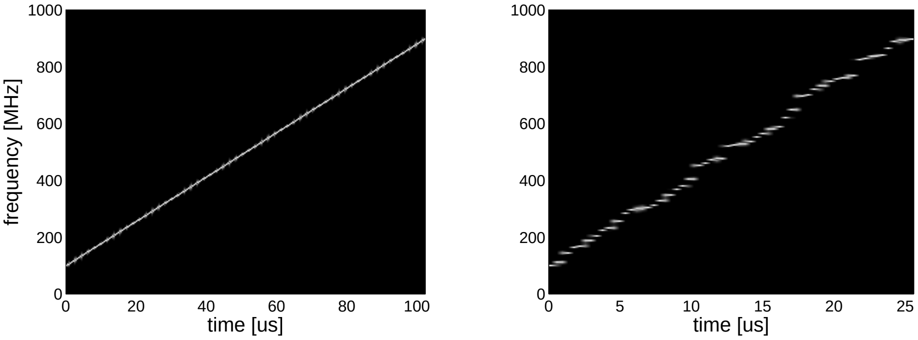
For our experiments, we designed a set of stepped frequency waveforms [59, 130] and the TX signal consists of a number of discrete frequencies f_m, ranging from 100 to 900MHz. In the Nyquist case (that represents unambiguous mapping of ranges to phases over the whole bandwidth) we transmit $N = 200$ frequencies over a bandwidth of 800 MHz. The achievable range resolution is therefore $\delta_R = 18.75$ cm. Each frequency is transmitted during 0.512 μs, thus implying a bandwidth of $B_f = 1.95$ MHz, and sequential frequencies are separated by $\Delta f = 4$ MHz, resulting in an unambiguous range of $R_{un} = 37.5$ m.

In the CS case, the number of TX frequencies is reduced from N to n ($n < N$). We used $n = 50$ and 100, that correspond, for $N = 200$ and $k = 5$, to $\delta = 0.5, 0.25$ and $\rho = 0.05, 0.1$. The subset of transmitted frequencies is chosen uniformly at random within the total transmitted bandwidth, with the constraints that we always use the first and last frequencies in the bandwidth (to span the same total bandwidth to preserve range resolution), and we also force at least two of the transmitted frequencies to be separated by the nominal frequency separation Δf, to guarantee that the unambiguous range is preserved. The spectrograms of the Nyquist waveform and of one of the CS TX waveforms for the case $n = 50$ ($\delta = 0.25$) are shown in Figure 4.3.

After reception and demodulation, each range bin, $r_i = r_0 + i\,\delta R$, $i = 1, \ldots, N$, maps to n phases proportional to the n transmitted frequencies $f_m = f_c + g(m)\Delta f$, $m = 1, \ldots, n$ and g is the vector containing the indices of the randomly selected frequencies. Therefore, the sensing matrix $\mathbf{A}$ can be represented as a partial Fourier

---

[2] Since we are interested in the detection problem from a single range measurements, we kept the targets fixed and did not perform any Doppler measurements.



(a) Nyquist waveform

(b) CS waveform

Figure 4.3: Spectrogram of TX waveform. Left: Nyquist waveform with $N = 200$; right: CS waveform with $n = 50$ and $\delta = 0.25$.

matrix, and the n samples y_m of the compressed measurement vector $\mathbf{y}$ are as in (2.11). While performing the measurements, we adjusted the per frequency transmitted power in a way that the same total power is transmitted in each burst, irrespective of the number of transmitted frequencies. This means that when the number of measurements is reduced by a factor δ, the power per transmitted frequency P_T is $1/\delta$ times higher than in the Nyquist waveform case, so that the total transmitted energy $(P_T \times n/B_f)$ is the same in all cases, irrespective of n. This enables us to analyze the effects on detection performance due only to measurement undersampling and not due to a reduction in the transmitted power as δ decreases.

4.2 CAMP detectors performance

In this section, we analyze the performance of the two proposed CAMP based detection schemes and compare them to both the MF (with full set of frequencies) and the CMF (with the same undersampling factor used in CAMP).

4.2.1 Reconstructed range profile

Figure 4.4 exhibits the signals reconstructed with Architectures 1 and 2 based on CAMP, in addition to SPGℓ_1 solver [28], CMF [130, 149], and MF. We set $\delta = 0.5$ for the CS algorithms and use $N = 200$ measurements for the MF. In this figure the five corner reflectors, indicated as T1–T5, are visible at ranges from 20m to 36m. For Architecture 1, τ_α was set using $\alpha = 10^{-4}$. For SPGℓ_1, ε was manually tuned

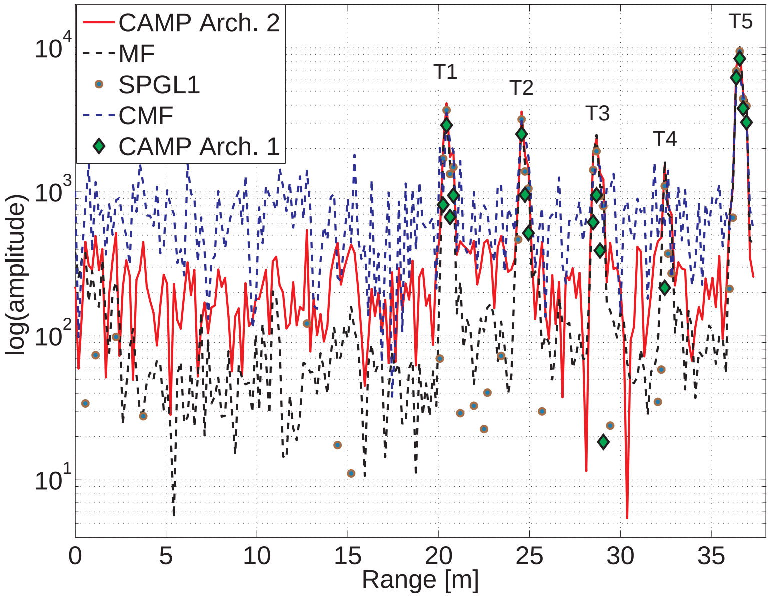


Figure 4.4: Reconstructed range profile using: CAMP Architectures 1 and 2, CMF, SPGℓ_1 and MF. For the MF, $N = 200$; for CS, $\delta = 0.5$.

to achieve the best reconstruction (by means of visual inspection). As expected, the signal reconstructed from the CMF is much more noisy than that of CAMP Architecture 2. In the CMF, the excessive noise outside the target bins is not actual system noise, but it is due to reconstruction from an undersampled spectrum (target sidelobes).

From Figure 4.4 we also observe that, as explained in Section 2.4.2, since the targets are not exactly on Fourier grid points, there is a leakage of target power into neighboring range bins both for MF and CS. In this case, the resolution of CAMP is comparable, and not superior, to that of the MF.

As can been seen in Figure 4.4, the SNR is very high for all the targets (in all cases above 20 dB). To evaluate the performance of the detectors at medium SNR values, we added white Gaussian noise (with $\sigma = 500$) to the raw frequency data samples. We set the output SNR of the MF to 17.2, 16.6, 14, 10.2 and 26 dB, from the closest corner to the farthest one.

For MF, CMF and CAMP Architecture 2, we estimate the SNR after reconstruction, i.e., the SNR that is input to the detector, separately for each target and for

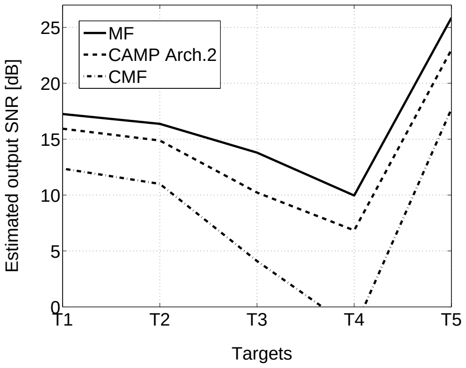


Figure 4.5: Estimated output SNR after reconstruction using MF (with Nyquist sampling), CAMP Architecture 2 and CMF (with $\delta = 0.5$) for targets T1–T5.

each value of δ. Again, we emphasize that the reconstruction SNR is the ratio of the estimated target power to the system- plus reconstruction noise. The estimated SNR for the five targets is shown in Figure 4.5 for $\delta = 0.5$. We observe that, as anticipated, the CMF SNR is lower than that of CAMP. In fact, as we estimate the noise (plus clutter and interference) level from the target free range bins, the presence of targets sidelobes at these range bins will act as excessive non Gaussian interference, which not only degrades the SNR but also compromises the performance of the CA-CFAR detector. Since a loss in SNR translates directly into a loss in the detection probability, for a given FAP, we predict that CAMP will perform better than CMF. So, while the total transmitted power remains fixed, the reconstruction SNR depends on the reconstruction scheme, in addition to the undersampling factor δ. In fact, for a fixed algorithm, the reconstruction SNR decreases as we reduce δ.

In Figure 4.6 we plot the estimated output noise standard deviation ($\hat{\sigma}_*$) for the experimental data where the range profiles were reconstructed using Median CAMP for different values of the threshold τ. The curve is obtained by averaging over all 300 measurements. Note that, for the same input noise variance, for $\delta = 0.25$ the output noise power is always higher than for $\delta = 0.5$, also implying that for the same target received power the SNR decreases with δ. We also see that the behavior of the estimated output noise standard deviation resembles the one shown is Section 3.5.3 for the simulated data.

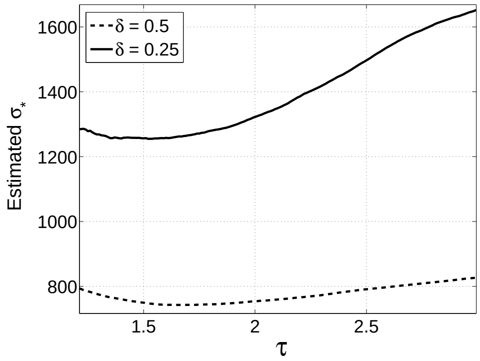


Figure 4.6: Estimated σ_* versus τ using Median CAMP.

4.2.2 ROC curves

We now present the result of detection performance of the newly proposed CAMP based architectures and compare them against the CMF. The ROC plots are shown for both the ideal case of fixed threshold (non-adaptive) detector for the CMF and CAMP Architectures 1 and 2 and for the adaptive detector case using a CA-CFAR processor for CMF and CAMP Architecture 2. For the P_d estimation, we used the detection at the location of the highest target peak. The CA-CFAR processor uses four guard cells and a CFAR window of length 20. An example of the reconstructed range profile using CAMP Architecture 2 with $\delta = 0.25$, together with the CA-CFAR adaptive threshold is shown in Figure 4.7. The CA-CFAR threshold multiplier is set for $\alpha = 10^{-4}$. The black circles in the figure indicate the range positions of the five targets that were used for estimation of the P_d.

In the ROC plots, the estimated detection probability for each target is plotted against the design FAP, that is used to set the detector threshold. For all detector cases (adaptive and non-adaptive), the CAMP reconstruction threshold τ_o of Architecture 2 is always adaptive. In Architecture 1 the threshold τ_α is set to $\tau_\alpha = \sqrt{-\log \alpha}$, where α is the design P_{fa}, and the same threshold $\tau_\alpha \hat{\sigma}$ is applied to all range bins inside Median CAMP. Recall that in Median CAMP the presence of targets in the received signal will result in overestimating the noise std $\hat{\sigma}_*$, with a consequent reduction in P_d compared to the ideal case, as described in Section 3.5.3. Furthermore, using Architecture 1 the highest FAP that can be achieved is equal to δ, since the sparse estimated signal $\widehat{\mathbf{x}}$ cannot have more than n out of N non-zero coefficients.

For the fixed threshold case, Figure 4.8 illustrates that, as claimed from our the-

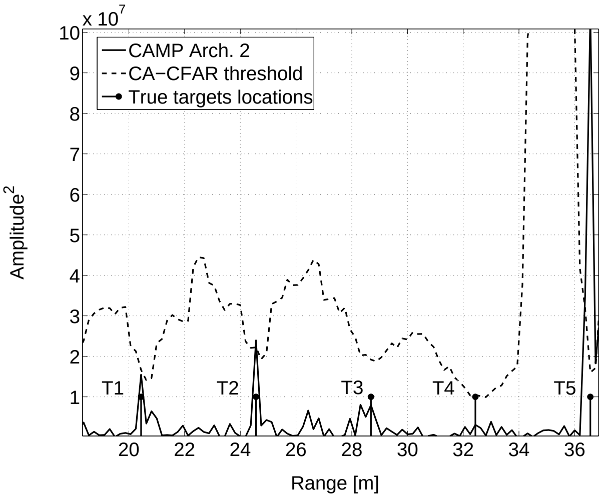


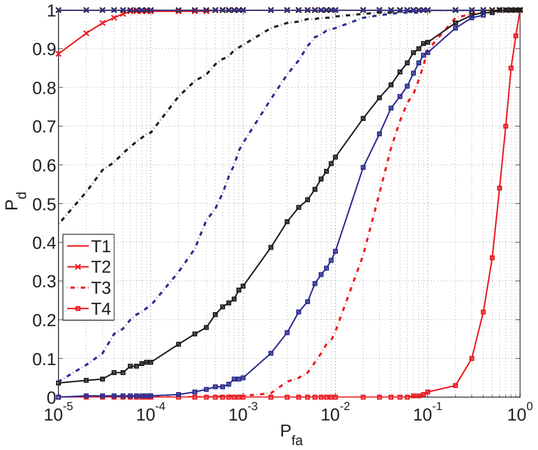
Figure 4.7: Zoom in of reconstructed range profile using CAMP Architecture 2 and CA-CFAR adaptive threshold. Black, solid line: squared amplitude of the reconstructed range profile using CAMP Architecture 2 with $\delta = 0.25$. Black, dashed line: CA-CFAR adaptive threshold for $\alpha = 10^{-4}$. Black circles: true targets positions.

oretical and simulated results, for a given FAP the detection probability of CAMP Architecture 2 is always higher than that of CAMP Architecture 1 (and hence any ℓ_1-solver). Also, as expected in a multiple target scenario, all CAMP architectures perform better than the CMF.

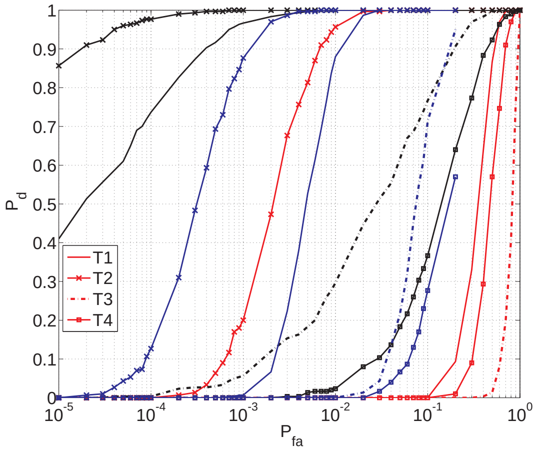
For the adaptive detector case, shown in Figure 4.9, it can be seen that, for a given FAP there is a loss in P_d when using the CA-CFAR processor compared to the fixed threshold detector. This is the well-known CFAR loss. Also with the CFAR processor, CAMP Architecture 2 performs still better than Architecture 1 and significantly better than CMF.

Note that in Figures 4.8 and 4.9 the ROC curves for target T5 are not shown as they are equal to 1 for all detectors in all cases. Furthermore, if we compare the cases $\delta = 0.5$ and 0.25, both in the ROC curves and in the range profiles illustrated in Figures 4.7 and 4.4, we observe that, for the same P_{fa}, target T1 has the highest P_d for the case $\delta = 0.5$ whereas target T2 has the highest P_d for the case $\delta = 0.25$. This is due to the fact that each measurement setting (TX waveform and power) results in sampling of the main target lobe at a slightly different location (due to small timing variations in the signal generation and transmission) and therefore a variation of the associated straddling loss.

If we compare the ROCs of the Adaptive CAMP CFAR detector to the ones of Architecture 1 that uses a fixed threshold, we observe that, in some particular



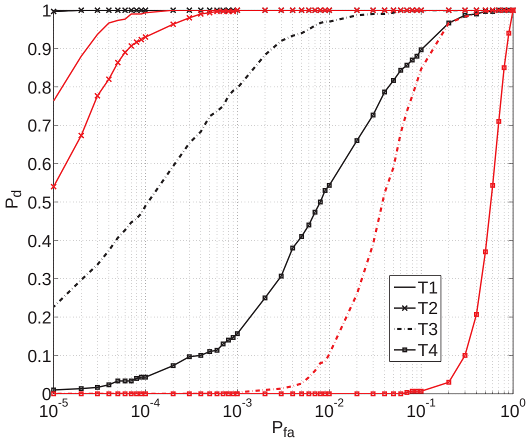
(a) $\delta = 0.5$



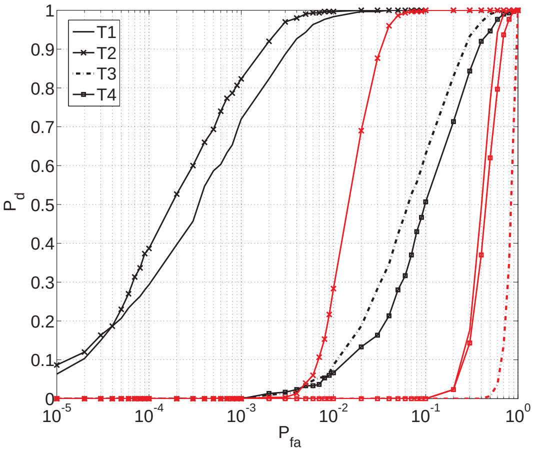
(b) $\delta = 0.25$

Figure 4.8: ROC curves using: CAMP Architecture 1 (blue lines), CAMP Architecture 2 with fixed threshold detector (black lines), CMF with fixed threshold detector (red lines). $\delta = 0.5$ (4.8(a)) and $\delta = 0.25$ (4.8(b)). Each line style corresponds to a different target, as indicated in the legend.

scenarios, the performance of the two detectors are very similar. However, at low FAPs and high P_d, which is the most relevant case in practical situations, Architecture 2 with the CFAR detector always outperforms Architecture 1. Furthermore, we observe that Architecture 1 is comparable to an OS-CFAR detector where the CFAR window is the entire signal, i.e., $2N_w = N$, including the CUT. Clearly, a serious disadvantage



(a) $\delta = 0.5$



(b) $\delta = 0.25$

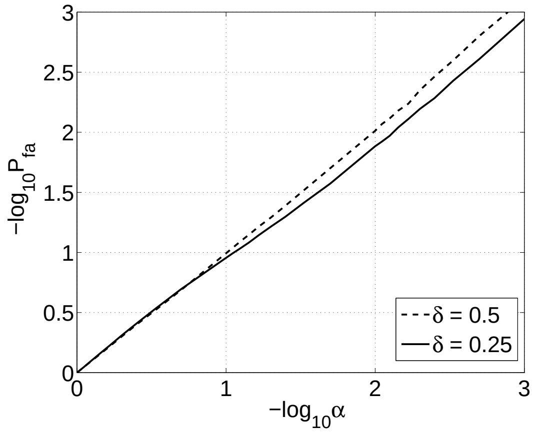
Figure 4.9: ROC curves using: CAMP Architecture 2 with CA-CFAR detector (black lines), CMF with CA-CFAR detector (red lines). $\delta = 0.5$ (4.9(a)) and $\delta = 0.25$ (4.9(b)). Each line style corresponds to a different target, as indicated in the legend.

of Architecture 1 is that, since the entire signal is used in the noise estimation and the threshold τ_α is fixed, it can not adapt to local variation of noise level. This makes Architecture 1 unsuitable for many radar applications. Architecture 2, in contrast, provides the flexibility to choose both the most appropriate CFAR processor and CFAR window length depending on the specific scenario. The discrepancy in performance between the two detectors depends also on the targets distribution in range. In fact, whereas Architecture 2 would have significant performance losses if multiple targets are present in the same CFAR window when using a CA-CFAR, Architecture 1 is insensitive to the target locations, and therefore would have the same performance. However, in Architecture 2 one has the freedom to choose the most appropriate CFAR processor for the scenario at hand, e.g., an OS-CFAR would be preferable in a multiple targets scenario. Instead, in the current implementation of Median CAMP that is used in Architecture 1, the lack of adaptivity to local variation of the noise and clutter levels represents a major drawback.

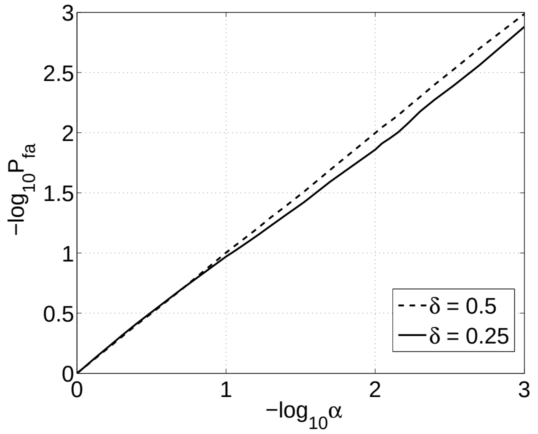
As we only performed measurements with targets, we are unable to evaluate the CFAR property of Architecture 1. However, for CAMP Architecture 2, we can demonstrate that our model $\widetilde{\mathbf{x}} = \mathbf{x} + \sigma_* \mathbf{w}$ is correct by estimating the FAP from the reconstructed noisy signal $\widetilde{\mathbf{x}}$ by excluding the range bins corresponding to the target locations plus four guard cells. If our model is correct, and the noise in the signal $\widetilde{\mathbf{x}}$ is Gaussian, then the estimated FAP should correspond to the design FAP used to set the detector threshold. This should be true for both the fixed threshold and the CFAR detector. This is demonstrated in figure 4.10, where the estimated P_{fa} is plotted versus the design FAP α for CAMP Architecture 2 using both the CA-CFAR and the fixed threshold detector. From this figure we observe that, as expected, the estimated FAP matches the design one, confirming that our model is correct.

4.3 Conclusions

In this chapter we investigated the use of the novel CAMP based detection schemes on the first experimental CS radar measurements. We have analyzed the performance of the proposed CAMP based detectors in the non-asymptotic setting using experimental measurements and compared the ROCs obtained with our adaptive and non-adaptive schemes with that of the Compressive Matched Filter (CMF) under equal settings. The results show that, as predicted by our theoretical findings, the approaches based on CAMP have significantly better performance than the CMF and that CAMP Architecture 2 achieves better performance than Architecture 1. Furthermore, we were able to implement the first fully adaptive CFAR CS radar detector, that requires no prior knowledge of either the number of targets or the noise level and demonstrated its CFAR property. We believe this work paves the way for the design and development of more sophisticated CS CFAR radar detectors, as further discussed in Section 7.3, page 128.



(a) Fixed threshold detector



(b) CA-CFAR detector

Figure 4.10: Estimated FAP versus design FAP α for CAMP Architecture 2 using fixed threshold detector (4.10(a)), and CA-CFAR detector (4.10(b)). $\delta = 0.5$ (dashed line) and $\delta = 0.25$ (solid line).

Chapter 5

Importance Sampling for Space Time Adaptive detectors

In this chapter we demonstrate the use of IS methods on the class of Normalized Matched Filter (NMF) Space Time Adaptive Processing (STAP) detectors and propose two variants of the square-law detector that use the Envelope (E) and Geometric Mean (GM). In the first part of this chapter, using the classical NMF detector, we demonstrate how IS can be used to both derive analytical closed form expressions and to speed up simulations for evaluation of false alarm probabilities. We extend the theoretical and numerical results, first presented in [101, 165], for deriving a powerful 2-dimensional biasing scheme that is used to evaluate the performance of the adaptive NMF, the Normalized Adaptive Matched Filter (NAMF) detector.

In the second part of this chapter, following a similar approach as in [96], we propose two novel NAMF detector variants, the envelope and geometric-mean NAMF detectors. Their CFAR property is established and threshold settings for the detectors for specified false alarm probability is accomplished using fast simulation. Performance analysis of these detectors reveals almost indistinguishable loss in detection probability in homogeneous Gaussian interference compared to conventional square-law STAP detector versions. In addition, they exhibit robust detection performance in the presence of interfering targets in the training data, for both non-fluctuating as well as fluctuating target models.[1]

---

[1] This chapter is based on articles [**J3**], [**C20**] and [**R1**] (a list of the author's publications is included at the end of this dissertation, p. 183.)

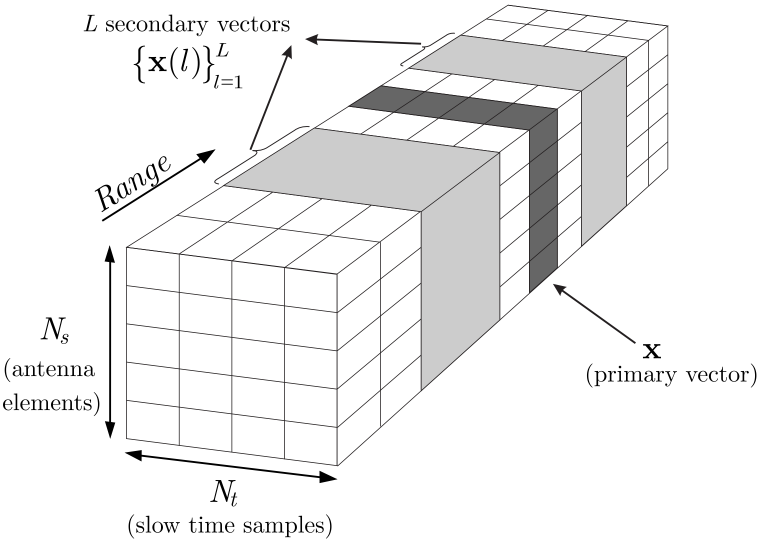


Figure 5.1: STAP data cube during one CPI.

5.1 Space Time Adaptive Processing

It is a well-established fact that the detection performance of many radar systems, typically operating in time (fast and slow) domain, can be significantly improved by simultaneously including a spatial domain (i.e., multiple antenna elements) in an adaptive way. This is the aim of Space Time Adaptive Processing [166]. As the principles of STAP are well understood, we proceed immediately to a description of the STAP data cube.

Consider a radar system consisting of a linear array of N_s antenna elements. A burst of N_t pulses is transmitted during a Coherent Processing Interval (CPI) and each element receives as many return samples in any one range gate. The complex (because of I and Q channel processing) samples measured during one CPI from all antenna elements in all range bins can be arranged in a data cube, as shown in Figure 5.1. The $N_sN_t = N$ samples in the range gate of interest are referred to as the primary data. They may contain a target and represent the range to be tested. The samples are rearranged in an $N \times 1$ column vector and denoted as $\mathbf{x}$. The target return is modeled as consisting of a known (direction-Doppler) vector $\mathbf{s}$, called the steering vector, with an unknown complex amplitude (a) in addition to clutter, interference, and noise. The steering vector $\mathbf{s}$ is given by the Kronecker product of the temporal vector $\mathbf{s}_t$ and the spatial vector $\mathbf{s}_s$ [166], i.e.,

$$\mathbf{s} = \mathbf{s}_t \otimes \mathbf{s}_s,$$

where

$$\begin{aligned}\mathbf{s}_t &= [1\ z_t\ z_t^2 \ldots z_t^{N_t-1}]^T, \\ \mathbf{s}_s &= [1\ z_s\ z_s^2 \ldots z_s^{N_s-1}]^T, \\ z_t &= e^{j2\pi f_d}, \\ z_s &= e^{j2\pi \frac{d}{\lambda}\cos\phi_t},\end{aligned}$$

and f_d is the target Doppler frequency normalized to the Pulse Repetition Frequency (PRF), ϕ_t is the look angle and d is the antenna elements spacing.[2]

There are L other N-length complex vectors, called the training or secondary data, obtained from as many nearby range gates and assumed to be free of target signal. These are denoted as $\mathbf{x}(l)$, $l = 1, \ldots, L$. It is assumed from now on that the training data is free of other targets or contamination[3]. In the absence of target, the primary and secondary data vectors are assumed to be jointly independent and complex Gaussian, sharing the $N \times N$ covariance matrix $\mathbf{R} = E\{\mathbf{x}\mathbf{x}^\dagger\}$, where the superscript † denotes complex conjugate transpose.

The binary hypothesis test for detecting the presence of a target can be set up as

$$\begin{aligned}&\mathrm{H}_0 : \mathbf{x} = \mathbf{n} \\ &\mathrm{H}_1 : \mathbf{x} = a\mathbf{s} + \mathbf{n}\end{aligned} \tag{5.1}$$

where $\mathbf{n}$ is the Gaussian noise plus clutter vector with covariance matrix $\mathbf{R}$.

For detecting the presence of a target from a given direction $\mathbf{s}$, the famous paper by Reed, Mallett and Brennan (RMB) [167] shows that the optimum filter weights vector $\mathbf{w}$ that maximizes the SNR has the form

$$\mathbf{w} = C\mathbf{R}^{-1}\mathbf{s}, \tag{5.2}$$

where C is a scalar constant. This is the well-known RMB beamformer.

Later, in 1974 Kelly [168] derived a test statistic to perform adaptive target detection based on the Generalize Likelihood Ratio Test (GLRT), since both the target amplitude and the noise covariance matrix are unknown. The GLRT test has the form

$$\Lambda_{\mathrm{GLRT}} \equiv \frac{|\mathbf{s}^\dagger \widehat{\mathbf{R}}^{-1}\mathbf{x}|^2}{\mathbf{s}^\dagger \widehat{\mathbf{R}}^{-1}\mathbf{s}\left(1 + \frac{1}{L}\mathbf{x}^\dagger \widehat{\mathbf{R}}^{-1}\mathbf{x}\right)} \underset{H_0}{\overset{H_1}{\gtrless}} L\eta. \tag{5.3}$$

---

[2] The steering vector $\mathbf{s}$ must be determined for each look angle and Doppler frequency of interest in a given range bin.

[3] In Section 5.4.3, detection performance is carried out for training data contaminated by interfering targets.

Since the second term inside the brackets in the denominator would impose a serious computational load in any real time implementation, a simplified version of the GLRT test was derived in [169], resulting in the well-known Adaptive Matched Filter (AMF) detector, given by

$$\Lambda_{\mathrm{AMF}} \equiv \frac{|\mathbf{s}^{\dagger}\mathbf{R}^{-1}\mathbf{x}|^2}{\mathbf{s}^{\dagger}\mathbf{R}^{-1}\mathbf{s}} \underset{H_0}{\overset{H_1}{\gtrless}} \eta. \tag{5.4}$$

Since then, many sophisticated STAP detectors have been designed, to deal with both computational load and with different types of clutter and interference scenarios. In [170] an overview can be found abut STAP related issues and proposed detectors solutions.

The first successful attempt to using IS for characterizing STAP detectors is presented in [101], and further extended in [96]. Building up upon that work, in this chapter we focus our attention to the class of NMF STAP detectors.

5.2 The Normalized Matched Filter detector

The NMF detection test is given by

$$\Lambda_{\mathrm{NMF}} \equiv \frac{|\mathbf{s}^{\dagger}\mathbf{R}^{-1}\mathbf{x}|^2}{(\mathbf{s}^{\dagger}\mathbf{R}^{-1}\mathbf{s})(\mathbf{x}^{\dagger}\mathbf{R}^{-1}\mathbf{x})} \underset{H_0}{\overset{H_1}{\gtrless}} \eta. \tag{5.5}$$

The NMF test statistic Λ_{NMF} is equivalent to the square magnitude of the correlation coefficient between the two transformed vectors $\mathbf{R}^{-1/2}\mathbf{x}$ and $\mathbf{R}^{-1/2}\mathbf{s}$. Its FAP in Gaussian interference is known, [171], and particularly easy to derive. It differs from the matched filter (for known interference covariance matrix $\mathbf{R}$) in the normalization term which is the second one in the denominator. The FAP of the detector is given by

$$\alpha_{\mathrm{NMF}} = (1-\eta)^{N-1},$$

and it has the CFAR property of being invariant to the interference covariance matrix $\mathbf{R}$. Furthermore, as is evident from (5.5), the FAP is also invariant to any scaling of the primary and secondary data.

5.2.1 The FAP of the NMF detector using IS

With a simple (re-) derivation of the FAP of this detector, we illustrate an elegant aspect of IS, by which it is sometimes possible to derive a perfect estimate of a rare-event probability.

It is assumed that $\mathbf{x} \sim \mathcal{CN}_N(0, \mathbf{R})$. With a whitened data vector defined as $\mathbf{x}_1 \equiv \mathbf{R}^{-1/2}\mathbf{x}$ and a transformed steering vector $\mathbf{s}_1 \equiv \mathbf{R}^{-1/2}\mathbf{s}$, the test statistic of

(5.5) takes the form

$$\Lambda_{\mathrm{NMF}} = \frac{|\mathbf{s}_1^\dagger \mathbf{x}_1|^2}{(\mathbf{s}_1^\dagger \mathbf{s}_1)(\mathbf{x}_1^\dagger \mathbf{x}_1)} = \frac{|y_1|^2}{\|\mathbf{x}_1\|^2}, \tag{5.6}$$

where $y_1 = \mathbf{u}_1^\dagger \mathbf{x}_1$ is a scalar random variable and $\mathbf{u}_1 = \mathbf{s}_1/\|\mathbf{s}_1\|$ is an N-dimensional unit vector. Using for example a Gram-Schmidt procedure, an N-dimensional basis can be formed by determining $N-1$ other unit vectors in the orthogonal subspace of $\mathbf{u}_1$. Denoting the former by $\mathbf{u}_i$, $i = 2, \ldots, N$, we define the corresponding random variables $y_i = \mathbf{u}_i^\dagger \mathbf{x}_1$. Then

$$|y_1|^2 + \sum_{i=2}^N |y_i|^2 = \sum_{i=1}^N |y_i|^2 = \mathbf{x}_1^\dagger \sum_{i=1}^N \mathbf{u}_i \mathbf{u}_i^\dagger \, \mathbf{x}_1 = \|\mathbf{x}_1\|^2, \tag{5.7}$$

with the $\{y_i\}_1^N$ being i.i.d. and distributed as $\mathcal{CN}_1(0,1)$. The test therefore becomes

$$\Lambda_{\mathrm{NMF}} = \frac{|y_1|^2}{|y_1|^2 + \sum_{i=2}^N |y_i|^2} \underset{H_0}{\overset{H_1}{\gtrless}} \eta, \tag{5.8}$$

which can be put in the form

$$\frac{u}{\sum_{i=2}^N v_i} \underset{H_0}{\overset{H_1}{\gtrless}} \eta_o \equiv \frac{\eta}{1-\eta}, \tag{5.9}$$

where

$$u \equiv |y_1|^2 \quad \text{and} \quad v_i \equiv |y_i|^2, \quad i = 2, \ldots, N.$$

The FAP of the NMF detector is then

$$\alpha_{\mathrm{NMF}} = P_{H_0}\big(U \geqslant \eta_o \sum_{i=2}^N V_i\big) = E\Big\{P_{H_0}\big(U \geqslant \eta_o \sum_{i=2}^N V_i \Big| V_2, \ldots, V_N\big)\Big\}. \tag{5.10}$$

The first line above has exactly the same form as the FAP of a CA-CFAR detector as, under H_0, U and each V_i are i.i.d. (unit) exponential random variables; the formula for this probability is well known. To re-derive the latter, a g-method estimator combined with IS can be written as

$$\widehat{\alpha}_{\mathrm{NMF}} = \frac{1}{K} \sum_1^K e^{-\eta_o \sum_2^N V_i} \cdot W(V_2, \ldots, V_N), \quad V_i \sim f_*. \tag{5.11}$$

Scaling each V_i with a and using the resulting weighting function

$$W(v_2, \ldots, v_N) = a^{N-1}\, e^{-(1-1/a)\sum_2^N v_i}, \tag{5.12}$$

yields the estimator

$$\widehat{\alpha}_{\mathrm{NMF}} = \frac{1}{K} \sum_1^K a^{N-1}\, e^{-(\eta_o + 1 - 1/a)\sum_2^N V_i}, \quad V_i \sim f_*. \tag{5.13}$$

If the scaling factor is chosen as $a = 1/(1+\eta_o)$ in this estimator, then

$$\widehat{\alpha}_{\text{NMF}} = \frac{1}{K}\sum_{1}^{K}\left(\frac{1}{1+\eta_o}\right)^{N-1} = (1-\eta)^{N-1} = \alpha_{\text{NMF}}.$$

and therefore the variance of this IS unbiased estimator is zero.

A novel biasing scheme based on rotation of the primary data vector was developed for the estimation of the FAP of the NMF detector using IS. The derivation of the FAP using this technique and the corresponding simulation results are given in Appendix C.

5.3 The Adaptive NMF STAP detector

The NAMF detector is the adaptive version of the NMF detector. Here, the true data covariance matrix is replaced by an estimate and the detection statistic is given by

$$\frac{|\mathbf{s}^{\dagger}\widehat{\mathbf{R}}^{-1}\mathbf{x}|^2}{(\mathbf{s}^{\dagger}\widehat{\mathbf{R}}^{-1}\mathbf{s})(\mathbf{x}^{\dagger}\widehat{\mathbf{R}}^{-1}\mathbf{x})} \underset{H_0}{\overset{H_1}{\gtrless}} \eta. \tag{5.14}$$

We now proceed to investigate the FAP of the NAMF detector using IS.

5.3.1 FAP estimation using IS

To estimate the FAP of the NAMF detector, we first rewrite the test in (5.14) in a suitable form to apply the $g-$method estimator. To this end, we perform some random variables transformation following a similar approach as in [168].

Define the vectors $\mathbf{z}$ and $\mathbf{z}(l), l = 1, \cdots, L$, distributed as $\mathcal{CN}_N(0, \mathbf{I})$ and independent. The vectors $\mathbf{z}$ and $\mathbf{z}(l)$ can be decomposed as

$$\mathbf{z} = \begin{bmatrix}\mathbf{z}_A \\ \mathbf{z}_B\end{bmatrix} \quad \text{and} \quad \mathbf{z}(l) = \begin{bmatrix}\mathbf{z}_A(l) \\ \mathbf{z}_B(l)\end{bmatrix},$$

where the A components are scalar and B components $(N-1)$-vector. The covariance matrix of the vectors $\mathbf{z}(l)$ is given by $\frac{1}{L}\mathcal{S}$, where

$$\mathcal{S} = \sum_{l=1}^{L}\mathbf{z}(l)\mathbf{z}(l)^{\dagger} = \begin{bmatrix}\mathcal{S}_{AA} & \mathcal{S}_{AB} \\ \mathcal{S}_{BA} & \mathcal{S}_{BB}\end{bmatrix}. \tag{5.15}$$

Furthermore, define

$$\begin{aligned} y &\equiv \mathbf{z}_A - \mathcal{S}_{AB}\mathcal{S}_{BB}^{-1}\mathbf{z}_B, \\ y(l) &\equiv \mathbf{z}_A(l) - \mathcal{S}_{AB}\mathcal{S}_{BB}^{-1}\mathbf{z}_B(l), \end{aligned} \tag{5.16}$$

$$\Sigma_B \equiv \mathbf{z}_B^\dagger \left(\sum_{l=1}^{L} \mathbf{z}_B(l) \mathbf{z}_B(l)^\dagger \right)^{-1} \mathbf{z}_B, \tag{5.17}$$

and

$$M_B \equiv 1 + \Sigma_B. \tag{5.18}$$

Using the definitions above and the transformations given in Appendix D, the test in (5.14) can be rewritten as

$$\frac{u}{\sum_{l=1}^{L-N+1} u(l)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\eta}{LM_B} (\mathbf{x}^\dagger \widehat{\mathbf{R}}^{-1} \mathbf{x}), \tag{5.19}$$

where u and $u(l), l = 1, \cdots, L-N+1$ are all i.i.d. unit exponential RVs. In addition, the normalization term in the RHS of (5.19) becomes

$$\mathbf{x}^\dagger \widehat{\mathbf{R}}^{-1} \mathbf{x} = L\, \mathbf{z}^\dagger \mathcal{S}^{-1} \mathbf{z}. \tag{5.20}$$

From the discussion on pages 120 and 121 of [168], it turns out that the quantity $\mathbf{z}^\dagger \mathcal{S}^{-1} \mathbf{z}$, denoted therein as Σ, can be expressed as

$$\begin{aligned} \mathbf{z}^\dagger \mathcal{S}^{-1} \mathbf{z} &= \frac{|y|^2}{\sum_{l=1}^{L} |y(l)|^2} + \Sigma_B \\ &\overset{d}{\equiv} \frac{|y|^2}{\sum_{l=1}^{L-N+1} |w(l)|^2} + \Sigma_B \\ &= M_B \cdot \frac{u}{\sum_{l=1}^{L-N+1} u(l)} + \Sigma_B, \end{aligned} \tag{5.21}$$

where $\{w(l)\}_1^{L-N+1}$ are i.i.d. each with distribution $\mathcal{CN}_1(0,1)$. Combining (5.19), (5.20), and (5.21) yields the test in the form

$$u \underset{H_0}{\overset{H_1}{\gtrless}} \eta_o \frac{\Sigma_B}{1+\Sigma_B} \sum_{l=1}^{L-N+1} u(l), \tag{5.22}$$

where $\eta_o = \eta/(1-\eta)$. Conditioned on Σ_B (i.e., on the B-vectors $\mathbf{z}_B$ and $\mathbf{z}_B(l)$) the test in (5.22) is in the form of a conventional CA-CFAR test, thus resulting in the FAP

$$\alpha_{\text{NAMF}} = E\{g_N(\Sigma_B)\}, \tag{5.23}$$

where

$$g_N(\Sigma_B) = \frac{1}{[1+\eta_o\Sigma_B/(1+\Sigma_B)]^{L-N+1}}. \tag{5.24}$$

The g-method estimator using IS for the FAP of the NAMF can then be set up as

$$\widehat{\alpha}_{\mathrm{NAMF}} = \frac{1}{K}\sum_{1}^{K} g_N(\Sigma_B)\, W(\mathbf{z}_B, \mathbf{z}_{BL}); \quad \sim f_\star. \tag{5.25}$$

To implement IS, biasing of the B-vectors contained in Σ_B must produce an increase in the value of the g-function in (5.24). This means that Σ_B must be made to decrease, which can easily be accomplished by scaling down the primary B-vector $\mathbf{z}_B$ and scaling up the secondary B-vectors $\mathbf{z}_B(l)$. A 2-dimensional biasing scheme results, which needs to be optimized adaptively. The primary and secondary scaling parameters are chosen as $a^{1/2}$ and $\theta^{1/2}$ respectively, where $0 < a \leq 1$ and $\theta \geq 1$, and the weighting function is given by

$$\begin{aligned} W(\mathbf{z}_B, \mathbf{z}_{BL}) = a^{N-1}\theta^{L(N-1)} \exp\big(-\mathbf{z}_B^\dagger \mathbf{z}_B(1-1/a)\big) \\ \cdot \exp\Big(-(1-1/\theta)\sum_{l=1}^{L} \mathbf{z}_B(l)^\dagger \mathbf{z}_B(l)\Big). \end{aligned} \tag{5.26}$$

Figures 5.2 to 5.6 show the simulation results for the 2-dimensional biasing scheme using the g-method estimator for the NAMF detector. In these figures we used $L = 128$ secondary vectors, $N = 64$ space-time samples, and $K = 10000$ realizations.

Figure 5.2 show the IS gain surface using the g-method estimator with 2 dimensional biasing versus the biasing parameters. From this figure it can be seen that, for each value of FAP, there is an optimal combination of the biasing parameters that achieves the maximum IS gain. The optimal parameters can be estimated using the adaptive IS procedure described in Section 1.2.2. An example of the adaptive IS recursion for estimating the optimal biasing parameters a and θ is shown in Figure 5.3.

Figure 5.4 shows the IS gain obtained using the g-method estimator with 2 dimensional biasing using the estimated optimal biasing parameters.

In Figure 5.5 the estimated threshold multipliers needed to achieve a design FAP α are shown versus the recursions of the inverse IS method described in Section 1.2.2. Recall that the inverse IS method, that can be used for the detector design, can only be applied in combination with the $g-$method estimator.

Finally, in Figure 5.6 the estimated threshold multipliers obtained using the inverse IS procedure are compared to those obtained with the analytical expression available in [171]. We observe that the inverse IS procedure correctly estimates the multiplier values.

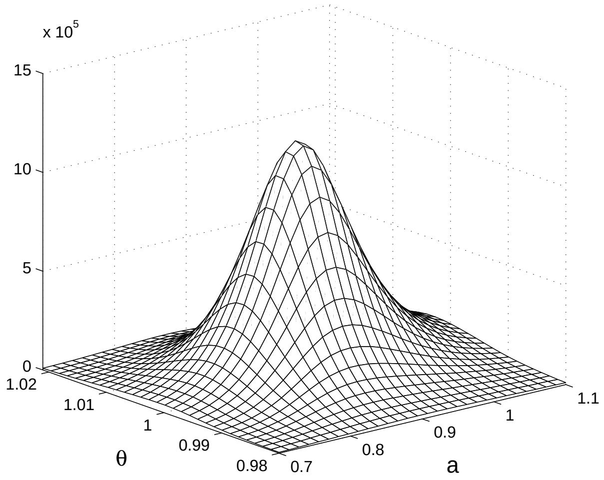
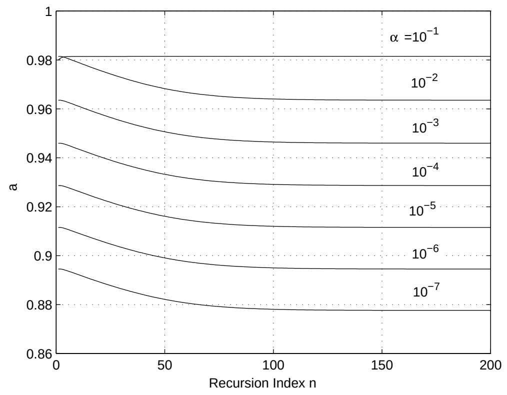


Figure 5.2: IS gain surface versus biasing parameters for $\alpha = 10^{-6}$ using the g-method with 2-dimensional biasing for the NAMF detector. $L = 128$, $N = 64$, and $K = 10000$.

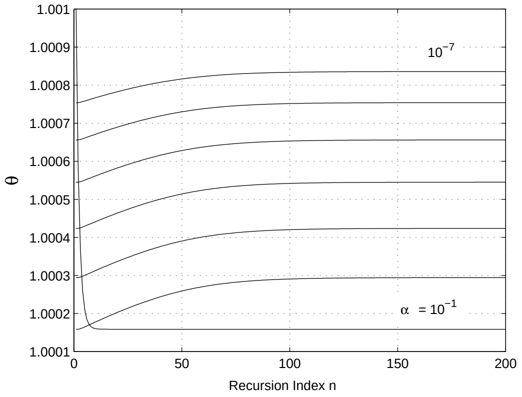
5.4 Envelope and Geometric Mean NAMF STAP detectors

In this section we propose two novel detectors for space-time adaptive processing. These are variants of the NAMF detector that use envelope-law and geometric mean (or logarithmic) processing, both being well-known concepts from conventional CFAR square-law radar detection [84–89, 106, 125–127, 172]. Linear-law CA-CFAR detectors using the amplitude of the received signal instead of the squared amplitude were studied in [127] and [84]. Later, it was also shown in [98] that mean-level CFAR processors including CA preceded by an envelope detector have more robust performance in the presence of Gaussian clutter power transitions and interfering targets as compared to the square-law detector. The geometric-mean (GM) variant of the CA-CFAR detector was already well known in the '70s and it is mostly known in the radar field as LOG/CFAR receiver, [85–89, 106, 172].

The so called LOG/CFAR detector was of great importance due to the capability of the receiver of operating over a larger dynamic range than a CA-CFAR receiver and to its capability of better suppressing Weibull distributed clutter and interfering targets with a reasonable CFAR loss. Later, a geometric-mean (GM) variant of the CA-CFAR detector was also considered in [93], where it was demonstrated that, in the scalar case (only 1 antenna element), the square-law CA-CFAR detector is better than the GM detector in homogeneous Gaussian background, but the GM detector



(a) Optimum scaling for primary B-vectors



(b) Optimum scaling for secondary B-vectors

Figure 5.3: Estimated optimum scaling parameters using adaptive IS versus recursion index using the g-method with 2-dimensional for the NAMF detector. $L = 128$, $N = 64$, and $K = 10000$.

performs better in the presence of interfering targets. Based on these observations, two STAP detector variants of the AMF using envelope-law and geometric mean processing were proposed in [96]. In this section we extend this idea and present two corresponding variants of the NAMF STAP detector [173], establish their CFAR property under the assumption of homogeneous Gaussian interference, and then carry out false alarm and detection probability performance estimation via IS.

The proposed envelope-law version of the NAMF detector, referred to as the E-

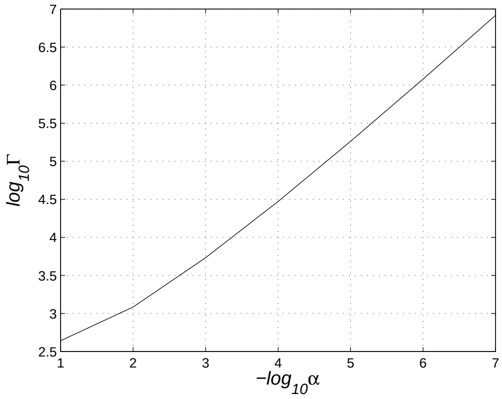


Figure 5.4: IS gain of the g-method with 2-dimensional biasing for the NAMF detector as function of FAP using optimum biasing parameters obtained from Figure 5.3. $L = 128$, $N = 64$, and $K = 10000$.

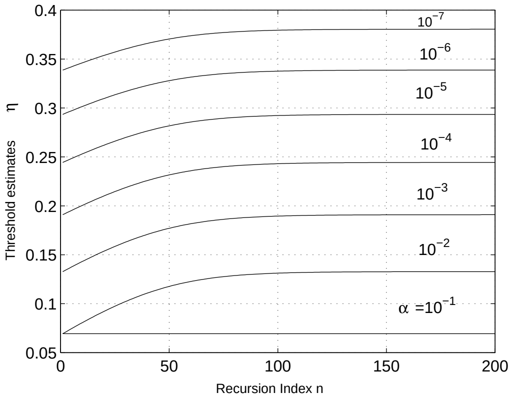


Figure 5.5: Threshold multipliers estimated via inverse IS using the g-method with 2-dimensional biasing for the NAMF detector. $L = 128$, $N = 64$, and $K = 10000$.

NAMF detector, [165, 173], is given by the test

$$\frac{|\mathbf{s}^\dagger \widehat{\mathbf{R}}^{-1}\mathbf{x}|}{\sqrt{\mathbf{x}^\dagger \widehat{\mathbf{R}}^{-1}\mathbf{x}}} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\eta_e}{L} \sum_{l=1}^{L} |\mathbf{s}^\dagger \widehat{\mathbf{R}}^{-1}\mathbf{x}(l)|, \tag{5.27}$$

where η_e denotes the detector threshold.

The geometric-mean version, referred to as the GM-NAMF detector, is defined as

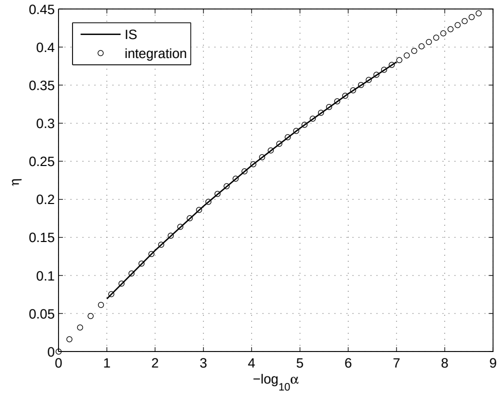


Figure 5.6: Comparison of inverse IS simulation and numerical integration [171] for determining the threshold multipliers for different FAP. $L = 128$, $N = 64$, and $K = 10000$.

$$\frac{|\mathbf{s}^\dagger \widehat{\mathbf{R}}^{-1}\mathbf{x}|}{\sqrt{\mathbf{x}^\dagger \widehat{\mathbf{R}}^{-1}\mathbf{x}}} \underset{H_0}{\overset{H_1}{\gtrless}} \eta_g \left(\prod_{l=1}^{L} |\mathbf{s}^\dagger \widehat{\mathbf{R}}^{-1}\mathbf{x}(l)| \right)^{1/L}, \tag{5.28}$$

where η_g is the threshold. The square-law version of the GM test in (5.28) is identical, being just a square of the above test expression.

5.4.1 Asymptotic thresholds

The asymptotic FAP behavior of the E-NAMF detector as $L \to \infty$ (or known covariance matrix $\mathbf{R}$) is obtained from

$$\begin{aligned} \alpha_{L\,\to\,\infty} &= P\left(\frac{|\mathbf{s}^\dagger \mathbf{R}^{-1}\mathbf{x}|}{\sqrt{\mathbf{x}^\dagger \mathbf{R}^{-1}\mathbf{x}}} \geq \eta_e E\{|\mathbf{s}^\dagger \mathbf{R}^{-1}\mathbf{x}(l)|\} \right) \\ &= P\left(\frac{|\mathbf{s}^\dagger \mathbf{R}^{-1}\mathbf{x}|^2}{\mathbf{x}^\dagger \mathbf{R}^{-1}\mathbf{x}} \geq \frac{\eta_e^2 \pi}{4}\, \mathbf{s}^\dagger \mathbf{R}^{-1}\mathbf{s} \right) \\ &= \left(1 - \frac{\pi}{4}\, \eta_e^2 \right)^{N-1}, \end{aligned} \tag{5.29}$$

the first line above following from convergence arguments and the law of large numbers applied to the independent and identically distributed sequence $\{\mathbf{s}^\dagger \mathbf{R}^{-1}\mathbf{x}(l)\}$, the second by noting that $\mathbf{s}^\dagger \mathbf{R}^{-1}\mathbf{x}(l)$ is distributed as $\mathcal{CN}_1(0, \mathbf{s}^\dagger \mathbf{R}^{-1}\mathbf{s})$, and the third from the fact that the second line represents the FAP $(1-\nu)^{N-1}$ of a normalized matched filter (NMF) STAP detector (the statistic of which can be obtained from the

LHS of (5.14) by replacing $\widehat{\mathbf{R}}$ with $\mathbf{R}$) with threshold ν ([171]) where $\nu = \eta_e^2\pi/4$. The asymptotic threshold can be calculated from the above expression.

Using similar arguments, the asymptotic FAP of the GM-NAMF detector is given by

$$\begin{aligned}\alpha_{L\,\rightarrow\,\infty} &= P\bigg(\log\frac{|\mathbf{s}^\dagger\mathbf{R}^{-1}\mathbf{x}|^2}{\mathbf{x}^\dagger\mathbf{R}^{-1}\mathbf{x}} \geq 2\log\eta_g + E\{\log|\mathbf{s}^\dagger\mathbf{R}^{-1}\mathbf{x}(l)|^2\}\bigg)\\ &= P\bigg(\log\frac{|\mathbf{s}^\dagger\mathbf{R}^{-1}\mathbf{x}|^2}{\mathbf{x}^\dagger\mathbf{R}^{-1}\mathbf{x}} \geq 2\log\eta_g - \gamma + \log(\mathbf{s}^\dagger\mathbf{R}^{-1}\mathbf{s})\bigg)\\ &= P\bigg(\frac{|\mathbf{s}^\dagger\mathbf{R}^{-1}\mathbf{x}|^2}{\mathbf{x}^\dagger\mathbf{R}^{-1}\mathbf{x}} \geq \eta_g^2\,e^{-\gamma}\,\mathbf{s}^\dagger\mathbf{R}^{-1}\mathbf{s}\bigg) = \big(1-\eta_g^2\,e^{-\gamma}\big)^{N-1} \end{aligned} \tag{5.30}$$

where $\gamma = 0.5772156...$ is the Euler-Mascheroni constant and in the second line we use the fact that the pdf of the logarithm of an exponential RV has a Generalized Extreme Value distribution. The CFAR property of both the E and GM NAMF detectors is proved in Appendix E

5.4.2 FAP estimation using IS

With some straightforward manipulations and using the definitions given in Section 5.3.1, the E-NAMF test in (5.27) can be rewritten as

$$|y|^2 \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\eta_e^2\Sigma_B}{LC_e}\left(\sum_{l=1}^{L}|y(l)|\right)^2, \tag{5.31}$$

where

$$C_e \equiv 1 - \frac{\eta_e^2}{L}\frac{\big(\sum_{l=1}^{L}|y(l)|\big)^2}{\sum_{l=1}^{L}|y(l)|^2}, \tag{5.32}$$

and the GM-NAMF in (5.28) as

$$|y|^2 \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\eta_g^2 L\Sigma_B}{C_g}\left(\prod_{l=1}^{L}|y(l)|^2\right)^{1/L}, \tag{5.33}$$

where

$$C_g \equiv 1 - L\eta_g^2\frac{\big(\prod_{l=1}^{L}|y(l)|^2\big)^{1/L}}{\sum_{l=1}^{L}|y(l)|^2}. \tag{5.34}$$

In writing (5.31) and (5.33) we have assumed that $C_e > 0$ and $C_g > 0$. This will be guaranteed if $\eta_e \leq 1$ and $\eta_g \leq 1$, respectively, and can be seen by an application of Hölder's inequality to the sums in (5.32) and using the fact that geometric means are smaller than arithmetic means in (5.34). As shown by simulation results, a wide range of FAPs is achieved for values of both thresholds less than unity and therefore we assume this restriction to hold. For brevity we denote the RHSs of both (5.31) and (5.33) by $\Sigma_B D$, where D is a function of $\begin{pmatrix} y(1) & \cdots & y(L) \end{pmatrix} \triangleq \mathbf{y}_L^T$ and corresponding constants that depend on the detector in question. With this (generic) notation, the FAP can be written as

$$\begin{aligned} \alpha &= P(|Y|^2 \geq \Sigma_B D) \\ &= E_B\{P(|Y|^2 \geq \Sigma_B D \,|\, \text{B-vec})\} \\ &= E_B\{E\{P(|Y|^2 \geq \Sigma_B D \,|\, \text{B-vec}, \mathbf{Y}_L)\}\} \\ &= E_B\{E\{e^{-\Sigma_B D/(1+\Sigma_B)} \,|\, \text{B-vec})\}\} \\ &= E_B\big\{E\big\{g(\Sigma_B, \mathbf{Y}_L)\Big|\, \text{B-vec}\big\}\big\}, \end{aligned} \tag{5.35}$$

where

$$g(\Sigma_B, \mathbf{Y}_L) \equiv e^{-\Sigma_B D/(1+\Sigma_B)}, \tag{5.36}$$

and E_B denotes expectation over the distribution of the B vectors and E the expectation over the conditional distribution of $\mathbf{Y}_L$. The fourth line above follows because conditioned on the B vectors, Y and $\mathbf{Y}_L$ are independent and $|Y|^2/(1+\Sigma_B)$ is unit exponential.

From (5.15) and (5.16) on page 90, y and $y(l)$ can be rewritten as

$$\begin{aligned} y &\equiv \mathbf{z}_A - \sum_{l=1}^{L} \mathbf{z}_A(l)\mathbf{z}_B(l)^\dagger\, \mathcal{S}_{BB}^{-1}\mathbf{z}_B, \\ y(l) &\equiv \mathbf{z}_A(l) - \sum_{i=1}^{L} \mathbf{z}_A(i)\mathbf{z}_B(i)^\dagger\, \mathcal{S}_{BB}^{-1}\mathbf{z}_B(l). \end{aligned} \tag{5.37}$$

Then the FAP in (5.35) can be further written as

$$\begin{aligned} \alpha &= E_B\big\{E\big\{g(\Sigma_B, \mathbf{Y}_L)\Big|\, \text{B-vec}\big\}\big\} = E\{g(\Sigma_B, Y(1), \ldots, Y(L))\} \\ &= E\{g(\Sigma_B, Z_A(1), \ldots, Z_A(L))\} \triangleq E\{g(\Sigma_B, \mathbf{Z}_{AL})\} \\ &= E_\star\{g(\Sigma_B, \mathbf{Z}_{AL})\, W(\mathbf{Z}_{AL})\} \end{aligned} \tag{5.38}$$

where $g(\Sigma_B, \mathbf{Z}_{AL}) = g(\Sigma_B, \mathbf{Y}_L)$ and is given by (5.36). A step by step derivation of the above equation is provided in Appendix F.

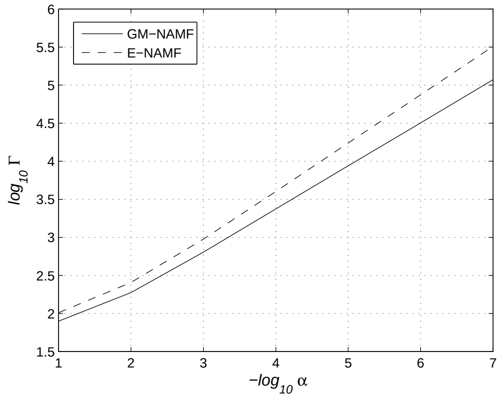


Figure 5.7: Estimated IS gains for the E- and GM-NAMF detectors. $L = 128$ and $N = 64$.

For simplicity, we perform IS only on the variables in $\mathbf{Z}_{AL} = (Z_A(1) \cdots Z_A(L))^T$. Noting that $\mathbf{Z}_{AL} \sim \mathcal{CN}_L(0, \mathbf{I})$, scaling down $\mathbf{Z}_{AL}$ with parameter θ leads to the weighting function

$$W(\mathbf{z}_{AL}) = \theta^L \, e^{-\sum_1^L z_A^\dagger(l) z_A(l)(1-1/\theta)}. \tag{5.39}$$

The IS estimator for the FAP is therefore

$$\widehat{\alpha} = \frac{1}{K}\sum_{i=1}^{K}[g(\Sigma_B, \mathbf{Z}_{AL})\, W(\mathbf{z}_{AL})]^{(i)}; \quad \mathbf{Z}_{AL} \sim f_\star(\mathbf{z}_{AL}), \text{B-vec} \sim f(\text{B-vec}) \tag{5.40}$$

5.4.3 Simulation results

In this section simulation results are shown for both E-NAMF and GM-NAMF detectors. The parameters used in the simulations are $L = 128$ and $N = 64$. In Figure 5.7 the estimated IS gain is plotted against the design FAP α. In Figure 5.8 the threshold multipliers for the two tests derived using the inverse IS procedure are shown versus the design FAP α.

P_D in homogeneous case

In this subsection we present the detection probability of the E-NAMF and GM-NAMF detectors in homogeneous i.i.d. Gaussian background and compare it with that of the NAMF detector. The FAP is set to 10^{-6}, $L = 128$ and $N = 64$. Detection probabilities are estimated using MC simulation with 100000 trials. The detection

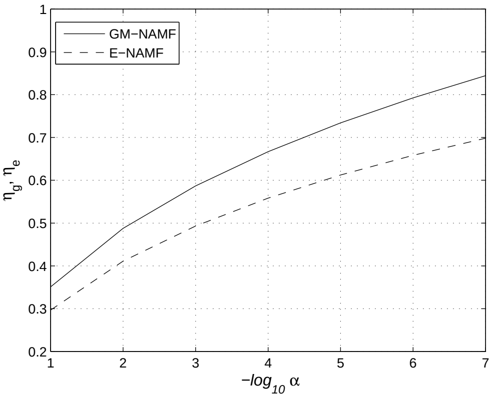


Figure 5.8: Threshold values for the E- and GM-NAMF detectors. $L = 128$ and $N = 64$.

probability for the case of both fluctuating and non-fluctuating target model is plotted in Figure 5.9 versus the SNR. In this figure is also plotted the detection probability for the NAMF detector obtained using the analytical formula of [171]. The figure shows that the three detectors perform almost the same in the presence of homogeneous clutter. In fact, the E-NAMF and NAMF perform exactly the same while the GM-NAMF has a very small loss compared to the other two detectors. Therefore, in homogeneous clutter background, regardless of the target fluctuation, both the E-NAMF detector and the GM-NAMF detector perform as well as the NAMF detector.

In Figure 5.10 the receiver operating characteristic is shown for a signal-to-noise ratio of 10 dB. The detection probability is plotted versus the FAP for E-NAMF and GM-NAMF for both Swerling 0 and Swerling 1 target models.

P_D in the presence of interfering targets

In this subsection the performance of the envelope, geometric mean and square law NAMF detectors are compared in the presence of two interfering targets in the secondary data. In the performance evaluations, correlated clutter and jammer have not been considered and the background consists of white Gaussian noise. In Figure 5.11 P_d performance are shown respectively for the Swerling 0 and Swerling 1 target model. In the simulations we assumed the interfering targets to have the same power and the same steering vector as the actual target. Figure 5.11(a) shows the comparison in the presence of 2 interferers when the target signal is constant, i.e., non-fluctuating. In Figure 5.11(b) is plotted the detection probability when both the target and the interferers are assumed to be fluctuating. For both target models, it can be seen that the NAMF detector, for a fixed P_d, has a significant detection loss compared to

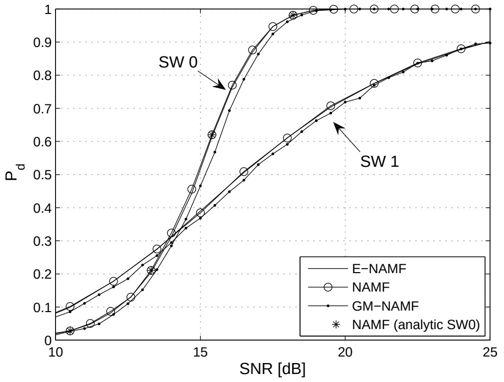


Figure 5.9: Detection probability versus SNR in homogeneous background for Swerling 0 and Swerling 1 target model. $L = 128$, $N = 64$ and $\alpha = 10^{-6}$.

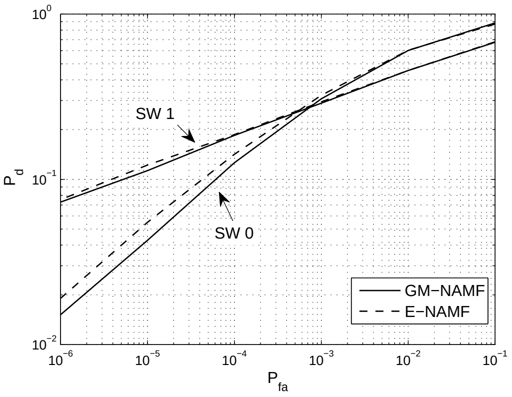
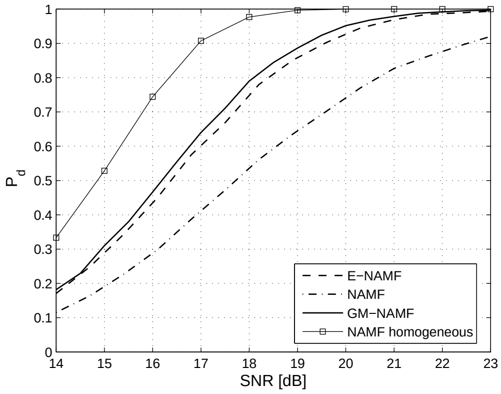
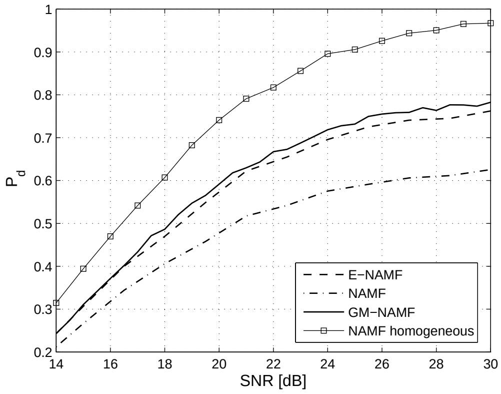


Figure 5.10: ROC for E-NAMF and GM-NAMF for both Swerling 0 and Swerling 1 target models. SNR=10 dB. $L = 128$ and $N = 64$.

the envelope and geometric mean variants and that the GM-NAMF performs slightly better than the E-NAMF detector. For example, at $P_d = 0.5$, the NAMF has a detection loss of about 1.2 dB w.r.t E-NAMF and 1.38 dB w.r.t GM-NAMF for the case of non-fluctuating target and about 2 dB w.r.t E-NAMF and 2.38 dB w.r.t GM-NAMF for the fluctuating model. This detection loss will increase for higher P_d. For example, in the case of Swerling 0 target model at $P_d = 0.9$, NAMF has about 3 dB loss w.r.t E-NAMF for the same case and 5.5 dB loss w.r.t NAMF in homogeneous Gaussian background. It has also to be noticed that in the fluctuating target scenario the presence of interfering targets produces a loss in terms of P_d performance much more significant than in the nonfluctuating case, especially in the region of high SNR.



(a) Swerling 0 target model



(b) Swerling 1 target model

Figure 5.11: P_d comparison for the NAMF, E-NAMF and GM-NAMF detectors with 2 interfering targets and NAMF detector without interfering targets (homogeneous). $\alpha = 10^{-6}$, $L = 128$ and $N = 64$.

As expected, in the presence of nonhomogeneities the E-NAMF and GM-NAMF detectors are more robust than the NAMF square low detector. Furthermore, they also perform better than the AMF detector and its envelope and geometric mean variants, whose performance are reported in [96]. This result is in agreement with previous works on LOG/CFAR processors, [85–89, 172] and reference therein, and due to the capability of the logarithmic receiver to suppress non Gaussian clutter and interfering targets.

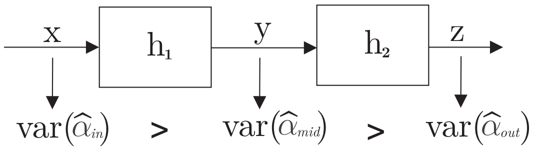


Figure 5.12: Variance of IS estimators when biasing is applied at different points in the processing chain.

5.5 Point of application of biasing

Some comments of a general nature regarding application of IS to signal processing algorithms are made here. They give some insight into the thinking behind our efforts to simulate the above mentioned algorithms. The simulation procedures described thus far in the previous sections, such as adaptive IS, are applicable to any detector. When IS biasing is performed on the input random variables then we refer to this technique as *input biasing.* Comparing the results for the different detectors presented in this thesis, it is clear that accuracy of the IS estimates and resulting simulation gains will of course depend on the particular detection algorithm under study. It is generally true (and intuitive) that better estimator performance can be obtained if IS biasing can be carried out closer to the point in the processing chain of the detector where the actual (rare event) decisionmaking is done, as shown in Figure 5.12. A formal proof of this fact can be found in [92]. This of course necessitates knowledge of density functions of the processes at the point where biasing is to be implemented. Often, input stochastic variables may have undergone transformations whose results are difficult to characterize statistically in analytical form, and we have to rely on the general method above. However, when such transformations can be characterized, then IS should be carried out using the modified processes. Therefore, it is desirable to perform biasing and IS as close to the final decisionmaking point as permitted by availability of knowledge of probability density functions. Another approach to biasing that is sometimes possible is to perform a series of (linear and/or nonlinear) transformations of the input processes as if one were carrying out a mathematical analysis of the algorithm. The transformations are carried out until the point beyond which it may not be possible to determine the density functions of the transformed processes without considerable mathematical effort. Biasing is then performed at this stage, the hope being that the g-method becomes applicable. This procedure may produce higher simulation gains than simple input biasing which, of course, is the easiest to implement. This point is illustrated in Figure 5.13 by comparing the gains obtained using the 2 different biasing schemes proposed in this thesis, namely input biasing for NMF (Appendix C) and g-method estimator with biasing of Σ_B for the NAMF detector (Section 5.3). Another example of this phenomenon is shown in

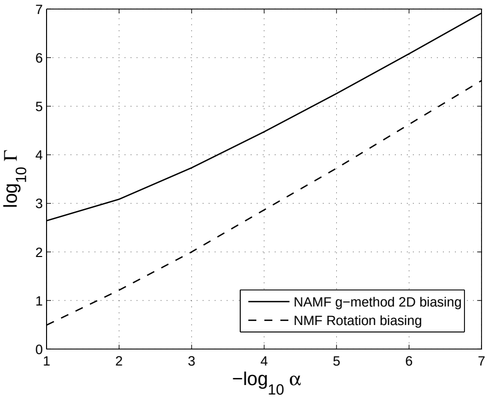


Figure 5.13: IS gain comparison for NMF using (input) biasing by rotation and NAMF using the g-method with 2 dimensional biasing. $N = 64$ and $L = 128$.

Appendix D, Figure D.3, by comparing the simulation gains of different IS schemes using one dimensional ([96]) and two dimensional biasing ([174]) applied to the AMF.

5.6 Conclusions

In this chapter we have introduced and developed IS techniques suitable for analysis of the NMF class of STAP detectors. Two dimensional biasing schemes using the g-method estimator have been used to produce new IS results for the NAMF detector and its variants, the envelope-law and geometric-mean. The E and GM NAMF detectors have been proposed here for the first time, and their respective thresholds have been determined using fast simulations. We have shown that these detectors have better performance in the presence of interfering targets in the training data while maintaining almost equal performance as the standard square-law version in homogeneous Gaussian interference. These detectors represent robust alternatives to conventional square-law processing.

Chapter 6

Fast simulations for Low Rank STAP detectors

In this chapter the STAP detector based on the low-rank approximation of the normalized adaptive matched filter (LRNAMF) is investigated for its performance. Being computationally efficient, the LRNAMF detector is a candidate for future implementation in STAP radar detection. Subject to an approximation for the disturbance covariance matrix in a clutter dominated scenario, the FAP of the LRNMF detector is known via a simple formula, [102, 103]. A brief description of the detector is reported in the first section in order to summarize the low-rank approximation. We then provide an analytical derivation, based on the g-method, of the exact FAP of the LRNMF detector for data possessing an arbitrary covariance matrix. The FAP derivation based on the g-method is obtained using both singular and non-singular Gaussian distributions. Using a nominal model for the eigenspectrum of the covariance matrix that is used to set up our simulations, we compare the results of the $g-$method estimator with the exact formula of the LRNMF FAP. We also discuss the adaptive LRNMF, the LRNAMF detector and show how the analytical results for the LRNMF detector can be used to predict its performance. Finally, we evaluate the detection performance of the adaptive detector when the secondary data are contaminated by interfering targets, and demonstrate that LR detectors are robust against outliers in the secondary data vectors.[1]

---

[1] This chapter is based on articles [**C19**] and [**R1**] (a list of the author's publications is included at the end of this dissertation, p. 183.)

6.1 The LRNMF detector

The low-rank approximation to the normalized matched filter is used for target detection in heterogeneous clutter scenarios. The background is assumed to consist of clutter plus white Gaussian noise. The binary hypothesis test can be written as

$$\begin{aligned} \mathrm{H}_0 : \mathbf{x} &= \mathbf{d} = \mathbf{c} + \mathbf{n} \\ \mathrm{H}_1 : \mathbf{x} &= a\mathbf{s} + \mathbf{d} = a\mathbf{s} + \mathbf{c} + \mathbf{n} \end{aligned} \tag{6.1}$$

where $\mathbf{x}$ is the primary data vector, $\mathbf{c}$ is the Gaussian clutter vector with covariance matrix $s\mathbf{R}_c$ with unknown level s and known structure, $\mathbf{n}$ denotes the additive white Gaussian noise vector with covariance matrix $\sigma^2\mathbf{I}$, where $\mathbf{I}$ is the $N \times N$ identity matrix and the noise power σ^2 is unknown, $\mathbf{s}$ is the steering vector and a is the unknown target amplitude. The vector $\mathbf{d}$ is used to represent the sum of the clutter and the white Gaussian noise. The covariance matrix of the disturbance $\mathbf{d}$ is

$$\mathbf{R}_d = s\mathbf{R}_c + \sigma^2\mathbf{I}. \tag{6.2}$$

In many real cases the clutter covariance matrix $\mathbf{R}_c$ has rank r which is less than N. This fact will be used to approximate the inverse of the disturbance covariance matrix which will be used in the NMF test. The matrix $\mathbf{R}_d$ can be expressed as

$$\mathbf{R}_d = \mathbf{U}\mathbf{D}\mathbf{U}^\dagger, \tag{6.3}$$

where $\mathbf{U}$ is the matrix whose columns are the normalized eigenvectors of $\mathbf{R}_d$ and $\mathbf{D}$ is the diagonal matrix of the eigenvalues of $\mathbf{R}_d$. When $\mathbf{R}_c$ has rank $r \ll N$, then $\mathbf{R}_d$ can be rewritten as

$$\mathbf{R}_d = \sum_{i=1}^{r}(s\lambda_i + \sigma^2)\mathbf{u}_i\mathbf{u}_i^\dagger + \sum_{i=r+1}^{N} \sigma^2\mathbf{u}_i\mathbf{u}_i^\dagger. \tag{6.4}$$

The inverse covariance matrix is

$$\mathbf{R}_d^{-1} = \sum_{i=1}^{r}(s\lambda_i + \sigma^2)^{-1}\mathbf{u}_i\mathbf{u}_i^\dagger + \sum_{i=r+1}^{N} \sigma^{-2}\mathbf{u}_i\mathbf{u}_i^\dagger. \tag{6.5}$$

The previous expression can be rewritten as

$$\begin{aligned}\mathbf{R}_d^{-1} &= \frac{1}{\sigma^2}\sum_{i=1}^{r}\frac{\mathbf{u}_i\mathbf{u}_i^\dagger}{(1+\frac{s\lambda_i}{\sigma^2})} + \sum_{i=r+1}^{N}\sigma^{-2}\mathbf{u}_i\mathbf{u}_i^\dagger \\ &= \frac{1}{\sigma^2}\sum_{i=1}^{N}\mathbf{u}_i\mathbf{u}_i^\dagger - \frac{1}{\sigma^2}\sum_{i=1}^{r}\frac{\frac{s\lambda_i}{\sigma^2}}{(1+\frac{s\lambda_i}{\sigma^2})}\mathbf{u}_i\mathbf{u}_i^\dagger \\ &= \frac{1}{\sigma^2}\Big(\mathbf{I}-\sum_{i=1}^{r}\frac{\frac{s\lambda_i}{\sigma^2}}{(1+\frac{s\lambda_i}{\sigma^2})}\mathbf{u}_i\mathbf{u}_i^\dagger\Big). \end{aligned} \tag{6.6}$$

For a clutter-to-noise ratio (CNR) much greater than one, i.e., $s\lambda_i \gg \sigma^2$, the inverse of the disturbance covariance matrix can be approximated as [175]

$$\mathbf{R}_d^{-1} \approx \frac{1}{\sigma^2}(\mathbf{I}-\mathbf{P}), \tag{6.7}$$

where

$$\mathbf{P} = \sum_{i=1}^{r}\mathbf{u}_i\mathbf{u}_i^\dagger \tag{6.8}$$

is a rank r projection matrix formed with the r eigenvectors corresponding to the r dominant eigenvalues of $\mathbf{R}_d$. Using the approximation in (6.7) in the NMF test of (5.5), we obtain the LRNMF test

$$\Lambda_{\mathrm{LR}} \equiv \frac{|\mathbf{s}^\dagger(\mathbf{I}-\mathbf{P})\mathbf{x}|^2}{(\mathbf{s}^\dagger(\mathbf{I}-\mathbf{P})\mathbf{s})(\mathbf{x}^\dagger(\mathbf{I}-\mathbf{P})\mathbf{x})} \underset{H_0}{\overset{H_1}{\gtrless}} \eta. \tag{6.9}$$

It can be noticed that the LRNMF test is invariant to the unknown clutter level s and to the noise power σ^2. Moreover, since $(\mathbf{I}-\mathbf{P})$ is also a projection matrix of rank $(N-r)$, $(\mathbf{I}-\mathbf{P})^2 = (\mathbf{I}-\mathbf{P})$, therefore we can define the transformed vectors $\mathbf{s}_1 = (\mathbf{I}-\mathbf{P})\mathbf{s}$, $\mathbf{x}_1 = (\mathbf{I}-\mathbf{P})\mathbf{x}$ and rewrite the test as

$$\Lambda_{\mathrm{LR}} \equiv \frac{|\mathbf{s}_1^\dagger\mathbf{x}_1|^2}{(\mathbf{s}_1^\dagger\mathbf{s}_1)(\mathbf{x}_1^\dagger\mathbf{x}_1)} \underset{H_0}{\overset{H_1}{\gtrless}} \eta. \tag{6.10}$$

It is important to observe that the low rank NMF test is the squared cosine of the angle between the transformed steering vector $\mathbf{s}_1$ and the transformed data vector $\mathbf{x}_1$. Fast simulation can be performed using rotation of the primary data vector $\mathbf{x}_1$ as described in Appendix C for the NMF detector; of course, since the two vectors in this case are not the same as for the NMF case, the IS performance will be different.

6.1.1 FAP approximation: low clutter rank and high CNR

The FAP for the LRNMF detector is derived in [103] and is given by

$$\alpha_{\text{LRNMF}} = (1-\eta)^{N-r-1}. \tag{6.11}$$

A crucial point of the LRNMF detector is the clutter rank estimation. The threshold for the LRNMF test, for a fixed FAP, will depend on the rank r of the clutter covariance matrix, specifically it increases with increasing r. A technique for clutter rank estimation can be found in [176]. The case $r = 0$ coincides with the full rank NMF test, which is invariant to the white noise level.

6.2 Exact FAP of the LRNMF detector using the $g-$method

In this section we derive two alternate expressions for the exact FAP of the LRNMF detector in (6.9). In particular, we do not assume that $(\mathbf{I} - \mathbf{P})$ in (6.7) whitens the primary vector $\mathbf{x}$, as is required for the derivation of the FAP formula in (6.11). Therefore, these exact forms are valid for any primary covariance matrix $\mathbf{R}_d$. The first expression involves singular multivariate Gaussian distributions whereas the second does not.

For ease of notation we denote the projection matrix $(\mathbf{I} - \mathbf{P})$ by $\mathbf{Q}$, which is idempotent. The LRNMF test of (6.9) can then be written as

$$\Lambda_{\text{LR}} \equiv \frac{|\mathbf{s}^\dagger \mathbf{Q}^2 \mathbf{x}|^2}{(\mathbf{s}^\dagger \mathbf{Q}^2 \mathbf{s})(\mathbf{x}^\dagger \mathbf{Q}^2 \mathbf{x})} \underset{H_0}{\overset{H_1}{\gtrless}} \eta. \tag{6.12}$$

Define the transformed vectors

$$\mathbf{x}_1 = \mathbf{Q}\mathbf{x} \quad \text{and} \quad \mathbf{s}_1 = \mathbf{Q}\mathbf{s}. \tag{6.13}$$

Then

$$\mathbf{R}_1 = E\{\mathbf{X}_1 \mathbf{X}_1^\dagger\} = \mathbf{Q} E\{\mathbf{X}\mathbf{X}^\dagger\}\mathbf{Q}^\dagger = \mathbf{Q}\mathbf{R}_d\mathbf{Q}^\dagger. \tag{6.14}$$

As $\mathbf{Q}$ is a singular matrix, the matrix $\mathbf{R}_1$ is also not of full rank. This follows from the property of the rank of a product of matrices.[2] Hence $\mathbf{x}_1 \sim \mathcal{SCN}_N(0, \mathbf{R}_1)$, where $\mathcal{SCN}$ indicates a singular multivariate Gaussian distribution.[3] Using the above transformations the test becomes

$$\Lambda_{\text{LR}} \equiv \frac{|\mathbf{s}_1^\dagger \mathbf{x}_1|^2}{(\mathbf{s}_1^\dagger \mathbf{s}_1)(\mathbf{x}_1^\dagger \mathbf{x}_1)} = \frac{|\mathbf{s}_1^\dagger \mathbf{x}_1|^2}{\|\mathbf{s}_1\|^2 \|\mathbf{x}_1\|^2} = \frac{|(\mathbf{s}_1^\dagger / \|\mathbf{s}_1\|)\ \mathbf{x}_1|^2}{\|\mathbf{x}_1\|^2} = \frac{|\mathbf{t}^\dagger \mathbf{x}_1|^2}{\|\mathbf{x}_1\|^2} \tag{6.15}$$

where $\mathbf{t} \equiv \mathbf{s}_1 / \|\mathbf{s}_1\|$ is a unit vector.

---

[2] $\text{rank}(\mathbf{AB}) \leq \min(\,\text{rank}(\mathbf{A}), \text{rank}(\mathbf{B})\,)$

[3] A complex Gaussian vector $\mathbf{x}$ of N components with covariance matrix $\mathbf{R}$ and mean vector μ is said to have a singular multivariate Gaussian distribution if $\text{rank}(\mathbf{R}) = p < N$. Using the

6.2.1 Exact FAP: using singular Gaussian distributions

Define a unitary transformation $\mathbf{t}_1 = \mathbf{Ht}$, such that the unit vector $\mathbf{t}_1$ has a single element equal to 1 and the remaining $(N-1)$ elements are zero, i.e., $\mathbf{t}_1 = (0, \ldots, 0, 1)^T$. The matrix $\mathbf{H}$ can be an Householder transformation matrix given by

$$\mathbf{H} = \mathbf{I} - \frac{2\mathbf{u}\mathbf{u}^\dagger}{\|\mathbf{u}\|^2}, \tag{6.16}$$

and

$$\mathbf{u} = \mathbf{t} + \frac{t_i\|\mathbf{t}\|}{|t_i|}\mathbf{e}_i, \tag{6.17}$$

where

$$\mathbf{e}_i = [0, \ldots, 0, 1, 0, \ldots, 0]^T \tag{6.18}$$

has 1 as the ith element and zero elsewhere. The transformation by the matrix $\mathbf{H}$ will eliminate all the elements of $\mathbf{t}$ except the element t_i. Note that $\mathbf{H}$ is an Hermitian unitary matrix

$$\mathbf{H}^{-1} = \mathbf{H}^\dagger = \mathbf{H}. \tag{6.19}$$

We will construct the $\mathbf{H}$ matrix in order to preserve only the Nth element of the vector $\mathbf{t}$. The likelihood ratio in (6.15) now takes the form

$$\Lambda_{\mathrm{LR}} = \frac{|\mathbf{t}_1^\dagger \mathbf{H}\mathbf{x}_1|}{\mathbf{x}_1^\dagger \mathbf{x}_1}. \tag{6.20}$$

Now define

$$\begin{aligned} \mathbf{y} &= \mathbf{H}\mathbf{x}_1, \\ \mathbf{R}_y &= E\{\mathbf{y}\mathbf{y}^\dagger\} = \mathbf{H}E\{\mathbf{x}_1\mathbf{x}_1^\dagger\}\mathbf{H}^\dagger = \mathbf{H}\mathbf{R}_1\mathbf{H}^\dagger, \end{aligned} \tag{6.21}$$

---

factorization

$$\mathbf{R} = \mathbf{U}\begin{bmatrix} \mathbf{D}_\lambda & 0 \\ 0 & 0 \end{bmatrix}\mathbf{U}^\dagger$$

the covariance matrix can be written as $\mathbf{R} = \mathbf{U}_1\mathbf{D}_\lambda\mathbf{U}_1^\dagger$, where $\mathbf{U} = [\mathbf{U}_1, \mathbf{U}_2]$ is an $N \times N$ orthogonal matrix, $\mathbf{U}_1$ is an $N \times p$ column orthogonal matrix, and $\mathbf{D}_\lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_p)$ with $\lambda_1 > 0$ for $i = 1, \ldots, p$. Define $\mathbf{R}^- \equiv \mathbf{U}_1\mathbf{D}_\lambda^{-1}\mathbf{U}_1^\dagger$, the generalized inverse of $\mathbf{R} = \mathbf{U}_1\mathbf{D}_\lambda\mathbf{U}_1^\dagger$. The p.d.f. of $\mathbf{x}$ is given by

$$f(\mathbf{x}) = \frac{1}{\pi^N|\mathbf{D}_\lambda|}\exp\{(\mathbf{x}-\mu)^\dagger\mathbf{R}^-(\mathbf{x}-\mu)\}$$

where $\mathbf{x}$ lies on the p-dimensional linear subspace defined by

$$\mathbf{L}^\dagger(\mathbf{x}-\mu) = 0;\ \mathbf{L} : N \times (N-p),\ \mathbf{L}^\dagger\mathbf{R} = 0,\ \mathbf{L}^\dagger\mathbf{L} = \mathbf{I}_{N-p}$$

where $\mathbf{y} \sim \mathcal{SCN}_N(0, \mathbf{R}_y)$. The vector $\mathbf{x}_1$ in the test can be replaced by

$$\mathbf{x}_1 = \mathbf{H}^{-1}\mathbf{y} = \mathbf{H}^{\dagger}\mathbf{y} = \mathbf{H}\mathbf{y}. \tag{6.22}$$

Then, the likelihood ratio of (6.20) and corresponding test can be rewritten as

$$\Lambda_{\mathrm{LR}} = \frac{|\mathbf{t}_1^{\dagger}\mathbf{y}|}{\mathbf{y}^{\dagger}\mathbf{H}^{\dagger}\mathbf{H}\mathbf{y}} = \frac{|\mathbf{t}_1^{\dagger}\mathbf{y}|}{\mathbf{y}^{\dagger}\mathbf{y}} = \frac{|y_N|^2}{\sum\limits_{i=1}^{N} |y_i|^2} \underset{H_0}{\overset{H_1}{\gtrless}} \eta \tag{6.23}$$

where y_i, $i = 1, \ldots, N$ are the elements of the vector $\mathbf{y}$. Now we can write the test in the form

$$\begin{aligned} |y_N|^2 \quad &\underset{H_0}{\overset{H_1}{\gtrless}} \quad \eta\Big(|y_N|^2 + \sum_{i=1}^{N-1} |y_i|^2\Big) \\ |y_N|^2(1-\eta) \quad &\underset{H_0}{\overset{H_1}{\gtrless}} \quad \eta \sum_{i=1}^{N-1} |y_i|^2 \\ |y_N|^2 \quad &\underset{H_0}{\overset{H_1}{\gtrless}} \quad \eta_0 \sum_{i=1}^{N-1} |y_i|^2 \\ |y_N| \quad &\underset{H_0}{\overset{H_1}{\gtrless}} \quad \Big(\eta_0 \sum_{i=1}^{N-1} |y_i|^2\Big)^{1/2} \end{aligned} \tag{6.24}$$

where $\eta_0 = \eta/1 - \eta$.

Define the $(N-1)$ vector $\tilde{\mathbf{y}} \equiv (y_1, \ldots, y_{N-1})^T$, $u_1 \equiv |y_N|$, and

$$b(\eta, \tilde{\mathbf{y}}) \equiv \Big(\eta_0 \sum_{i=1}^{N-1} |y_i|^2\Big)^{1/2}. \tag{6.25}$$

Then, the FAP is given by

$$\begin{aligned} \alpha_{\mathrm{LRNMF}} \quad &= \quad P\Big(|Y_N| \geq \ \Big(\eta_0 \sum_{i=1}^{N-1} |Y_i|^2\Big)^{1/2}\Big) \\ &= \quad E\Big\{P\Big(U_1 \geq \ b(\eta, \tilde{\mathbf{Y}})\big|\tilde{\mathbf{Y}}\Big)\Big\} \triangleq E\{g(\eta, \tilde{\mathbf{Y}})\} \end{aligned} \tag{6.26}$$

and the g function can be expressed in integral form as

$$g(\eta, \tilde{\mathbf{y}}) = \int_{b(\eta, \tilde{\mathbf{y}})}^{\infty} f(u_1|\tilde{\mathbf{y}}) du_1. \tag{6.27}$$

The conditional distribution of Y_N is also Gaussian and we have to determine the conditional mean and variance. The vector $\mathbf{y}$ can be written as $\mathbf{y} = [\tilde{\mathbf{y}}; y_N]$ and the covariance matrix of $\mathbf{y}$ can be partitioned as

$$\mathbf{R}_y = \begin{bmatrix} E\{\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^\dagger\} & E\{\tilde{\mathbf{Y}}Y_N^\dagger\} \\ E\{Y_N\tilde{\mathbf{Y}}^\dagger\} & E\{Y_N Y_N^\dagger\} \end{bmatrix} = \begin{bmatrix} \mathbf{R}_{\tilde{\mathbf{y}}} & \mathbf{R}_{\tilde{\mathbf{y}}y_N} \\ \mathbf{R}_{y_N\tilde{\mathbf{y}}} & \sigma^2_{y_N} \end{bmatrix} \tag{6.28}$$

where $\mathbf{R}_{\tilde{\mathbf{y}}}$ is the $(N-1)\times(N-1)$ singular covariance matrix of the vector $\tilde{\mathbf{y}}$ and $\sigma^2_{y_N}$ is the variance of the random variable Y_N. The conditional mean and variance of Y_N are

$$\begin{aligned} E\{Y_N|\tilde{\mathbf{y}}\} &= \mathbf{R}_{y_N\tilde{\mathbf{y}}}\mathbf{R}^-_{\tilde{\mathbf{y}}}\tilde{\mathbf{y}} = \mathbf{c}^T\tilde{\mathbf{y}} = \sum_{i=1}^{N-1} c_i\, y_i, \\ \operatorname{var}\{Y_N|\tilde{\mathbf{y}}\} &= \sigma^2_{y_N} - \mathbf{R}_{y_N\tilde{\mathbf{y}}}\mathbf{R}^-_{\tilde{\mathbf{y}}}\mathbf{R}_{\tilde{\mathbf{y}}y_N}, \end{aligned} \tag{6.29}$$

where $\mathbf{c}^T \equiv \mathbf{R}_{y_N\tilde{\mathbf{y}}}\mathbf{R}^-_{\tilde{\mathbf{y}}}$, $c_i, i = 1,\ldots,N-1$ are the elements of the vector $\mathbf{c}$, and $\mathbf{R}^-_{\tilde{\mathbf{y}}}$ is the g-inverse of $\mathbf{R}_{\tilde{\mathbf{y}}}$.

Conditioned on the vector $\tilde{\mathbf{Y}}$, the random variable Y_N is Gaussian with mean and variance given in the expressions above. Then the random variable $U_1 = |Y_N|$ has a Rice distribution with noncentrality parameter

$$s^2 = \Big|\sum_{i=1}^{N-1} c_i\, y_i\Big|^2, \tag{6.30}$$

and parameter

$$\sigma^2 = \operatorname{var}\{Y_N|\tilde{\mathbf{y}}\}/2. \tag{6.31}$$

Hence, the integral in (6.27) can be written as

$$\begin{aligned} g(\eta,\tilde{\mathbf{y}}) &= \int_{b(\eta,\tilde{\mathbf{y}})}^{\infty} \frac{u_1}{\sigma^2}\exp\big[-(u_1^2+s^2)/2\sigma^2\big] I_0(u_1 s/\sigma^2)\, du_1 \\ &= Q_1\left(\frac{s}{\sigma},\frac{b(\eta,\tilde{\mathbf{y}})}{\sigma}\right), \end{aligned} \tag{6.32}$$

where $Q_1(\cdot,\cdot)$ is the Marcum-q function, and the FAP becomes

$$\alpha_{\mathrm{LRNMF}} = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} Q_1\left(\frac{s}{\sigma},\frac{b(\eta,\tilde{\mathbf{y}})}{\sigma}\right) f(\tilde{\mathbf{y}})d\tilde{\mathbf{y}} \tag{6.33}$$

where $f(\tilde{\mathbf{y}})$ is the density corresponding to the singular Gaussian distribution $\mathcal{SCN}_{N-1}(0,\mathbf{R}_{\tilde{\mathbf{y}}})$. This is one expression for the FAP of the LRNMF detector for arbitrary covariance matrix $\mathbf{R}_d$ of the primary data vector.

6.2.2 Exact FAP: using nonsingular Gaussian distributions

We start with the covariance matrix $\mathbf{R}_1$ of $\mathbf{x}_1 \sim \mathcal{SCN}_N(0, \mathbf{R}_1)$ having some rank say $p < N$ in (6.14) and write its decomposition as

$$\mathbf{R}_1 = \mathbf{Q}\mathbf{R}_d\mathbf{Q}^\dagger = \mathbf{V}\Lambda\mathbf{V}^\dagger \tag{6.34}$$

where $\mathbf{V}$ is the matrix of eigenvectors corresponding to the eigenvalues $\{\lambda_i\}_1^N$ in the diagonal matrix Λ, which can be written as

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & . & . & . & 0 \\ 0 & . & . & . & . & . \\ . & . & \lambda_p & . & . & . \\ . & . & . & 0 & . & . \\ . & . & . & . & . & 0 \\ 0 & . & . & . & 0 & 0 \end{pmatrix}. \tag{6.35}$$

Define

$$\Lambda_p \equiv \begin{pmatrix} \lambda_1 & 0 & . & . & 0 \\ 0 & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & 0 \\ 0 & . & . & 0 & \lambda_p \end{pmatrix}, \tag{6.36}$$

and

$$\begin{aligned} \mathbf{V}_p &\equiv \text{upper left block of } \mathbf{V} \in R^{p\times p}, \\ \mathbf{V}_{N-p} &\equiv \text{lower left block of } \mathbf{V} \in R^{(N-p)\times p}. \end{aligned} \tag{6.37}$$

Using these definitions we can write $\mathbf{R}_1$ as

$$\mathbf{R}_1 = \begin{pmatrix} \mathbf{V}_p \Lambda_p \mathbf{V}_p^\dagger & \mathbf{V}_p \Lambda_p \mathbf{V}_{N-p}^\dagger \\ \mathbf{V}_{N-p} \Lambda_p \mathbf{V}_p^\dagger & \mathbf{V}_{N-p} \Lambda_p \mathbf{V}_{N-p}^\dagger \end{pmatrix} \tag{6.38}$$

which also turns out to be the covariance matrix of a new vector $\mathbf{x}_1$ defined as

$$\mathbf{x}_1 \equiv \underbrace{\begin{bmatrix} \mathbf{V}_p \\ \mathbf{V}_{N-p} \end{bmatrix}}_{N\times p} \Lambda_p^{1/2}\, \mathbf{W}; \quad \mathbf{W} \sim \mathcal{CN}_p(0, \mathbf{I}) \tag{6.39}$$

and hence this new $\mathbf{x}_1$ is statistically identical to the vector $\mathbf{x}_1$ in the beginning of this subsection. This is a well known representation [177] for expressing a singular Gaussian vector in terms of a nonsingular Gaussian vector with independent components.

Substituting this definition into (6.15) yields the likelihood ratio

$$\begin{aligned} \mathit{\Lambda}_{\mathrm{LR}} &= \frac{|\mathbf{t}^{\dagger}\mathbf{x}_1|^2}{\|\mathbf{x}_1\|^2} \\ &= \frac{\left| \mathbf{t}^{\dagger} \begin{bmatrix} \mathbf{V}_p \\ \mathbf{V}_{N-p} \end{bmatrix} \Lambda_p^{1/2}\, \mathbf{W} \right|^2}{\mathbf{W}^{\dagger}\Lambda_p^{1/2} \begin{bmatrix} \mathbf{V}_p^{\dagger} & \mathbf{V}_{N-p}^{\dagger} \end{bmatrix} \begin{bmatrix} \mathbf{V}_p \\ \mathbf{V}_{N-p} \end{bmatrix} \Lambda_p^{1/2}\, \mathbf{W}} \\ &= \frac{\left| \mathbf{t}^{\dagger} \begin{bmatrix} \mathbf{V}_p \\ \mathbf{V}_{N-p} \end{bmatrix} \Lambda_p^{1/2}\, \mathbf{W} \right|^2}{\mathbf{W}^{\dagger}\Lambda_p\mathbf{W}}. \end{aligned} \tag{6.40}$$

Now define

$$\mathbf{Y} \equiv \Lambda_p^{1/2}\mathbf{W} \tag{6.41}$$

and the $p \times 1$ unit vector

$$\mathbf{t}_1 \equiv \begin{bmatrix} \mathbf{V}_p^{\dagger} & \mathbf{V}_{N-p}^{\dagger} \end{bmatrix} \mathbf{t}. \tag{6.42}$$

Then $Y \sim \mathcal{CN}_p(0, \mathbf{\Lambda_p})$ and the likelihood ratio becomes

$$\mathit{\Lambda}_{\mathrm{LR}} = \frac{|\mathbf{t}_1^{\dagger}\mathbf{y}|^2}{\mathbf{y}^{\dagger}\mathbf{y}}. \tag{6.43}$$

Next we find a unitary transformation $\mathbf{H}$ (not to be confused with the unitary transformation $\mathbf{H}$ defined on page 109) such that

$$\mathbf{e}_1 \equiv \mathbf{H}\,\mathbf{t}_1 = \begin{pmatrix} 1 & \cdot & \cdot & \cdot & 0 \end{pmatrix}^T. \tag{6.44}$$

Then the likelihood ratio can be written as

$$\mathit{\Lambda}_{\mathrm{LR}} = \frac{|\mathbf{e}_1^{\dagger}\mathbf{H}\mathbf{y}|^2}{\mathbf{y}^{\dagger}\mathbf{y}}. \tag{6.45}$$

Defining

$$\mathbf{Z} \equiv \mathbf{H}\,\mathbf{Y} \tag{6.46}$$

leads to

$$\mathit{\Lambda}_{\mathrm{LR}} = \frac{|\mathbf{e}_1^{\dagger}\mathbf{z}|^2}{\mathbf{z}^{\dagger}\mathbf{z}}, \tag{6.47}$$

with $\mathbf{Z} \sim \mathcal{CN}_p(0, \mathbf{R}_z)$ where

$$\mathbf{R}_z = \mathbf{H}\Lambda_p\mathbf{H}^{\dagger}. \tag{6.48}$$

The likelihood ratio test can therefore be written as

$$\frac{|z_1|^2}{\sum_{i=1}^{p}|z_i|^2} \overset{H_1}{\underset{H_0}{\gtrless}} \eta \tag{6.49}$$

where the z_i's denote the components of $\mathbf{z}$. Rearranging yields

$$|z_1| \overset{H_1}{\underset{H_0}{\gtrless}} \Big(\eta_o \sum_{i=2}^{p}|z_i|^2\Big)^{1/2} \tag{6.50}$$

where $\eta_o = \eta/(1-\eta)$, which can be considered (somewhat loosely speaking) as the nonsingular version of the test in (6.24) on page 110. The rest of the derivation for the FAP, which is given below, is similar to the one in Section 6.2.1.

Defining $\tilde{\mathbf{z}} \equiv \begin{pmatrix} z_2 & \cdot & \cdot & \cdot & z_p \end{pmatrix}^T$, the covariance matrix $\mathbf{R}_z$ can be partitioned as

$$\begin{aligned} \mathbf{R}_z &= E\left\{\begin{bmatrix} Z_1 \\ \tilde{\mathbf{Z}} \end{bmatrix}\begin{bmatrix} Z_1^\star & \tilde{\mathbf{Z}}^\dagger \end{bmatrix}\right\} \\ &= \begin{bmatrix} E\{|Z_1|^2\} & E\{Z_1\tilde{\mathbf{Z}}^\dagger\} \\ E\{\tilde{\mathbf{Z}}Z_1^\star\} & E\{\tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^\dagger\} \end{bmatrix} \\ &\triangleq \begin{bmatrix} \sigma_1^2 & \mathbf{R}_{z_1\tilde{\mathbf{z}}} \\ \mathbf{R}_{\tilde{\mathbf{z}}z_1} & \mathbf{R}_{\tilde{\mathbf{z}}} \end{bmatrix}. \end{aligned} \tag{6.51}$$

The random variable Z_1 is conditionally Gaussian with density $f(z_1 \mid \tilde{\mathbf{z}})$ corresponding to the distribution $\mathcal{CN}_1(\mu_c, \sigma_c^2)$ where the conditional mean and variance μ_c and σ_c^2 are given by

$$\begin{aligned} \mu_c = E\{Z_1 \mid \tilde{\mathbf{z}}\} &= \mathbf{R}_{z_1\tilde{\mathbf{z}}}\,\mathbf{R}_{\tilde{\mathbf{z}}}^{-1}\,\tilde{\mathbf{z}}, \\ \sigma_c^2 = \mathrm{var}(Z_1 \mid \tilde{\mathbf{z}}) &= \sigma_1^2 - \mathbf{R}_{z_1\tilde{\mathbf{z}}}\,\mathbf{R}_{\tilde{\mathbf{z}}}^{-1}\,\mathbf{R}_{\tilde{\mathbf{z}}z_1}. \end{aligned} \tag{6.52}$$

The random variable $U \equiv |Z_1|$ is then conditionally Rice with density

$$f(u \mid \tilde{\mathbf{z}}) = \frac{u}{\sigma^2}\exp\big[-(u^2+s^2)/2\sigma^2\big]I_0(us/\sigma^2), \quad u \geq 0 \tag{6.53}$$

where $s = |\mu_c|$ and $\sigma^2 = \sigma_c^2/2$. Defining

$$b(\eta, \tilde{\mathbf{z}}) \equiv \Big(\eta_0 \sum_{i=2}^{p}|z_i|^2\Big)^{1/2} \tag{6.54}$$

the FAP of the test in (6.50) can be expressed as

$$\begin{aligned}\alpha_{\text{LRNMF}} &= P\Big(|Z_1| \geq \Big(\eta_0 \sum_{i=2}^{p} |Z_i|^2\Big)^{1/2}\Big) \\ &= E\Big\{P\Big(U \geq \, b(\eta, \tilde{\mathbf{Z}})|\tilde{\mathbf{Z}}\Big)\Big\} \\ &\triangleq E\big\{g(\eta, \tilde{\mathbf{Z}})\big\} \qquad (6.55)\end{aligned}$$

where

$$\begin{aligned}g(\eta, \tilde{\mathbf{z}}) &= \int_{b(\eta,\tilde{\mathbf{z}})}^{\infty} \frac{u}{\sigma^2} \exp\big[-(u^2+s^2)/2\sigma^2\big] I_0(us/\sigma^2)\, du \\ &= Q_1\left(\frac{s}{\sigma}, \frac{b(\eta,\tilde{\mathbf{z}})}{\sigma}\right) \qquad (6.56)\end{aligned}$$

using (6.53). Then, similarly to (6.33), the FAP can be written as

$$\alpha_{\text{LRNMF}} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} Q_1\left(\frac{s}{\sigma}, \frac{b(\eta,\tilde{\mathbf{z}})}{\sigma}\right) f(\tilde{\mathbf{z}})\, d\tilde{\mathbf{z}} \qquad (6.57)$$

where $f(\tilde{\mathbf{z}})$ corresponds to the (nonsingular) density of $\mathcal{CN}_{p-1}(0, \mathbf{R}_{\tilde{\mathbf{z}}})$. This is the second expression for the exact FAP of the LRNMF detector.

6.3 Nominal statistical model for simulation and threshold setting

A nominal statistical model for the radar returns in the (target-free) primary data vector is required in order to be able to specify a covariance matrix $\mathbf{R}_d$ under the Gaussian assumption. Such a model can then be used to derive threshold settings for desired FAP values for the detector either through simulation or through direct computation of FAP expressions.

For the disturbance covariance $\mathbf{R}_d = s\mathbf{R}_c + \sigma^2\mathbf{I}$ in (6.2) we start again with the low clutter-rank approximation of (6.4) on page 106 as

$$\mathbf{R}_d = \sum_{i=1}^{r} (s\lambda_i + \sigma^2)\, \mathbf{u}_i\mathbf{u}_i^\dagger + \sum_{i=r+1}^{N} \sigma^2\, \mathbf{u}_i\mathbf{u}_i^\dagger = \mathbf{UDU}^\dagger, \qquad (6.58)$$

and identify the matrix of eigenvalues $\mathbf{D}$ as

$$\begin{aligned}\mathbf{D} &= \operatorname{diag}\left(s\lambda_1 + \sigma^2, \ldots, s\lambda_r + \sigma^2, \sigma^2, \ldots, \sigma^2 \right) \\ &= \sigma^2 \operatorname{diag}\left(1 + s\lambda_1/\sigma^2, \ldots, 1 + s\lambda_r/\sigma^2, \underbrace{1, \ldots, 1}_{N-r} \right).\end{aligned} \tag{6.59}$$

Now, choice of the unitary matrix $\mathbf{U}$ of eigenvectors will determine the covariance matrix $\mathbf{R}_d$. The simplest choice that can be made (at the risk of seeming somewhat subjective) is $\mathbf{U} = \mathbf{I}$. In such a case, $\mathbf{R}_d = \mathbf{D}$. That is

$$\mathbf{R}_d = \mathbf{D} = \sigma^2 \begin{pmatrix} 1 + s\lambda_1/\sigma^2 & 0 & . & . & . & 0 \\ 0 & . & . & . & . & . \\ . & . & 1 + s\lambda_r/\sigma^2 & . & . & . \\ . & . & . & 1 & . & . \\ . & . & . & . & . & 0 \\ 0 & . & . & . & 0 & 1 \end{pmatrix}. \tag{6.60}$$

It turns out that this choice is not without some practical significance. In a special case, using the high clutter-to-noise ratio assumption

$$\frac{s\lambda_i}{\sigma^2} \gg 1$$

with

$$\sigma^2 = 1; \; \lambda_i = 1, \, i = 1, \ldots, r$$

leads to a covariance matrix of the form

$$\mathbf{R}_d = \mathbf{D} = \begin{pmatrix} s & 0 & . & . & 0 \\ 0 & s & . & . & . \\ . & . & . & . & . \\ . & . & . & 1 & 0 \\ 0 & . & . & 0 & 1 \end{pmatrix}. \tag{6.61}$$

A plot of the eigenspectrum of $\widehat{\mathbf{R}}_d$ estimated from Gaussian data generated according to the above $\mathbf{R}_d$ is shown in Figure 6.1. This eigenspectrum shows a close resemblance to the eigenspectrum estimated from KASSPER data, [103]. We note of course that this property of our model is due to the shape of the matrix of eigenvalues $\mathbf{D}$ (rather than the particular choice $\mathbf{U} = \mathbf{I}$).

A consequence of the structure of the nominal covariance matrix $\mathbf{R}_d$ in (6.60) is described in the following subsection.

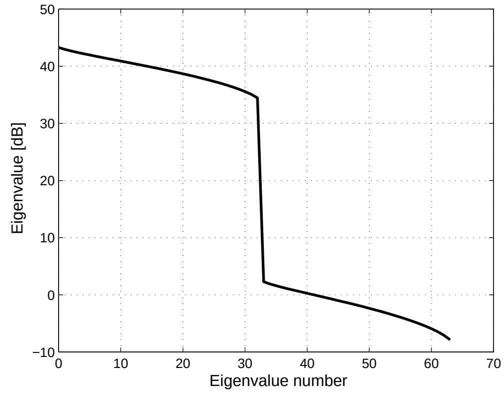


Figure 6.1: Estimated eigenspectrum for data having covariance matrix in (6.61). The clutter rank is $r = 33$ and CNR $= 40$dB.

6.3.1 Exact FAP of LRNMF detector: nominal $\mathbf{R}_d$

We calculate here the FAP of the LRNMF detector when the covariance matrix $\mathbf{R}_d$ of the primary data vector is as in (6.60). The assumption of high clutter-to-noise ratio is not used. Therefore, the only restriction is on the shape of $\mathbf{R}_d$. As $\mathbf{R}_d$ is diagonal, it follows that its eigenvector matrix $\mathbf{U}$ are composed of the columns of $\mathbf{I}$ and hence, for a given r, the projection matrix $\mathbf{Q}$ is given by

$$\begin{aligned}\mathbf{Q} &= \mathbf{I} - \sum_{i=1}^{r} u_i u_i^{\dagger} \\ &= \sum_{i=1}^{N-r} u_i u_i^{\dagger} \\ &= \begin{pmatrix} 0 & . & . & . & . & 0 \\ . & . & . & . & . & . \\ . & . & 0 & . & . & . \\ . & . & . & 1 & . & . \\ . & . & . & . & . & 0 \\ 0 & . & . & . & 0 & 1 \end{pmatrix} \qquad (6.62)\end{aligned}$$

which is of rank $N - r$. Applying this and $\mathbf{R}_d$ in (6.60) to (6.34) yields

$$\mathbf{R}_1 = \mathbf{V}\Lambda\mathbf{V}^{\dagger} = \mathbf{Q}. \qquad (6.63)$$

This implies, from (6.35), that

$$\Lambda = \begin{pmatrix} 1 & 0 & . & . & . & 0 \\ 0 & . & . & . & . & . \\ . & . & 1 & . & . & . \\ . & . & . & 0 & . & . \\ . & . & . & . & . & 0 \\ 0 & . & . & . & 0 & 0 \end{pmatrix} \tag{6.64}$$

$p = N - r$, and thus Λ_p in (6.36) is $\Lambda_p = \mathbf{I}_{N-r}$. Using this in (6.48) yields $\mathbf{R}_z = \mathbf{I}_{N-r}$. Therefore the test in (6.50) represents a simple CA-CFAR test with FAP given by

$$\alpha_{\text{LRNMF}} = (1 - \eta)^{N-r-1}$$

which coincides with the expression for the LRNMF detector in (6.11) on page 108.

Therefore we can conclude that whereas the above formula is an approximation (albeit a good one) for the FAP under the assumptions of low clutter-rank and high clutter-to-noise ratios, it is an exactitude if the covariance matrix of the data possesses the structure in (6.60). In the latter case the clutter-to-noise ratio does not matter.

6.4 The LRNAMF detector

The low-rank normalized adaptive matched filter (LRNAMF) detector is described by

$$\mathit{\Lambda}_{\text{LR-A}} \equiv \frac{|\mathbf{s}^\dagger(\mathbf{I} - \widehat{\mathbf{P}})\mathbf{x}|^2}{(\mathbf{s}^\dagger(\mathbf{I} - \widehat{\mathbf{P}})\mathbf{s})(\mathbf{x}^\dagger(\mathbf{I} - \widehat{\mathbf{P}})\mathbf{x})} \underset{H_0}{\overset{H_1}{\gtrless}} \eta \tag{6.65}$$

where

$$\widehat{\mathbf{P}} = \sum_{i=1}^{r} \mathbf{u}_i \mathbf{u}_i^\dagger \tag{6.66}$$

and $\{\mathbf{u}_i\}_1^N$ are eigenvectors of the covariance matrix estimate $\widehat{\mathbf{R}}_d$ obtained from the secondary vectors with r being the estimated rank of the clutter component of the disturbance.

6.4.1 FAP approximation: low clutter rank and high CNR

The main difference between the LRNMF and LRNAMF detectors (the former given in (6.9)), for a fixed value of r used in both detectors, lies in the set of eigenvectors $\{\mathbf{u}_i\}_1^N$ used to compose the matrix $\mathbf{P}$ (or $\widehat{\mathbf{P}}$). However, for the low-rank clutter model of (6.4) and subject to the high clutter-to-noise ratio approximation of (6.7), the FAP performance of the LRNMF detector is given by the expression of (6.11) and this expression is independent of the eigenvectors used in the detector. This implies that the LRNMF detector is CFAR with respect to the structure of covariance matrices $\mathbf{R}_d$ of data that arise from models satisfying these assumptions. Therefore, for data having the same eigenspectrum as in the above model, and for the same value of r used in both detectors, the FAP performance of the LRNAMF detector will be, largely, equal to that of the LRNMF detector and given by the same formula for each realization of the sample covariance matrix $\widehat{\mathbf{R}}_d$. The LRNAMF detector will also be approximately CFAR under the above assumptions on clutter rank and power. Nevertheless, it must be emphasized that the FAP expression of (6.11) is an approximation.

6.4.2 Exact FAP: arbitrary covariance $\mathbf{R}_d$

If the model assumptions do not hold, specifically in cases where the CNR is not high, then the expressions for exact FAP of the LRNMF detector derived in (6.33) and (6.57) can be used to determine the exact FAP of the LRNAMF detector. Whereas the influence of the secondary vectors on the LRNAMF detector is felt through the eigenvectors in $\widehat{\mathbf{P}}$, via singular value decomposition of the covariance matrix estimate $\widehat{\mathbf{R}}_d$, the FAP is affected by the covariance matrix $\mathbf{R}_{\mathbf{z}}$ in the expression (6.57). This matrix is derived from the eigenvectors in $\widehat{\mathbf{P}}$ through a series of transformation described in Sections 6.2 and 6.2.2 and is, strictly speaking, a random matrix whose statistical properties are related to those of the secondary vectors. Therefore the FAP expression for the LRNMF detector in (6.57) can be considered as a conditional probability expression; the average of this over the statistics of the covariance matrix $\mathbf{R}_{\mathbf{z}}$ will produce the exact FAP performance of the LRNAMF detector. Although it does not appear to be analytically tractable, this fact can be formally expressed by

$$\alpha_{\text{LRNAMF}} \quad = \quad E_{\mathbf{R}_{\mathbf{z}}}\{\alpha_g(\mathbf{R}_{\mathbf{z}})\} = E_{\mathbf{R}_{\mathbf{z}}}\{E\{g(\eta,\tilde{\mathbf{Z}})\big|\mathbf{R}_{\mathbf{z}}\}\} \tag{6.67}$$

with $g(\eta,\tilde{\mathbf{z}})$ given in (6.56).

6.5 Simulations for LR detectors

We report here the simulation results obtained for FAP and threshold estimation for both LRNMF and LRNAMF detectors having the nominal covariance matrix in (6.61). The two FAP expressions, (6.33) and (6.57), have been used to set up estimators for FAP. The estimator for the LRNMF detector can be written as

$$\widehat{\alpha}_{\text{LRNMF}} = \frac{1}{K}\sum_{1}^{K} g(\eta,\tilde{\mathbf{Z}}); \quad \tilde{\mathbf{Z}} \sim \mathcal{CN}_{p-1}(0,\mathbf{R}_{\tilde{\mathbf{z}}}). \tag{6.68}$$

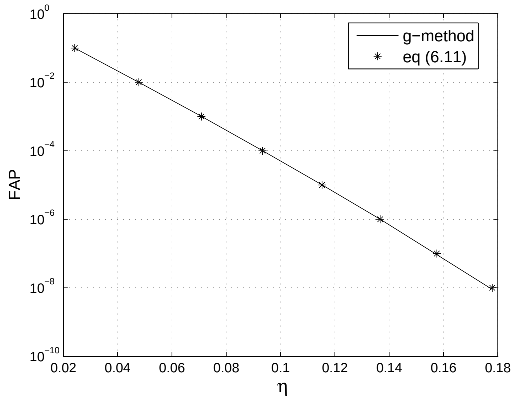


Figure 6.2: Comparison of FAP using the g-method estimator and the formula in (6.11) for the LRNMF detector. $N = 128, r = 33$, $K = 100000$ (for the g-method estimator).

For simplicity we have not used any IS. The FAP estimator for the LRNAMF detector is configured as

$$\begin{aligned} \widehat{\alpha}_{\mathrm{LRNAMF}} &= \frac{1}{K_2}\sum_{1}^{K_2}\left[\frac{1}{K_1}\sum_{1}^{K_1}[g(\eta,\tilde{\mathbf{Z}})]^{(i)}\right]^{(j)}; \\ & \tilde{\mathbf{Z}} \sim \mathcal{CN}_{p-1}(0,\mathbf{R}_{\tilde{\mathbf{z}}}),\ \mathbf{R}_{\tilde{\mathbf{z}}} \sim f(\text{secondary vectors}) \qquad (6.69) \end{aligned}$$

where the inner simulation is conditioned on the covariance matrix $\mathbf{R}_{\tilde{\mathbf{z}}}$, which depends on the data covariance $\mathbf{R}_d$ in place.

Detector thresholds have been estimated using the inverse g-method described in Section 1.2.2.

Results for FAP

The first result for FAP of the LRNMF detector is shown in Figure 6.2, for $N = 128$ and $r = 33$. As expected, for the nominal covariance model the simulation results using the g-method coincide with the formula of (6.11), indicated by stars in the figure. Corresponding simulation gains for the g-method are shown in Figure 6.3.

Figure 6.4 shows the results of the estimated FAP versus the threshold multipliers for LRNAMF detector for different values of r and for $N = 64$ and $L = 128$. The FAP of the LRNMF detector is also shown, and it is obtained using the theoretical formula in (6.11).

A somewhat surprising observation is that for the same data eigenspectrum, the LRNMF and LRNAMF detectors match very closely in terms of achieved FAP values

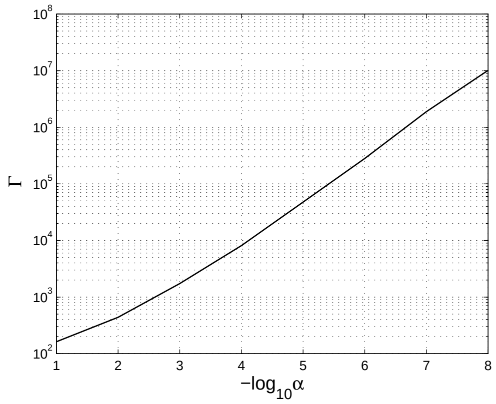


Figure 6.3: Gain of the g-method estimator for the LRNMF detector. $N = 128, r = 33$, $K = 100000$.

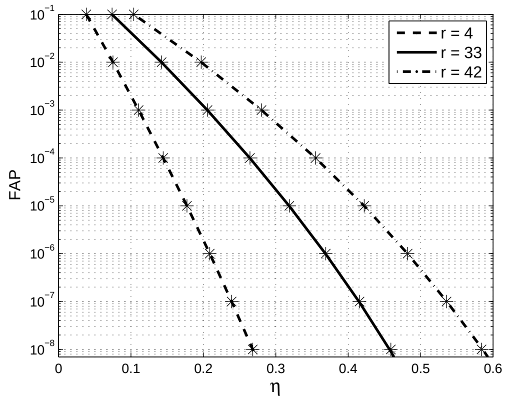


Figure 6.4: Estimated FAP versus threshold multipliers for the LRNMF detector using the theoretical formula in (6.11) for the values of r reported in the legend (red stars) and for the LRNAMF detector using the g-method estimator. $r = 4, 33$, and 42, $N = 64$, $L = 128$, and $K = 100000$.

for given thresholds. A plausible explanation for this behavior lies in the following arguments. Firstly, LR detectors seem to ignore some of the information available in the secondary vectors, depending solely on the eigenvectors obtained through spectral decomposition and not on the actual estimated eigenvalues. Furthermore, for a sufficiently large number L of secondary vectors, the estimated eigenspectrum resembles the actual eigenspectrum in shape. This fact may manifest itself through the eigenvectors of the spectral decomposition.

The estimated simulation gains for the LRNAMF detector using the g-method

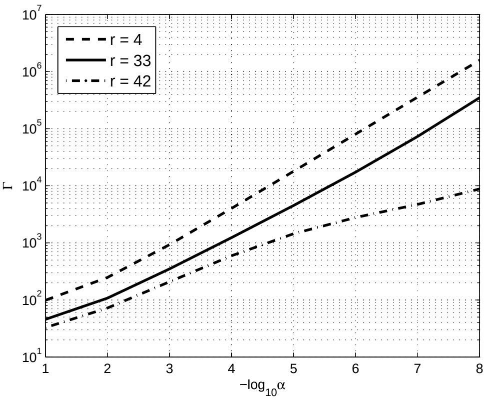


Figure 6.5: Gain of the g-method estimator for the LRNAMF detector. $N = 64, L = 128$.

estimator is shown in Figure 6.5. From this figure we observe that, even without applying IS, the $g-$method alone can achieve considerable gains compared to classical MC. Also, it can be seen that the gain of the $g-$method varies significantly with the rank r of the clutter covariance matrix. This is due to the change of the covariance matrix $\mathbf{R_z}$ that changes, via Λ_p, with the rank r.

Results for detection probability

In Figure 6.6 the detection probability for the adaptive LR detector for both Swerling 0 and 1 models for the target amplitude distribution are shown for the adaptive LR detector.[4] For the covariance matrix, we used in our simulations CNR = 40dB and $r = 33$. Although in a practical scenario the clutter rank would have to be estimated from the data, in the following simulations we assume r to be known (or correctly estimated). Furthermore, for the P_d estimation we used conventional MC simulations, without IS or g-method. We also consider the case when two interfering targets are presents in the secondary vectors. For simplicity, we assume that the two interfering targets share the same steering vector and have SNR = 25dB. Figure 6.6 shows that, for both target amplitude models, the detection performance is almost unchanged by the presence of interfering targets in the secondary data. This can be explained by the fact that, if the rank is correctly estimated, the 2 interfering targets with SNR = 25 dB have no impact on the first r eigenvalues of the covariance matrix, which is clutter dominated. Therefore, as long as the SNR of the interfering targets is smaller

---

[4] The detection performance of the non adaptive LRNMF detector is known in closed form and available in [102].

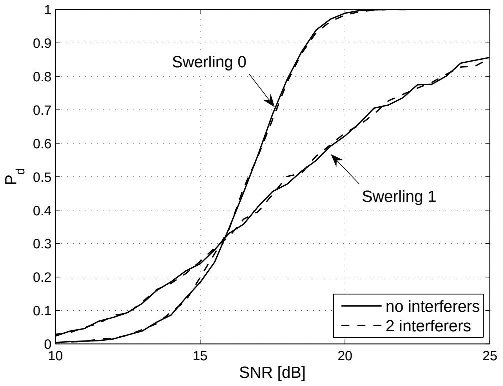


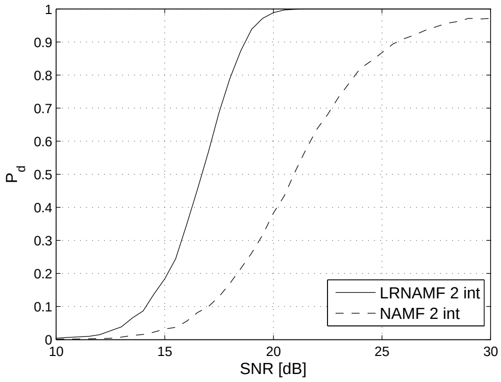
Figure 6.6: Detection probability versus SNR for LRNAMF detector with and without interfering targets in the secondary data. The interfering targets have SNR = 25 dB. $N = 64, L = 128, r = 33$ and $P_{fa} = 10^{-6}$.

than the CNR, and the clutter rank is correctly estimated, the adaptive LR detector is robust against the presence of outliers in the training data.

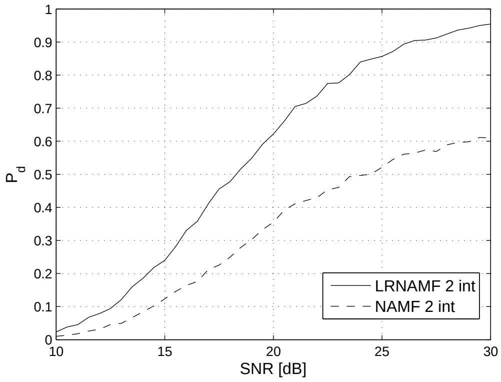
Figure 6.7 compares the detection performance of the adaptive LR and NAMF detectors in the presence of two interfering targets in the training data. We observe that, when the clutter has low rank, than the LRNAMF detector outperforms the NAMF detector, since excluding from the estimate the eigenvalues associated with the interferers results in a better estimate of the clutter covariance matrix of the primary data vector.

6.6 Conclusions

In this chapter some theoretical investigations have been made into the performance of low-rank STAP detectors. We have been able to characterize the FAP performance of LRNMF detectors in terms of detection thresholds using the $g-$method estimator. We showed that the threshold settings for the adaptive detector are very similar to those of the non adaptive LR detector. Furthermore, we reported some results for the detection performance of the LRNAMF detector, and observed that, thanks to the low rank approximation of the clutter covariance matrix, the detector is robust against the presence of outliers in the training data.



(a) Swerling 0



(b) Swerling 1

Figure 6.7: Detection probability versus SNR for the adaptive NAMF and LRNAMF detectors in the presence of two interfering targets in the training data. $N = 64, r = 33$ and $P_{fa} = 10^{-6}$.

Chapter 7

Conclusions

In the research presented in this thesis we focus on novel sampling techniques that can be used for the design, analysis and implementation of radar systems.

In the first part of the thesis we investigate the use of the recently developed Compressive Sensing (CS) technique to radar systems. CS represents an alternative to the Shannon-Nyquist theorem relying on 'sparseness' in some signal domain, making it possible to relax the requirements on sampling rates. However, before CS can be implemented in any operational system, a number of practical problems that have not been addressed before must be solved. One of these is how to perform adaptive target detection from CS measurements. In the work presented here, we devise and compare different architectures for automatic detection of targets embedded in noise from a single CS stepped frequency measurement (range), using both theoretical, simulated and experimental data. We envisage that the proposed architectures can be further extended to include angle- and Doppler processing.

In the second part of the thesis we investigate the use of Importance Sampling (IS) techniques for the design of robust and efficient Space Time Adaptive Processing (STAP) detectors. Space time adaptive processing is a technique used in multi-channel radar systems for improving clutter and interference suppression. STAP produces significant processing gain at the cost of an increased computational complexity. IS represents a fast alternative to classical Monte Carlo simulation for the analysis of such detectors. However, in order to apply IS, an efficient way to bias the probability density function of the test statistic such that the rare event occurs more often must be designed specifically for the problem at hand. We examine several STAP detectors from the standpoint of applying importance sampling to characterize their performances. Various biasing techniques are also devised and implemented, such as biasing by rotation and two dimensional biasing, resulting in significant speed-ups in performance evaluation compared to conventional MC methods.

In the remainder of this chapter we summarize the main contributions of this research in both areas and conclude with some remarks on future work.

7.1 Compressive Sensing radar detectors

The main research question that we address in this thesis is the design and analysis of compressive sensing radar detectors. The efforts towards this goal have resulted in two main contributions:

- the design of the first architecture for adaptive CS radar detection with constant false alarm rate properties, whose performance is also validated with experimental CS measurements;
- the development of a theoretical framework that provides a methodology to predict the performance of the proposed detector, which makes it possible to design practical CS-based radar systems.

The first objective, namely the design of a CS radar detector with CFAR properties, is achieved by exploiting the properties of the Complex Approximate Message Passing (CAMP) algorithm, which is an extension of the AMP algorithm that was carried out during this thesis for ℓ_1-norm recovery from complex data. The CAMP algorithm allows a statistical characterization of the recovered signal and, using this property, we are able to derive closed form expressions for the detection and false alarm probabilities. Furthermore, with a simple modification to the algorithm, we are able to combine the non-linear ℓ_1-norm reconstruction with conventional CFAR detection, which provides a fully adaptive detection scheme. By comparing theoretical and simulated results we also characterize the behavior of CAMP for sensing matrices other than Gaussian, for which the theory was developed, and show that our conclusions still hold for the case of partial Fourier sensing matrices, which are very suitable for radar measurements and for which, unfortunately, no theoretical claims can yet be made. From our analysis and experiments, we conclude that using a separate detector after ℓ_1-norm recovery results in improved detection performance compared to the ℓ_1-detector and the compressive matched filter, as in our architecture both the estimation and detection problems can be optimized separately.

The second main contribution is the derivation of a design methodology for CS-based radar. This is based on closed form expressions for the CAMP output SNR as a function of the system and target parameters, which are obtained in this thesis. These relations can be used to obtain CS link budget plots that allow the system designer to evaluate the trade off between transmit power and undersampling. Moreover, as our novel architecture based on CAMP can be combined with conventional radar detectors, the characterization of the recovery SNR permits the use of classical radar design tools even in the CS framework. Our design methodology plays an important role in determining when and how CS can be applied and at what cost. In fact, the sparsity-undersampling trade offs show that CS comes at a price. If the signal is not sparse enough or the number of measurements is too small, there will be losses in SNR and therefore in performance, the impact of which can be evaluated with the proposed methodology. However, if the signal sparsity is of the order of 1% with respect to the number of degrees of freedom, the signal can be significantly compressed compared to the Nyquist case without noticeable performance loss. In some radar applications where the SNR is not critical, much higher compression can be applied even at lower sparsity.

We believe this work paves the way for the development and implementation of CS in operational radar systems.

7.2 Importance Sampling for analysis and design of STAP detectors

The main objective of this work is to demonstrate the use of importance sampling techniques for the efficient synthesis and analysis of space time adaptive processing radar detectors. The main contributions of this thesis in this direction are:

- the development of ad hoc fast simulation methods for the analysis of the class of Normalized Adaptive Matched Filter STAP detectors;
- the design of two novel STAP detectors, the Geometric Mean and Envelope NAMF detectors, and the investigation of their performance using IS.

In this work, we devise novel biasing schemes for the class of Normalized Adaptive Matched Filter detector and its variants, such as biasing of the probability density function by rotation of the primary data vector and two-dimensional biasing combined with the g-method estimator. Using both adaptive and inverse IS, we characterize the false alarm probability (FAP) performance of the NAMF detectors in terms of detection thresholds and disturbance backgrounds. With some transformations of the test statistic, and using the g-method estimator, we also derive general expressions for the FAP of the low rank NAMF detector that do not make use of the low-rank approximation. By means of simulations, we demonstrate that the adaptive and non-adaptive LR detectors have equal threshold settings under the assumption of low rank and high CNR covariance matrix. We also show that, thanks to the low rank approximation, the LRNAMF detector is robust to the presence of outliers in the secondary data. Furthermore, in this scenario, if the clutter rank is correctly estimated, the

LR detector outperforms the NAMF. By investigating different methods of IS, we are able to gain knowledge on the performance, in terms of gains over conventional MC simulations, of different biasing schemes. Although the design of an efficient IS method requires a meticulous mathematical analysis of the detector under investigation, by comparing our results with traditional Monte Carlo it is clear that IS can achieve tremendous gain in terms of computational time.

Building upon the work in [96], we also present here for the first time the envelope-law (E) and geometric-mean (GM) NAMF STAP detectors. The E and GM NAMF detectors are investigated using fast simulations, and their respective thresholds are determined using two-dimensional biasing schemes with the g-method estimator. The proposed detectors are shown to have better performance in the presence of interfering targets in the training data compared to the standard (square-law) NAMF detector while maintaining almost equal performance in homogeneous Gaussian background. These detectors represent robust alternatives to conventional square-law processing.

7.3 Recommendations and future work

CS is a topic that has attracted much interest in the last few years, resulting in significant developments especially in the theoretical and algorithmic areas. In the field of radar, but also in other areas where CS has been successfully demonstrated on synthetic data, the necessary step to bring the CS theory to the next level is to design CS based hardware and processing chains that can be tested on actual operational systems and practical conditions.

We foresee that in the field of radar CS could be especially useful for achieving super-resolution in imaging radars and for reducing cost, power and weight on small platforms, such as UAVs. Our research has made the first step in this direction, by demonstrating the possibility of adaptive target detection on experimental CS radar measurements and by providing a theoretical framework to gauge the impact of CS on radar design. However, toward this goal several major challenges still need to be addressed. Implementing a multi-dimensional CS based radar system that can work in operational conditions requires dealing with issues such as clutter, interference, and non-homogeneity of the environment. The long history of radar systems has to some extent addressed all these issues for classical Nyquist-based radars, resulting for example in very advanced STAP algorithms for multi-channel, coherent radar systems. However, these issues might have to be reconsidered and have not been addressed yet for CS radars. The challenges that would need to be tackled include dealing with non-gaussian and non-white clutter and interference; extending our approach to multiple signal domains (time, Doppler, channel); characterizing the tradeoffs between compression and power, and analyzing the impact on the radar architecture. Further, in a multi-dimensional approach it is of great interest to explore the possibilities of adaptive compressive sensing, in such a way that the available resources can be efficiently divided amongst the different radar tasks, depending on the environment and target scenario.

Concerning our outlook on application of IS to STAP detectors, it would be particularly interesting to further investigate the design of novel biasing schemes for the analysis of STAP detectors whose performance is not known, such as the nonhomogeneity and parametric detectors. The so far developed IS methods could provide a starting point for the analysis of such detectors. Another important extension would be to consider different target models and more general clutter distributions, such as the K-distributed models, for clutter backgrounds that are homogeneous as well as nonhomogeneous. Simultaneously, the conceptualization and development of new detection algorithms from a robustness perspective together with necessary performance characterization could be carried out, as it was done in this thesis for the envelope and geometric mean variants of the NAMF STAP detector. Another desirable development would be the inclusion of emerging knowledge-aided STAP detection techniques under the IS umbrella.

Finally, with further research into CS-based techniques for multi-dimensional radar applications, it is very likely that new detection algorithms capable of dealing with a number of practical situations will be developed. It is also likely that many of these algorithms will not lead to simple and straightforward analytical performance characterizations, especially with regard to false alarm and detection probabilities that are standard metrics in radar system design. Thus, the designer will have to resort to Monte Carlo simulations necessitating the use of IS for evaluating CS based CFAR algorithms. To achieve this combination of CS-based system design and IS-based performance analysis, significant effort will be required to develop accurate statistical characterizations of the inputs and outputs of these new detection algorithms. It is clear from this that there are new and challenging directions for future research in the areas dealt with in this thesis.

Appendix A

Risk of the soft thresholding function

We derive an analytical expression for (3.2) for the case when the non-zero elements of the vector $\mathbf{x}$ all have equal amplitude a and (possibly non-zero) phase θ, i.e., $x_i = ae^{j\theta}$, $i = 1, \cdots, k$. We also demonstrate that, for the complex signal case, the function $\Psi(\sigma_*)$ is independent of the phase of the non-zero coefficients. At the end of the appendix, we also explain how the calculations can be easily generalized to the more general setting where the non-zero elements have different amplitude distributions.

Using the above distribution for the non-zero coefficients of $\mathbf{x}$ in (3.2), we obtain

$$\begin{aligned}
\sigma_*^2 &= \sigma^2 + \frac{1}{\delta}\mathbb{E}_{Z,X}\{|(\eta(X+\sigma_* Z;\tau\sigma_*) - X)|^2\} \\
&= \sigma^2 + \frac{1}{\delta}\Big\{\mathbb{E}_Z\{|(\eta(\sigma_* Z;\tau\sigma_*))|^2\}P(x=0) \\
&\qquad\qquad + \mathbb{E}_Z\{|(\eta(ae^{j\theta}+\sigma_* Z;\tau\sigma_*) - ae^{j\theta})|^2\}P(x=ae^{j\theta})\Big\} \\
&= \sigma^2 + \frac{\sigma_*^2}{\delta}\Big\{(1-\delta\rho)\mathbb{E}_Z\{|(\eta(Z;\tau))|^2\} + \delta\rho\mathbb{E}_Z\{|(\eta(\mu e^{j\theta}+Z;\tau) - \mu e^{j\theta})|^2\}\Big\},
\end{aligned} \tag{A.1}$$

where $\mu = a/\sigma_*$, and $Z \sim \mathcal{CN}(0,1)$ and can be decomposed as $Z = (Z_r + jZ_c)$, where $Z_r, Z_c \sim \mathcal{N}(0, 1/2)$. Define the two independent random variables $W = |Z| \sim$ Rayleigh($\frac{1}{2}$) and $\vartheta = \measuredangle(Z) \sim$ that is uniformly distributed between 0 and 2π. The first expectation in (A.1) can be computed as

$$\begin{aligned}\mathbb{E}_Z\{|\eta(Z;\tau)|^2\} &= \int_{\vartheta}\int_{w>\tau}(w-\tau)^2 f_w(w) f_\vartheta(\vartheta)\,dw\,d\vartheta = 2\int_{w>\tau} w(w-\tau)^2\, e^{-w^2}dw\\ &= 2\sqrt{\pi}\left(\int_{w>\tau} w^3\,\frac{e^{-w^2}}{\sqrt{\pi}}dw - 2\tau\int_{w>\tau} w^2\,\frac{e^{-w^2}}{\sqrt{\pi}}dw + \tau^2\int_{w>\tau} w\,\frac{e^{-w^2}}{\sqrt{\pi}}dw\right),\end{aligned} \tag{A.2}$$

where each of the integrals in the last line is the incomplete moment of a Gaussian random variable with parameters $(0, \frac{1}{\sqrt{2}})$ of order 3, 2, and 1, respectively. These integrals can be computed numerically.

Now we want to compute the second expectation in (A.1), which is given by

$$\mathbb{E}(|\eta(\mu \mathrm{e}^{j\theta} + Z_r + jZ_c;\tau) - \mu\mathrm{e}^{j\theta}|^2), \tag{A.3}$$

where μ, and θ are fixed numbers and the expectation is with respect to the RV Z_r, Z_c, which are independent and distributed as $N(0, 1/2)$. Expanding the expected value, and using the definition of complex soft thresholding function, we obtain

$$\begin{aligned}&\mathbb{E}(|\eta(\mu \mathrm{e}^{j\theta} + Z_r + jZ_c;\tau) - \mu\mathrm{e}^{j\theta}|^2)\\ =& \int_{(\mu cos\theta+z_r)^2+(\mu sin\theta+z_c)^2\le\tau^2} \mu^2\frac{1}{\pi}\mathrm{e}^{-(z_r^2+z_c^2)}dz_r dz_c\\ +& \int_{(\mu cos\theta+z_r)^2+(\mu sin\theta+z_c)^2\ge\tau^2} |(\sqrt{(\mu\cos\theta+z_r)^2+(\mu\sin\theta+z_c)^2}-\tau)\mathrm{e}^{j\tan^{-1}\frac{\mu\sin\theta+z_c}{\mu\cos\theta+z_r}}\\ -& \mu\cos\theta - j\mu\sin\theta|^2\frac{1}{\pi}\mathrm{e}^{-(z_r^2+z_c^2)}dz_r dz_c.\end{aligned} \tag{A.4}$$

We first simplify the second integral in (A.4). Using the equalities

$$\begin{aligned}\cos\big(\tan^{-1}\frac{\mu\sin\theta+z_c}{\mu\cos\theta+z_r}\big) &= \frac{\mu\cos\theta+z_r}{\sqrt{(\mu\cos\theta+z_r)^2+(\mu\sin\theta+z_c)^2}}\\ \sin\big(\tan^{-1}\frac{\mu\sin\theta+z_c}{\mu\cos\theta+z_r}\big) &= \frac{\mu\sin\theta+z_c}{\sqrt{(\mu\cos\theta+z_r)^2+(\mu\sin\theta+z_c)^2}}\end{aligned}$$

we can rewrite the second integral in (A.4) as

$$\begin{aligned}
& \int\limits_{(\mu cos\theta+z_r)^2+(\mu sin\theta+z_c)^2\geq\tau^2} |(\sqrt{(\mu\cos\theta+z_r)^2+(\mu\sin\theta+z_c)^2}-\tau)\mathrm{e}^{j\tan^{-1}\frac{\mu\sin\theta+z_c}{\mu\cos\theta+z_r}} \\
- & \ \mu\cos\theta - j\mu\sin\theta|^2\frac{1}{\pi}\mathrm{e}^{-(z_r^2+z_c^2)}dz_r dz_c \\
= & \int\limits_{(\mu cos\theta+z_r)^2+(\mu sin\theta+z_c)^2\geq\tau^2} |z_r+jz_c-\frac{\tau(\mu\cos\theta+z_r)}{\sqrt{(\mu\cos\theta+z_r)^2+(\mu\sin\theta+z_c)^2}} \\
- & \ j\frac{\tau(\mu\sin\theta+z_c)}{\sqrt{(\mu\cos\theta+z_r)^2+(\mu\sin\theta+z_c)^2}}|^2\frac{1}{\pi}\mathrm{e}^{-(z_r^2+z_c^2)}dz_r dz_c. \qquad \text{(A.5)}
\end{aligned}$$

Define $x = \mu\cos\theta + z_r$ and $y = \mu\sin\theta + z_c$. Using this change of variables and transforming to polar coordinates, we can rewrite (A.5) as

$$\begin{aligned}
& \int\limits_{\sqrt{x^2+y^2}>\tau} |x-\mu\cos\theta+j(y-\mu\sin\theta)-\frac{\tau x}{\sqrt{x^2+y^2}}-j\frac{\tau y}{\sqrt{x^2+y^2}}|^2 \\
& \frac{1}{\pi}\mathrm{e}^{-(x-\mu\cos\theta)^2-(y-\mu\sin\theta)^2}dxdy \\
= & \int\limits_{\phi=0}^{2\pi}\int\limits_{r>\tau} |r\cos\phi-\mu\cos\theta+j(r\sin\phi-\mu\sin\theta)-\tau\cos\phi-j\tau\sin\phi|^2 \\
& \frac{1}{\pi}\mathrm{e}^{-(r\cos\phi-\mu\cos\theta)^2-(r\sin\phi-\mu\sin\theta)^2}rdrd\phi \\
= & \int\limits_{\phi=0}^{2\pi}\int\limits_{r>\tau} |(r-\tau)\cos\phi-\mu\cos\theta+j((r-\tau)\sin\phi-\mu\sin\theta)|^2 \\
& \frac{1}{\pi}\mathrm{e}^{-(r\cos\phi-\mu\cos\theta)^2-(r\sin\phi-\mu\sin\theta)^2}rdrd\phi \\
= & \int\limits_{\phi=0}^{2\pi}\int\limits_{r>\tau} \frac{1}{\pi}[(r-\tau)^2+\mu^2-2\mu(r-\tau)\cos(\theta-\phi)] \\
& \mathrm{e}^{-r^2-\mu^2+2r\mu\cos(\theta-\phi)}rdrd\phi. \qquad \text{(A.6)}
\end{aligned}$$

Using the same transformation as above, we now simplify the first integral in

(A.4).

$$\begin{aligned}
&\int\limits_{(\mu cos\theta+z_r)^2+(\mu sin\theta+z_c)^2\leq\tau^2} \mu^2\frac{1}{\pi}\mathrm{e}^{-(z_r^2+z_c^2)}dz_rdz_c\\
&= \int\limits_{x^2+y^2\leq\tau^2} \mu^2\frac{1}{\pi}\mathrm{e}^{-(x-\mu\cos\theta)^2+(y-\mu\sin\theta)^2}dxdy\\
&= \int\limits_{\phi=0}^{2\pi}\int\limits_{r\leq\tau} \mu^2\frac{1}{\pi}\mathrm{e}^{-(r\cos\phi-\mu\cos\theta)^2+(r\sin\phi-\mu\sin\theta)^2}rdrd\phi\\
&= \int\limits_{\phi=0}^{2\pi}\int\limits_{r\leq\tau} \mu^2\frac{1}{\pi}\mathrm{e}^{-r^2-\mu^2+2r\mu\cos(\theta-\phi)}rdrd\phi. \qquad (A.7)
\end{aligned}$$

For the periodicity of the cosine function, the integrals in (A.6) and (A.7) are independent of θ and, therefore, in Chapter 3 (and in the remainder of this demonstration) we have consider the case $\theta = 0$.

Now, combining the 2 integrals in (A.6) and (A.7) and using the following relations [178, 179]

$$\begin{aligned}
\int\limits_0^{2\pi}\mathrm{e}^{z\cos x}dx &= 2\pi I_0(z),\\
\int\limits_0^{2\pi}\cos x\mathrm{e}^{z\cos x}dx &= 2\pi I_1(z),\\
M_{-\frac{1}{2},0}(\mu^2) &= \mu\mathrm{e}^{-\mu^2/2}M_{1,1}(\mu^2) = \mu\mathrm{e}^{-\mu^2/2}\mathrm{e}^{\mu^2} = \mu\mathrm{e}^{\mu^2/2},\\
2\mu^2\mathrm{e}^{-\mu^2}\int\limits_0^{\infty}r\mathrm{e}^{-r^2}I_0(2r\mu)dr &= \mu^2\mathrm{e}^{-\mu^2}\int\limits_0^{\infty}\mathrm{e}^{-x}I_0(2\sqrt{\mu^2x})dx = \mu^2\mathrm{e}^{-\mu^2}\left(\frac{\mathrm{e}^{\mu^2/2}}{\mu}M_{-\frac{1}{2},0}(\mu^2)\right)\\
&= \mu\mathrm{e}^{-\mu^2/2}M_{-\frac{1}{2},0}(\mu^2) = \mu^2,
\end{aligned}$$

where I_0 and I_1 are modified Bessel functions of the first kind of order 0 and 1 and M is the Whittaker M function (pp.505 and 509 of [178]), the expectation in (A.3) can be written as

$$
\begin{aligned}
&\mathbb{E}(|\eta(\mu + z_r + jz_c;\tau) - \mu|^2) = \int\limits_{\phi=0}^{2\pi} \int\limits_{r\le\tau} \mu^2 \frac{1}{\pi} \mathrm{e}^{-r^2-\mu^2+2r\mu\cos(\phi)} r dr d\phi \\
&+ \frac{1}{\pi} \int\limits_{\phi=0}^{2\pi} \int\limits_{r>\tau} [(r-\tau)^2 + \mu^2 - 2\mu(r-\tau)\cos(\phi)]\mathrm{e}^{-r^2-\mu^2+2r\mu\cos(\phi)} r dr d\phi \\
&= 2\mu^2 \mathrm{e}^{-\mu^2} \int\limits_{0}^{\tau} r\mathrm{e}^{-r^2} I_0(2r\mu)dr + 2\mathrm{e}^{-\mu^2} \int\limits_{\tau}^{\infty} r(r-\tau)^2 \mathrm{e}^{-r^2} I_0(2r\mu)dr \\
&+ 2\mu^2 \mathrm{e}^{-\mu^2} \int\limits_{\tau}^{\infty} r\mathrm{e}^{-r^2} I_0(2r\mu)dr - 2\mathrm{e}^{-\mu^2} \int\limits_{\tau}^{\infty} 2r\mu(r-\tau)\mathrm{e}^{-r^2} I_1(2r\mu) \\
&= 2\mu^2 \mathrm{e}^{-\mu^2} \int\limits_{0}^{\infty} r\mathrm{e}^{-r^2} I_0(2r\mu)dr + \mathrm{e}^{-\mu^2} \int_{\tau}^{\infty} r(r-\tau)\mathrm{e}^{-r^2} \Big(- 4\mu I_1(2r\mu) + 2(r-\tau) I_0(2r\mu)\Big) dr \\
&= \mu^2 + \mathrm{e}^{-\mu^2} \int_{\tau}^{\infty} r(r-\tau)\mathrm{e}^{-r^2} \Big(- 4\mu I_1(2r\mu) + 2(r-\tau) I_0(2r\mu)\Big) dr \qquad \text{(A.8)}
\end{aligned}
$$

The integral in the last line of (A.8) can be implemented in Matlab.

Finally, note that if the amplitudes of the non-zero coefficients are drawn from an arbitrary distribution G, then we can also calculate the expected value of the expressions in (A.8) with respect to $\mu \sim G$.

Appendix B

Proof of Theorem 3.2.2

To prove Theorem 3.2.2, we derive the detection and false alarm probabilities of Architectures 1 and 2. Since the thresholds τ_α and τ_o are different for Architectures 1 and 2, in what follows, $\sigma_{*,\alpha}$ and $\sigma_{*,o}$ denote the fixed point solutions of SE for Architectures 1 and 2, respectively.

Architecture 1

Recall from Figure 3.3(a) that in Architecture 1 the non-zero coefficients in $\widehat{\mathbf{x}}$ represent the final detections and the threshold τ_α is selected so as to achieve the desired FAP α. In this architecture, the output of CAMP is given by

$$\widehat{\mathbf{x}} = \eta(\mathbf{x} + \sigma_{*,\alpha}\mathbf{z}; \tau_\alpha \sigma_{*,\alpha}), \tag{B.1}$$

where $\mathbf{z} \sim \mathcal{CN}(0, \mathbf{I})$. The Gaussianity of the noise is due to the assumption of the Theorem, and it holds in the asymptotic settings according to Theorem 3.1.2. The test statistic at bin $i, i = 1, \cdots, N$, is given by

$$|\widehat{x}_i| \underset{H_0}{\overset{H_1}{\gtreqqless}} 0. \tag{B.2}$$

Therefore, we have

$$P_{fa_1} = P(|\widehat{x}_i| \neq 0 | H_0) = P(|\sigma_{*,\alpha} z| > \tau_\alpha \sigma_{*,\alpha}) = P(|z| > \tau_\alpha) = e^{-\tau_\alpha^2},$$

where $z \sim \mathcal{CN}(0, 1)$. This leads us to the following straightforward parameter tuning: $\tau_\alpha = \sqrt{-\ln \alpha}$.

For evaluating the detection probability, we define a as the square root of the power received from a target at location i. We then have

$$\begin{aligned} P_{d_1} &= P(|\widehat{x}_i| \neq 0|H_1) = P(|\eta(\tilde{x}_i; \tau_\alpha \sigma_{*,\alpha})| \neq 0|H_1) = P(|x_{o,i} + \sigma_{*,\alpha} z| > \tau_\alpha \sigma_{*,\alpha}) \\ &= P(|a + \sigma_{*,\alpha} z| > \tau_\alpha \sigma_{*,\alpha}) = Q_1(\sqrt{2\mathrm{SNR}_{CS,1}}, \sqrt{-2\ln\alpha}), \end{aligned} \tag{B.3}$$

where $\mathrm{SNR}_{CS,1} = a^2/\sigma^2_{*,\alpha}$, $\sigma^2_{*,\alpha}$ is evaluated using (3.2) with threshold parameter τ_α, and $Q_1(.,.)$ is the Marcum-Q function.

Architecture 2

Suppose that after the estimation block we obtain the following noisy estimate of $\mathbf{x}$

$$\widetilde{\mathbf{x}} = \mathbf{x} + \sigma_{*,o}\mathbf{z}.$$

For Architecture 2, the decision statistic at bin i is given by

$$|\widetilde{x}_i| \underset{H_0}{\overset{H_1}{\gtrless}} \kappa.$$

Therefore,

$$P_{fa_2} = P(|\widetilde{x}_i| > \kappa|H_0) = P(|\sigma_{*,o} z| > \kappa) = P(|z| > \kappa/\sigma_{*,o}) = e^{-\kappa^2/\sigma^2_{*,o}}.$$

The last equality comes from the fact that $|z| \sim \mathrm{Rayleigh}(1/2)$. Thus, for a desired FAP α, the detector threshold can be set as $\kappa = \sigma_{*,o}\sqrt{-\ln\alpha}$, where as before $\sigma^2_{*,o}$ can be computed from (3.2) with threshold parameter τ_o. The detection probability is given by

$$\begin{aligned} P_{d_2} &= P(|\widetilde{x}_i| > \kappa|H_1) = P(|x_{o,i} + \sigma_{*,o} z| > \kappa) = P(|a + \sigma_{*,o} z| > \kappa) \\ &= Q_1(\sqrt{2\mathrm{SNR}_{CS,2}}, \sqrt{-2\ln\alpha}), \end{aligned} \tag{B.4}$$

where $\mathrm{SNR}_{CS,2} = a^2/\sigma^2_{*,o}$.

The proof of Theorem 3.2.2 is now straightforward. Since $\sigma^2_{*,o} \leq \sigma^2_{*,\alpha}$, for the same target received power a^2, $\mathrm{SNR}_{CS,2} \geq \mathrm{SNR}_{CS,1}$ and, therefore, for the same FAP α, $P_{d_2} \geq P_{d_1}$, with equality if and only if $\alpha = e^{-\tau_o^2}$. Note that the above arguments are independent of the distribution G of the non-zero elements of $\mathbf{x}$.

Appendix C

FAP estimation of the NMF by rotation of primary vector

In this appendix, we developed a method of IS for estimating the FAP of the NMF detector in (5.5) that is based on rotation of the primary data vector

If we want to estimate α_{NMF} using IS with, e.g., biasing of the input data vector, then it is clear from the test statistic in (5.5) that a simple scaling of the elements of the primary data vector $\mathbf{x}$ will be useless. A form of biasing can however be developed using the (well known) fact that the test statistic is actually a cosine-squared one, the concerned angle being that between the transformed steering (or unit) and whitened primary data vectors $\mathbf{s}_1$ (or $\mathbf{u_1}$) and $\mathbf{x}_1$ respectively. With these transformation, the NMF test statistic in (5.5) takes the form

$$\mathit{\Lambda}_{\text{NMF}} = \frac{|\mathbf{s}_1^\dagger \mathbf{x}_1|^2}{(\mathbf{s}_1^\dagger \mathbf{s}_1)(\mathbf{x}_1^\dagger \mathbf{x}_1)} = \frac{|y_1|^2}{\|\mathbf{x}_1\|^2}, \tag{C.1}$$

The frequency of false alarm events in a simulation can therefore be increased by biasing $\mathbf{x}_1$ so as to decrease the angle toward zero or increase it toward π. This can be accomplished by a rotation of $\mathbf{x}_1$ which, since it is being assumed for this detector that the actual data covariance matrix $\mathbf{R}$ is completely known, is equivalent to input biasing (of the primary data vector $\mathbf{x}$).

Consider biasing the whitened primary data vector $\mathbf{x}_1$ by rotation with an $N \times N$ matrix $\mathbf{A}$. The biased vector is $\mathbf{A}\mathbf{x}_1$ and will be distributed as $\mathcal{CN}_N(0, \mathbf{R}_\star)$ with covariance matrix $\mathbf{R}_\star$ given by $\mathbf{R}_\star = \mathbf{A}\mathbf{A}^\dagger$, since $\mathbf{x}_1 \sim \mathcal{CN}_N(0, \mathbf{I})$. The weighting function is then given by

$$W(\mathbf{x}_1; \mathbf{A}) = \frac{f(\mathbf{x}_1)}{f_\star(\mathbf{x}_1)} = |\mathbf{R}_\star| \exp\big(-\mathbf{x}_1^\dagger(\mathbf{I} - \mathbf{R}_\star^{-1})\mathbf{x}_1\big), \tag{C.2}$$

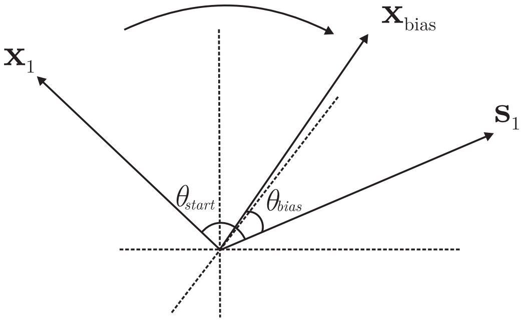


Figure C.1: Effect of biasing by rotation of the primary data vector.

where $|\cdot|$ denotes matrix determinant. The FAP estimator for the test in (C.1) then takes the form

$$\widehat{\alpha}_{\text{NMF}} = \frac{1}{K}\sum_{1}^{K} 1\big(|\mathbf{u}_1^\dagger \mathbf{x}_1|^2 \geq \eta\, \|\mathbf{x}_1\|^2\big)\, W(\mathbf{x}_1;\mathbf{A}); \quad \mathbf{x}_1 \sim f_\star. \tag{C.3}$$

The problem now centers around determining an effective rotation or biasing matrix $\mathbf{A}$. There is clearly an unbounded number of rotation matrices to choose from. To narrow the choice, it is most convenient to make the matrix dependent on a single (biasing) parameter, say a, and search for the optimum value of the latter by minimizing the associated I-function which is given by

$$I(a) = E_\star\big\{1(\mathcal{A})\, W^2(\mathbf{x}_1;\mathbf{A})\big\} = E\big\{1(\mathcal{A})\, W(\mathbf{x}_1;\mathbf{A})\big\}, \tag{C.4}$$

where $\mathcal{A} \equiv 1\big(|\mathbf{u}_1^\dagger \mathbf{x}_1|^2 \geq \eta\, \|\mathbf{x}_1\|^2\big)$ is the false alarm event. In actual simulation we will determine a minimizer for an estimate of $I(a)$, that is

$$a_{\text{opt}} = \arg\min_a \Big(\widehat{I}(a) = \frac{1}{K}\sum_{1}^{K} 1(\mathcal{A})\, W^2(\mathbf{x}_1;\mathbf{A});\ \sim f_\star\Big). \tag{C.5}$$

The form or structure of $\mathbf{A}$ has to be now decided. We assume that $\mathbf{A} = \mathbf{I} + a\mathbf{T}$ for $0 \leq a < 1$ where $\mathbf{T}$ is some $N \times N$ matrix. The biasing matrix $\mathbf{A}$ is to be constructed so as to move the data vector $\mathbf{x}_1$ in a controlled manner towards $\mathbf{u}_1$, as shown in Figure C.1. Owing to this, the matrix will depend not only on the parameter a but also on the components of $\mathbf{u}_1$. For the general case $\mathbf{R} \neq \mathbf{I}$, we make the choice that $\mathbf{A} = \mathbf{I} + a\mathbf{T}$ for $0 \leq a < 1$, where

$$\mathbf{T} = \begin{pmatrix} t_1 - 1 & t_1 & . \;\; . & t_1 \\ t_2 & t_2 - 1 & . \;\; . & t_2 \\ . & & & . \\ t_N & . & . \;\; . & t_N - 1 \end{pmatrix},$$

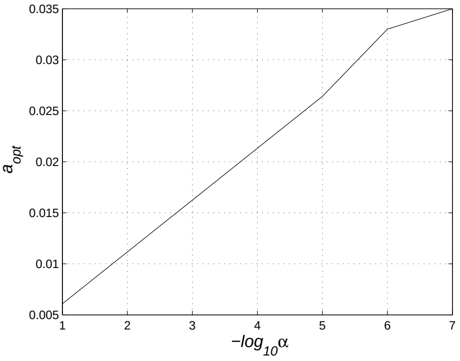
and t_i, $i = 1, \ldots, N$, denote the elements of $\mathbf{u}_1$. Then, the biasing matrix becomes

$$\mathbf{A} = \begin{pmatrix} 1+(t_1-1)a & t_1 a & . & . & t_1 a \\ t_2 a & & . & . & t_2 a \\ . & & & & . \\ t_N a & & . & . & . \quad 1+(t_N-1)a \end{pmatrix}.$$

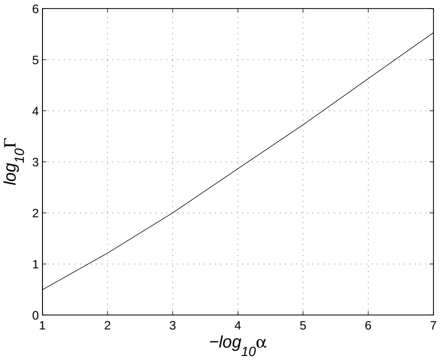
The components of the biased data vector $\mathbf{x}_1$ thus become correlated. For the special case $\mathbf{R} = \mathbf{I}$, $\mathbf{A}$ reduces to a matrix with unity on the diagonal and a everywhere else, assuming that $\mathbf{s}$=$[1\ldots1]'/\sqrt{N}$. When $a = 1$ the biased vector is just $\big(\sum_{i=1}^{N} x_{1i}\big)\mathbf{u}_1$ and is collinear with $\mathbf{u}_1$, where x_{1i}'s denote the elements of the unbiased $\mathbf{x}_1$. That is, the fully biased vector would be aligned in the direction of $\mathbf{u}_1$ or opposite to it, in $\mathcal{C}^N$. However we avoid simulating with $a = 1$ since it obviously leads to a singular $\mathbf{A}$ matrix. When $a = 0$, no biasing or rotation takes place. Estimating an optimum value of a that maximizes the simulation gain for a given threshold η is an easy problem, and can be solved using the adaptive IS method described in Section 1.2.2.

An interesting phenomenon takes place with this biasing scheme. The weighting function in (C.2) depends on the matrix $\mathbf{A}$ which in turn depends (apart from on the biasing parameter a) on the elements of the unit vector $\mathbf{u}_1$ through the $\mathbf{T}$ matrix. The IS estimator in (C.3) estimates $\alpha_{\text{NMF}} = E\{1(\mathcal{A})\}$ which, from (5.14), is known to be independent of the particular unit vector $\mathbf{u}_1$ being used. Hence, if the estimator $\widehat{\alpha}_{\text{NMF}}$ is a reasonably good one, then it will be largely unaffected by the choice of the biasing matrix $\mathbf{A}$ (and hence of $\mathbf{T}$). However, the gain of the proposed IS scheme depends on the I-function given in (C.4) and this depends on the weighting function and therefore on the matrix $\mathbf{T}$. Thus the maximum gain for the optimized IS scheme will be affected by the choice of unit vector $\mathbf{u}_1$. Since $\mathbf{u}_1 = \mathbf{R}^{-1/2}\mathbf{s}/\|\mathbf{R}^{-1/2}\mathbf{s}\|$ it follows that the IS performance of this rotation biasing scheme depends on the data covariance matrix $\mathbf{R}$ in force; this is despite the fact that the detector FAP is independent of $\mathbf{R}$. Decoupling the biasing matrix $\mathbf{A}$ from $\mathbf{u}_1$ is of course not possible. Thus there may exist some $\mathbf{R}$ which gives a best IS simulation gain for given detector constants η and N. Furthermore, it may be possible to achieve some invariance of IS gain to covariance matrix $\mathbf{R}$ by using a matrix normalization for the $\mathbf{T}$ or $\mathbf{A}$ matrices. In the interests of expediency, the results presented here are obtained by simulating the case $\mathbf{R} = \mathbf{I}$. However, it must be stated here that in the general case of $\mathbf{R} \neq \mathbf{I}$ this rotation biasing scheme would be the only possible one to use if the probability distributions of the clutter and interference returns departed from the Gaussian.

The simulation results for the optimum biasing parameter a_{opt} and the IS gain Γ as functions of the FAP α for the rotation biasing scheme are shown in Figures C.2(a) and C.2(b) respectively.



(a) Optimum biasing parameter a_{opt}



(b) IS simulation gain Γ

Figure C.2: Simulation results using biasing by rotation for the NMF detector. $\mathbf{R} = \mathbf{I}$, $N = 128$, and $K = 5000$.

Appendix D

FAP Estimation of the AMF revisited

D.1 FAP estimation using 2 dimensional biasing

The AMF detection test, as obtained in [169], is given by

$$\frac{|\mathbf{s}^\dagger \widehat{\mathbf{R}}^{-1}\mathbf{x}|^2}{\mathbf{s}^\dagger \widehat{\mathbf{R}}^{-1}\mathbf{s}} \underset{H_0}{\overset{H_1}{\gtrless}} \eta. \tag{D.1}$$

As shown in [169], the test in (D.1) can be rewritten as

$$\begin{aligned} |\mathbf{s}^\dagger \widehat{\mathbf{R}}^{-1}\mathbf{x}|^2 \quad & \underset{H_0}{\overset{H_1}{\gtrless}} \quad \eta\, \mathbf{s}^\dagger \widehat{\mathbf{R}}^{-1}\mathbf{s} \\ & = \quad \eta\, \mathbf{s}^\dagger \widehat{\mathbf{R}}^{-1}\widehat{\mathbf{R}}\widehat{\mathbf{R}}^{-1}\mathbf{s} \\ & = \quad \eta\, \mathbf{s}^\dagger \widehat{\mathbf{R}}^{-1}\frac{1}{L}\sum_{l=1}^{L}\mathbf{x}(l)\mathbf{x}(l)^\dagger \widehat{\mathbf{R}}^{-1}\mathbf{s} \\ & = \quad \frac{\eta}{L}\sum_{l=1}^{L}\mathbf{s}^\dagger \widehat{\mathbf{R}}^{-1}\mathbf{x}(l)\mathbf{x}(l)^\dagger \widehat{\mathbf{R}}^{-1}\mathbf{s} \\ & = \quad \frac{\eta}{L}\sum_{l=1}^{L}|\mathbf{s}^\dagger \widehat{\mathbf{R}}^{-1}\mathbf{x}(l)|^2. \end{aligned} \tag{D.2}$$

This is in the form of a vector (or, array) version of the usual CA-CFAR test. The LHS is a square law detector, being the output of a matched filter (matched to the direction $\mathbf{s}$ in which the array is steered) for incoherent detection using the so-called sample matrix inversion (SMI) beamformer weights $\widehat{\mathbf{R}}^{-1}\mathbf{s}$. The RHS represents a cell averaging term. Define the random variables[1]

$$G \equiv \mathbf{s}^{\dagger}\widehat{\mathbf{R}}^{-1}\mathbf{x} \quad \text{and} \quad G(l) \equiv \mathbf{s}^{\dagger}\widehat{\mathbf{R}}^{-1}\mathbf{x}(l), \tag{D.3}$$

for $l = 1, \ldots, L$. Using (D.3) the AMF test in (D.2) can be rewritten as

$$|G|^2 \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\eta}{L} \sum_{l=1}^{L} |G(l)|^2. \tag{D.4}$$

In [96], the following proposition was proven.

Proposition D.1.1. [96] *Any STAP detection algorithm that uses only the random variables G and $\{G(l)\}_1^L$ defined in* (D.3) *for its description such that the algorithm itself is unchanged by arbitrary but equal scaling of all these variables, has a FAP which is independent of the target-free data covariance* $\mathbf{R}$.

Using the transformations $\mathbf{u} = \mathbf{R}^{-1/2}\mathbf{s}$, $\mathbf{y} = \mathbf{R}^{-1/2}\mathbf{x}$, and $\mathbf{y}(l) = \mathbf{R}^{-1/2}\mathbf{x}(l)$, leads to

$$G = \mathbf{u}^{\dagger}\widetilde{\mathbf{R}}^{-1}\mathbf{y} \quad \text{and} \quad G(l) = \mathbf{u}^{\dagger}\widetilde{\mathbf{R}}^{-1}\mathbf{y}(l), \tag{D.5}$$

where $\widetilde{\mathbf{R}} \equiv \mathbf{R}^{-1/2}\widehat{\mathbf{R}}\mathbf{R}^{-1/2}$. The whitened vectors $\mathbf{Y}$ and $\mathbf{Y}(l)$ are both distributed $\mathcal{CN}_N(0, \mathbf{I})$. It turns out that $\widetilde{\mathbf{R}}$ has the complex Wishart distribution[2] $\mathcal{CW}(L, N; \frac{1}{L}\mathbf{I})$. Further, a unitary transformation $\mathbf{U}$ can be found which rotates the new signal vector $\mathbf{u}$ into an elementary vector $\mathbf{e}$ as

$$d\mathbf{e} = \mathbf{U}^{\dagger}\mathbf{u},$$

such that $\mathbf{e} = [1, 0, \ldots, 0]^{\dagger}$ and where $d^2 = \|\mathbf{U}^{\dagger}\mathbf{u}\| = \mathbf{s}^{\dagger}\mathbf{R}^{-1}\mathbf{s}$. The first column of $\mathbf{U}$ is the new signal vector $\mathbf{u}$. The remaining columns comprise an orthonormal basis determined, for example, by a Gram-Schmidt procedure. Let $\mathbf{z} = \mathbf{U}^{\dagger}\mathbf{y}$ and $\mathbf{z}(l) = \mathbf{U}^{\dagger}\mathbf{y}(l)$. Applying these to (D.5) yields the variables

$$G = \frac{d}{L}\mathbf{e}^{\dagger}\mathcal{S}^{-1}\mathbf{z} \quad \text{and} \quad G(l) = \frac{d}{L}\mathbf{e}^{\dagger}\mathcal{S}^{-1}\mathbf{z}(l), \tag{D.7}$$

---

[1] In this Appendix we have reproduced for the reader's convenience certain definitions and transformations from [168], [169], and [96] that are required for the present analysis. We have used several results from these three papers and have attempted to maintain the same notation.

[2] When $\mathbf{X} \sim \mathcal{CN}_N(0, \mathbf{R})$, the (Wishart) matrix $\mathbf{W} = \sum_{l=1}^{L} \mathbf{X}\mathbf{X}^{\dagger} = L\widehat{\mathbf{R}}$ has the complex Wishart distribution $\mathcal{CW}(L, N; \mathbf{R})$ specified by the density

$$f_{\mathbf{W}}(\mathbf{w}) = \begin{cases} \frac{(\det \mathbf{w})^{L-N}}{J(\mathbf{R})} \exp(-\operatorname{tr}(\mathbf{R}^{-1}\mathbf{w})), & \text{if } \mathbf{R} \text{ is positive definite} \\ 0, & \text{otherwise} \end{cases} \tag{D.6}$$

where $J(\mathbf{R}) = \pi^{N(N-1)/2} \prod_{n=1}^{N} \Gamma(L - n + 1)(\det \mathbf{R})^{L}$. The covariance estimate $\widehat{\mathbf{R}}$ is distributed as $\mathcal{CW}(L, N; \frac{1}{L}\mathbf{R})$. If $\mathbf{B}$ is an $N \times N$ nonsingular complex matrix, then $\mathbf{V} = \mathbf{B}^{\dagger}\mathbf{W}\mathbf{B}$ is distributed as $\mathcal{CW}(L, N; \mathbf{B}^{\dagger}\mathbf{R}\mathbf{B})$. See [180] for more on Wishart distributions.

where $\mathcal{S} \equiv L\,\mathbf{U}^\dagger\widetilde{\mathbf{R}}\mathbf{U}$. While $\mathbf{z}$ and $\mathbf{z}(l)$ are distributed as $\mathcal{CN}_N(0,\mathbf{I})$ and are independent, $\mathcal{S}$ has the distribution $\mathcal{CW}(L,N;\mathbf{I})$. The vectors $\mathbf{z}$ and $\mathbf{z}(l)$ are decomposed as

$$\mathbf{z} = \begin{bmatrix}\mathbf{z}_A \\ \mathbf{z}_B\end{bmatrix} \quad \text{and} \quad \mathbf{z}(l) = \begin{bmatrix}\mathbf{z}_A(l) \\ \mathbf{z}_B(l)\end{bmatrix},$$

where the A components are scalar and B components $(N-1)$-vector. Correspondingly, $\mathcal{S}$ is decomposed as

$$\mathcal{S} = \sum_{l=1}^{L} \mathbf{z}(l)\mathbf{z}(l)^\dagger = \begin{bmatrix}\mathcal{S}_{AA} & \mathcal{S}_{AB} \\ \mathcal{S}_{BA} & \mathcal{S}_{BB}\end{bmatrix}, \tag{D.8}$$

with

$$\mathcal{P} \equiv \mathcal{S}^{-1} = \begin{bmatrix}\mathcal{P}_{AA} & \mathcal{P}_{AB} \\ \mathcal{P}_{BA} & \mathcal{P}_{BB}\end{bmatrix}.$$

The entries of $\mathcal{P}$ can be expressed as

$$\begin{aligned}
\mathcal{P}_{AA} &= (\mathcal{S}_{AA} - \mathcal{S}_{AB}\mathcal{S}_{BB}^{-1}\mathcal{S}_{BA})^{-1}, \\
\mathcal{P}_{BA} &= -\mathcal{S}_{BB}^{-1}\mathcal{S}_{BA}\mathcal{P}_{AA}, \\
\mathcal{P}_{AB} &= \mathcal{P}_{BA}^{-1}, \\
\mathcal{P}_{BB} &= \mathcal{S}_{BB}^{-1} + \mathcal{P}_{AA}^{-1}\mathcal{P}_{BA}\mathcal{P}_{AB},
\end{aligned}$$

as shown in [168], page 120. Using these definitions and relations in (D.7) gives

$$G = \frac{d}{L}\,\mathbf{e}^\dagger\mathcal{S}^{-1}\mathbf{z} = \frac{d}{L}\,\mathcal{P}_{AA}(\mathbf{z}_A - \mathcal{S}_{AB}\mathcal{S}_{BB}^{-1}\mathbf{z}_B) = \frac{d}{L}\,\mathcal{P}_{AA}\,y, \tag{D.9}$$

and

$$G(l) = \frac{d}{L}\,\mathbf{e}^\dagger\mathcal{S}^{-1}\mathbf{z}(l) = \frac{d}{L}\,\mathcal{P}_{AA}\big(\mathbf{z}_A(l) - \mathcal{S}_{AB}\mathcal{S}_{BB}^{-1}\mathbf{z}_B(l)\big) = \frac{d}{L}\,\mathcal{P}_{AA}\,y(l), \tag{D.10}$$

where

$$\begin{aligned}
y &\equiv \mathbf{z}_A - \mathcal{S}_{AB}\mathcal{S}_{BB}^{-1}\mathbf{z}_B, \\
y(l) &\equiv \mathbf{z}_A(l) - \mathcal{S}_{AB}\mathcal{S}_{BB}^{-1}\mathbf{z}_B(l).
\end{aligned} \tag{D.11}$$

The AMF test of (D.4) can therefore be written as

$$|y|^2 \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\eta}{L}\sum_{l=1}^{L}|y(l)|^2. \tag{D.12}$$

Conditioned on $\mathbf{z}_B$ and $\{\mathbf{z}_B(l)\}_1^L$ (or, for short, the B-vectors), the random variables Y and $\{Y(l)\}_1^L$ are (in the absence of a target) uncorrelated and Gaussian with zero means and variances that can be calculated[3] as

$$E_B\{|Y|^2\} = 1 + \mathbf{z}_B^\dagger \mathcal{S}_{BB}^{-1} \mathbf{z}_B = 1 + \mathbf{z}_B^\dagger \left(\sum_{l=1}^{L} \mathbf{z}_B(l) \mathbf{z}_B(l)^\dagger \right)^{-1} \mathbf{z}_B, \tag{D.13}$$

using (D.8) in the second step, and

$$E_B\{|Y(l)|^2\} = 1 - \mathbf{z}_B(l)^\dagger \mathcal{S}_{BB}^{-1} \mathbf{z}_B(l), \qquad l = 1, \ldots, L, \tag{D.14}$$

with E_B denoting conditional expectation. Further, the conditional covariance of the variables $Y(l)$ is given by

$$E_B\{Y(k)Y(n)^\star\} = -\,\mathbf{z}_B(n)^\dagger \mathcal{S}_{BB}^{-1} \mathbf{z}_B(k), \quad k \neq n. \tag{D.15}$$

Hence the set of conditionally jointly Gaussian zero mean random variables Y and $\{Y(l)\}_1^L$ have individual variances and covariances that are functions of the random vectors $\mathbf{z}_B$ and $\{\mathbf{z}_B(l)\}_1^L$. The latter are all jointly independent, each being distributed as $\mathcal{CN}_{N-1}(0, \mathbf{I})$. The probability of any event defined on the random variables Y and $\{Y(l)\}_1^L$ in (D.11) can thus be determined by performing an averaging operation over the distributions of $\mathbf{z}_B$ and $\{\mathbf{z}_B(l)\}_1^L$ and this probability will be independent of the data covariance $\mathbf{R}$. This statement is also true for the random variables G and $\{G(l)\}_1^L$ in (D.9) and (D.10) with the caveat that any constant scaling of these variables should leave the event unchanged. The preceding arguments therefore constitute proof of Preposition D.1.1.

From [168] it is known that

$$\sum_{l=1}^{L} |y(l)|^2 \overset{d}{\equiv} \sum_{l=1}^{L-N+1} |w(l)|^2, \tag{D.16}$$

where the $w(l)$ are i.i.d. each with distribution $\mathcal{CN}_1(0,1)$. Moreover, $Y/M_B^{1/2}$ is, conditioned on the B-vectors, also distributed as $\mathcal{CN}_1(0,1)$ where

$$M_B \equiv 1 + \Sigma_B, \tag{D.17}$$

and

$$\Sigma_B = \mathbf{z}_B^\dagger \left(\sum_{l=1}^{L} \mathbf{z}_B(l) \mathbf{z}_B(l)^\dagger \right)^{-1} \mathbf{z}_B. \tag{D.18}$$

---

[3] See pages 121 and 122 of [168].

The test in (D.12) then takes the form

$$u \underset{H_0}{\overset{H_1}{\gtrless}} \eta' \sum_{l=1}^{L-N+1} u(l), \tag{D.19}$$

where $U \equiv |Y|^2/M_B$ and $\{U(l) \equiv |w(l)|^2\}_1^{L-N+1}$ are all unit exponential and i.i.d., and $\eta' \equiv \eta/(LM_B)$. Once again, this is in the form of the usual CA-CFAR test when it is conditioned on the B-vectors. Hence the FAP of the AMF detector can be written as

$$\begin{aligned} \alpha_{\text{AMF}} &= P\Big(U \geq \eta' \sum_{l=1}^{L-N+1} U(l)\Big) \\ &= E\Big\{P\Big(U \geq \eta' \sum_{l=1}^{L-N+1} U(l)\Big| \text{B-vectors}\Big)\Big\} \\ &\triangleq E\{g(\Sigma_B)\}, \end{aligned} \tag{D.20}$$

where

$$g(\Sigma_B) = \frac{1}{[1+\eta/(LM_B)]^{L-N+1}}. \tag{D.21}$$

We can therefore estimate the FAP using the g-method with an IS simulation that biases the B-vectors. This estimator is

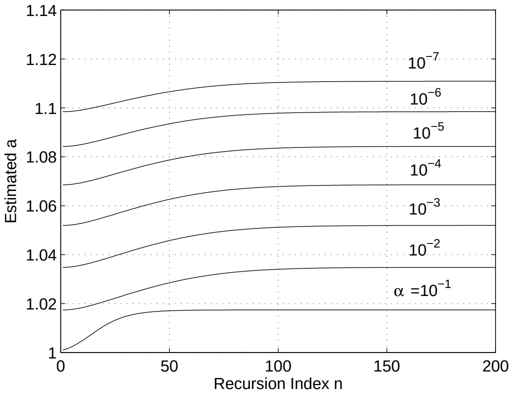
$$\widehat{\alpha}_{\text{AMF}} = \frac{1}{K}\sum_{1}^{K} g(\Sigma_B)\, W(\mathbf{z}_B, \mathbf{z}_{BL}); \quad \sim f_\star, \tag{D.22}$$

where $\mathbf{z}_{BL} \equiv (\mathbf{z}_B(1), \ldots \mathbf{z}_B(L))'$.

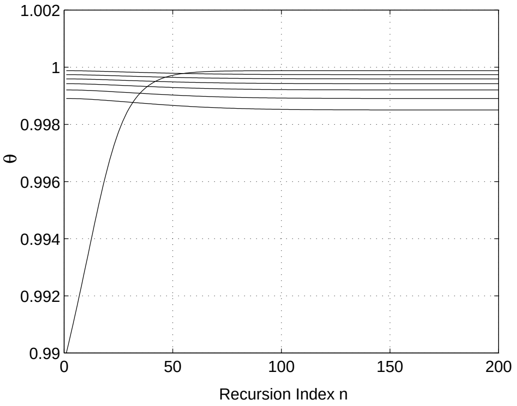
Biasing of the B-vectors contained in Σ_B must produce an increase in the value of the g-function in (D.21). This means that Σ_B must be made to increase, which can easily be accomplished by scaling up the primary B-vector $\mathbf{z}_B$ and scaling down the secondary B-vectors $\mathbf{z}_B(l)$. A 2-d biasing scheme results, which needs to be optimized adaptively. The primary and secondary scaling parameters are chosen as $a^{1/2}$ and $\theta^{1/2}$ respectively, with the weighting function W being the same as in (5.26) except that the roles of the two biasing parameters are exchanged, that is, $0 < \theta \leq 1$ and $a \geq 1$. The rest of the optimization procedure is as discussed in Section 1.2.2.

D.2 Simulation results

Figure D.1 contains the result of the IS optimization for estimating the biasing parameters a and θ for the AMF detector using the 2-d $g-$method estimator described



(a) Optimum scaling for primary B-vectors



(b) Optimum scaling for secondary B-vectors

Figure D.1: Optimum scaling parameters for $g-$method with 2-d biasing for the AMF detector. $K = 10000$, $L = 128$ and $N = 64$.

previously. Figure D.2 shows an example of the 2-d IS gain surface versus the biasing parameters for $\alpha = 10^{-6}$.

In Figure D.3 the gain obtained from the 2-d IS biasing method presented here is compared with the g-method used with input biasing, the technique which was developed in [96]. From this figure it is clear that applying the g-method with biasing of B-vectors is a more powerful IS scheme than the corresponding method used with input biasing of all secondary vectors. This example also illustrates the point made in Section 5.5.

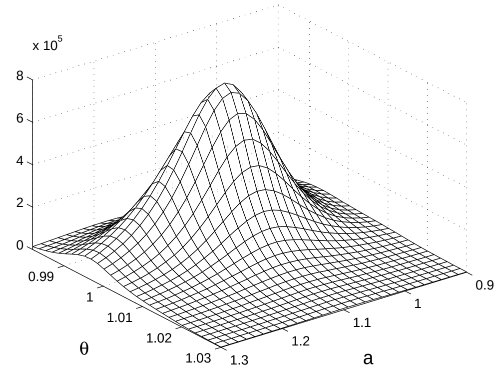


Figure D.2: IS gain surface for 2-d biasing of B-vectors for the AMF detector. $K = 10000$, $\alpha = 10^{-6}$, $L = 128$ and $N = 64$.

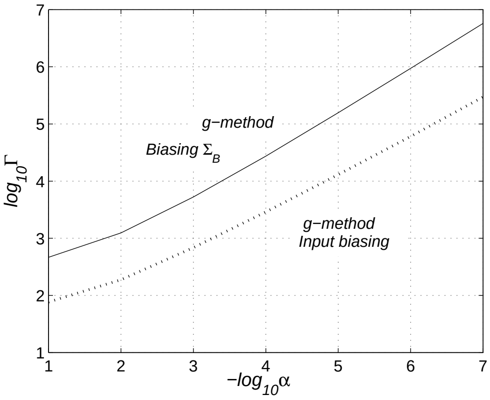


Figure D.3: IS gain comparison of the g-method with 2-d biasing derived here and the g-method with input biasing presented in [96]. For both schemes $K = 10000$, $L = 128$ and $N = 64$.

Appendix E

CFAR property of the E and GM NAMF detectors

The E-NAMF detector is given by

$$\frac{|\mathbf{s}^\dagger \widehat{\mathbf{R}}^{-1}\mathbf{x}|}{\sqrt{\mathbf{x}^\dagger \widehat{\mathbf{R}}^{-1}\mathbf{x}}} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\eta_e}{L}\sum_{l=1}^{L}|\mathbf{s}^\dagger \widehat{\mathbf{R}}^{-1}\mathbf{x}(l)|, \tag{E.1}$$

where η_e denotes the detector threshold.

Using the transformation in Appendix D, the E-NAMF test of (E.1) can be rewritten as

$$\frac{|y|}{\sqrt{\mathbf{x}^\dagger \widehat{\mathbf{R}}^{-1}\mathbf{x}}} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\eta_e}{L}\sum_{l=1}^{L}|y(l)|. \tag{E.2}$$

Then, combining (5.20) and the first line of (5.21) and substituting for $\mathbf{x}^\dagger \widehat{\mathbf{R}}^{-1}\mathbf{x}$ in the above yields the test

$$\frac{|y|}{\sqrt{L\Sigma_B + L|y|^2 \Big/ \sum_{l=1}^{L}|y(l)|^2}} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\eta_e}{L}\sum_{l=1}^{L}|y(l)|. \tag{E.3}$$

So if we replace y and $y(l)$ by G and $G(l)$ as defined in (D.3) respectively in the above, then the test remains unchanged. Therefore, from Proposition D.1.1, it follows that the test is CFAR.

The GM-NAMF detector is defined as

$$\frac{|\mathbf{s}^\dagger \widehat{\mathbf{R}}^{-1}\mathbf{x}|}{\sqrt{\mathbf{x}^\dagger \widehat{\mathbf{R}}^{-1}\mathbf{x}}} \underset{H_0}{\overset{H_1}{\gtrless}} \eta_g\left(\prod_{l=1}^{L}|\mathbf{s}^\dagger \widehat{\mathbf{R}}^{-1}\mathbf{x}(l)|\right)^{1/L}, \tag{E.4}$$

where η_g is the threshold.

Just as for the E-NAMF detector, the GM-NAMF test of (E.4) can be rewritten as

$$\frac{|y|}{\sqrt{\mathbf{x}^\dagger \widehat{\mathbf{R}}^{-1}\mathbf{x}}} \underset{H_0}{\overset{H_1}{\gtrless}} \eta_g \left(\prod_{l=1}^{L} |y(l)| \right)^{1/L}, \tag{E.5}$$

and then as

$$\frac{|y|}{\sqrt{L\Sigma_B + L|y|^2 \Big/ \sum_{l=1}^{L} |y(l)|^2}} \underset{H_0}{\overset{H_1}{\gtrless}} \eta_g \left(\prod_{l=1}^{L} |y(l)| \right)^{1/L}. \tag{E.6}$$

For the same reasons as for the E-NAMF detector, the GM-NAMF test is also CFAR.

Appendix F

Derivation of (5.38)

Equation (5.38) in Chapter 5, page 98 can be derived using the following equalities.

$$
\begin{aligned}
\alpha &= E_B\{E\{g(\Sigma_B, \mathbf{Y}_L)\big|\,\text{B-vec}\}\} \\
&= \int\int g(\Sigma_B, \mathbf{Y}_L)\, f(\mathbf{Y}_L\big|\,\text{B-vec})\, f(\text{B-vec})\, d\mathbf{Y}_L\, d\text{B-vec} \\
&= \int\int g(\Sigma_B, \mathbf{Y}_L)\, f(\mathbf{Y}_L, \text{B-vec})\, d\mathbf{Y}_L\, d\text{B-vec} \\
&= E\{g(\Sigma_B, \mathbf{Y}_L)\} \\
&= E\{g(\Sigma_B, Y(1), \ldots, Y(L))\} \\
&= E\{g(\Sigma_B, Z_A(1), \ldots, Z_A(L))\} \\
&\triangleq E\{g(\Sigma_B, \mathbf{Z}_{AL})\} \\
&= \int\int g(\Sigma_B, \mathbf{z}_{AL})\, f(z_{AL}, \text{B-vec})\, d\mathbf{z}_{AL}\, d\text{B-vec} \\
&= \int\int g(\Sigma_B, \mathbf{z}_{AL})\, f(\mathbf{z}_{AL})\, f(\text{B-vec})\, d\mathbf{z}_{AL}\, d\text{B-vec} \\
&= \int\int g(\Sigma_B, \mathbf{z}_{AL})\, \frac{f(\mathbf{z}_{AL})}{f_\star(\mathbf{z}_{AL})}\, f_\star(\mathbf{z}_{AL})\, f(\text{B-vec})\, d\mathbf{z}_{AL}\, d\text{B-vec} \\
&\triangleq \int\int g(\Sigma_B, \mathbf{z}_{AL})\, W(\mathbf{z}_{AL})\, f_\star(\mathbf{z}_{AL})\, f(\text{B-vec})\, d\mathbf{z}_{AL}\, d\text{B-vec} \\
&= E_\star\{g(\Sigma_B, \mathbf{Z}_{AL})\, W(\mathbf{Z}_{AL})\} \qquad \text{(F.1)}
\end{aligned}
$$

List of Symbols and Acronyms

ADC	Analog to Digital Converter, page 1
AMF	Adaptive Matched Filter, page 17
AMP	Approximate Message Passing, page 48
AOA	Angle Of Arrival, page 23
BER	Bit Error Rate, page 11
BPDN	Basis Pursuit Denoising, page 9
CA	Cell Averaging, page 15
CFAR	Constant False Alarm Rate, page 17
CLT	Central Limit Theorem, page 11
CMF	Compressive Matched Filter, page 38
CPI	Coherent Processing Interval, page 86
CS	Compressive Sensing, page 2
CUT	Cell Under Test, page 2
DFT	Discrete Fourier Transform, page 7
DOA	Direction Of Arrival, page 23
E	Envelope, page 19
FAP	False Alarm Probability, page 3
GCV	Generalized Cross Validation, page 34
GLRT	Generalized Likelihood Ratio Test, page 24

GM	Geometric Mean, page 19
i.i.d.	independent and identically distributed, page 6
IHT	Iterative Hard Thresholding, page 9
IS	Importance Sampling, page 3
ISL	Integrated Sidelobe Level, page 26
IST	Iterative Soft Thresholding, page 33
KS	Kolmogorov-Smirnov, page 61
LASSO	Least Absolute Shrinkage and Selection Operator, page 9
LFM	Linear Frequency Modulation, page 24
LR	Low Rank, page 19
LS	Least Square, page 7
MC	Monte Carlo, page 3
MF	Matched Filter, page 23
MP	Matching Pursuit, page 9
MSE	Mean Square Error, page 34
NAMF	Normalized Adaptive Matched Filter, page 19
NMF	Normalized Matched Filter, page 19
OS	Order Statistic, page 27
PC	pulse compression, page 24
pdf	probability density function, page 3
PN	pseudorandom noise, page 24
PRF	Pulse Repetition Frequency, page 87
RCS	Radar Cross Section, page 22
RIP	Restricted Isometry Property, page 5
RV	random variable, page 11

RX	receive or receiver, page 74
SE	State Evolution, page 51
SF	Stepped Frequency, page 19
SLL	sidelobes level, page 26
SNR	Signal-to-Noise Ratio, page 10
SPGℓ_1	Spectral Projected Gradient ℓ_1, page 34
STAP	Space Time Adaptive Processing, page 1
SURE	Stein's unbiased risk estimator, page 34
TX	transmit or transmitter, page 74

Bibliography

[1] E. T. Whittaker. On the functions which are represented by the expansions of the interpolation theory. *Proc. Roy. Soc. Edinburgh, Sec. A*, 35:181–194, 1915.

[2] H. Nyquist. Certain topics in telegraph transmission theory. *Trans. AIEE*, 47:617–644, Apr. 1928.

[3] C. E. Shannon. Communication in the presence of noise. *Proc. Institute of Radio Engineers (IRE)*, 37(1):10–21, Jan. 1949.

[4] E. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inf. Theory*, 52(2):489–509, Feb. 2006.

[5] M. Rangaswamy, J. H. Michels, and B. Himed. Statistical analysis of the non-homogeneity detector for STAP applications. *J. Digital Signal Process.*, 14(3):253–267, May 2004.

[6] R. Srinivasan. Some results in importance sampling and an application to detection. *J. Signal Process.*, 65(1):73–88, Feb. 1998.

[7] D. L. Donoho. Compressed sensing. *IEEE Trans. Inf. Theory*, 52(4):1289 – 1306, Apr. 2006.

[8] R. Baraniuk. Compressive sensing. *IEEE Signal Process. Mag.*, 24(4):118–121, Jul. 2007.

[9] E. Candès and M. Wakin. An introduction to compressive sampling. *IEEE Signal Process. Mag.*, 25(2):21–30, Mar. 2008.

[10] E. Candès and T. Tao. Near optimal signal recovery from random projections: Universal encoding strategies? *IEEE Trans. Inf. Theory*, 52(12):5406–5425, Dec. 2006.

[11] E. Candès, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.*, 59(8):1207–1223, Aug. 2006.

[12] A. Cohen, W. Dahmen, and R. DeVore. Compressed sensing and best k-term approximation. *J. Amer. Math. Soc.*, 22:211–231, 2009.

[13] D. S. Taubman and M. W. Marcellin. *JPEG2000 image compression fundementals, standards and practice.* Kluwer Academic Publishers, 2002.

[14] E. Candès and T. Tao. Decoding by linear programming. *IEEE Trans. Inf. Theory*, 51(12):4203–4215, Dec. 2005.

[15] D. L. Donoho and X. Huo. Uncertainity principles and ideal atomic decomposition. *IEEE Trans. Inf. Theory*, 47(7):2845–2862, Nov. 2001.

[16] H. Taylor, S. Banks, and J. McCoy. Deconvolution with ℓ_1-norm. *Geophysics Magazine*, 44(1):39–52, 1979.

[17] S. Levy and P. Fullagar. Reconstruction of a sparse spike train from a portion of its spectrum and application to high resolution deconvolution. *Geophysics Magazine*, 46(9):1235–1243, 1981.

[18] F. Santosa and W.W. Symes. Linear inversion of band-limited seismograms. *SIAM J. on Sci. Computing*, 7(4):1307–1330, 1986.

[19] E. Candès and J. Romberg. Sparsity and incoherence in compressive sampling. *Inverse Problems*, 23(3):969–985, 2007.

[20] G. Reeves and M. Gastpar. Difference between observation and sampling error in sparse signal reconstruction. In *Proc. IEEE Work. Stat. Signal Process.*, 2007.

[21] J. Laska, P. Boufounos, M. Davenport, and R. Baraniuk. Democracy in action: quantization, saturation, and compressive sensing. *Appl. Comput. Harmon. Anal.*, 31(3):429–443, Nov. 2011.

[22] C. Studer, P. Kuppinger, G. Pope, and H. Bölcskei. Recovery of sparsely corrupted signals. *IEEE Trans. Inf. Theory*, 58(5):3115–3130, May 2012.

[23] J. Treichler, M. Davenport, and R. Baraniuk. Application of compressive sensing to the design of wideband signal acquisition receivers. In *U.S./Australia Joint Work. Defense Apps. of Signal Proces. (DASP)*, 2009.

[24] E. Arias-Castro and Y.C. Eldar. Noise folding in compressed sensing. *IEEE Signal Process. Lett.*, 18(8):478–481, Aug. 2011.

[25] M. A. Davenport, J. N. Laska, J. R. Treichler, and R. Baraniuk. The pros and cons of compressive sensing for wideband signal acquisition: Noise folding vs. dynamic range. Preprint 2011.

[26] R. Tibshirani. Regression shrinkage and selection via the LASSO. *J. Roy. Stat. Soc., Series B*, 58(1):267–288, 1996.

[27] S.S. Chen, D.L. Donoho, and M.A. Saunders. Atomic decomposition by basis pursuit. *SIAM J. on Sci. Computing*, 20(1):33–61, 1998.

[28] E. van den Berg and M. P. Friedlander. Probing the Pareto frontier for basis pursuit solutions. *SIAM J. on Sci. Computing*, 31(2):890–912, 2008.

[29] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani. Least angle regression. *Ann. Stat.*, 32(2):407–499, 2004.

[30] S. J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky. A method for large-scale ℓ_1-regularized least squares. *IEEE J. Sel. Topics Sig. Proc.*, 1(4):606–617, Dec. 2007.

[31] A. Beck and M. Teboulle. A fast iterative shrinkage thresholding algorithm for linear inverse problems. *SIAM J. Imag. Sci.*, 2(1):183–202, 2009.

[32] S. Becker, J. Bobin, and E. Candès. NESTA: a fast and accurate first-order method for sparse recovery. *SIAM J. Imag. Sci.*, 4(1):1–39, 2009.

[33] M. Figueiredo, R. Nowak, and S. Wright. Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. *IEEE J. Sel. Topics Sig. Proc.*, 1(4):586–598, 2007.

[34] I. Loris. On the performance of algorithms for the minimization of ℓ_1-penalized functionals. *Inverse Problems*, 25(3), 2009.

[35] E. Hale, W. Yin, and Y. Zhang. Fixed point continuation method for ℓ_1-minimization with application to compressed sensing. *Rice University Technical Report TR07-07*, 2007.

[36] W. Yin, S. Osher, D. Goldfarb, and J. Darbon. Bregman iterative algorithms for ℓ_1-minimization with applications to compressed sensing. *SIAM J. Imag. Sci.*, 1(1):143–168, 2008.

[37] A. Maleki, L. Anitori, Y. Zai, and R. Baraniuk. Asymptotic analysis of complex LASSO via complex approximate message passing (CAMP). Submitted to *IEEE Trans. Inf. Theory*, 2011.

[38] D. L. Donoho and Y. Tsaig. Fast solution of ℓ_1-norm minimization problems when the solution may be sparse. *IEEE Trans. Inf. Theory*, 54(11):4789–4812, 2008.

[39] J. A. Tropp. Just relax: Convex programming methods for identifying sparse signals. *IEEE Trans. Inf. Theory*, 52(3):1030–1051, Mar. 2006.

[40] Z. Yang, C. Zhang, J. Deng, and W. Lu. Orthonormal expansion ℓ_1-minimization algorithms for compressed sensing. *IEEE Trans. Signal Process.*, 59(12):6285–6290, Dec. 2011.

[41] T. Blumensath and M. Davies. Iterative thresholding for sparse approximations. *J. Fourier Anal. Applicat.*, 2008.

[42] W. Dai and O. Milenkovic. Subspace pursuit for compressive sensing signal reconstruction. *IEEE Trans. Inf. Theory*, 55(5):2230–2249, May 2009.

[43] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Comm. Pure Appl. Math.*, 75:1412–1457, 2004.

[44] D. L. Donoho, I. Drori, Y. Tsaig, and J. L. Starck. Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit. Technical report, Standford Univ., Statistic Dept., 2006.

[45] S. Mallat and S. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Trans. Signal Process.*, 41(12):3397–3415, Dec. 1993.

[46] D. Malioutov, M. Cetin, and A.S. Willsky. A sparse signal reconstruction perspective for source localization with sensor arrays. *IEEE Trans. Signal Process.*, 53(8):3010–3022, Aug. 2005.

[47] D. Needell and R. Vershynin. Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit. *IEEE J. Sel. Topics Sig. Proc.*, 4(2):310–316, Apr. 2009.

[48] D. Needell and J. A. Tropp. CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. *Appl. Comput. Harmon. Anal.*, 26(3):301–321, 2008.

[49] J. A. Tropp. Greed is good: Algorithmic results for sparse approximation. *IEEE Trans. Inf. Theory*, 50(10):2231–2242, Oct. 2004.

[50] J. A. Tropp and A. C. Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Trans. Inf. Theory*, 53(12):4655–4666, Dec. 2007.

[51] A. C. Gilbert J. A. Tropp and M. Strauss. Algorithms for simultaneous sparse approximation. Part I: Greeedy pursuit. *J. Signal Process.*, 86(3):572–588, 2006.

[52] T. Blumensath and M. E. Davies. How to use the iterative hard thresholding algorithm. *Proc. Work. Struc. Parc. Rep. Adap. Signaux (SPARS)*, 2009.

[53] R. Baraniuk and T. P. H. Steeghs. Compressive radar imaging. *Proc. IEEE Radar Conf.*, 2007.

[54] M. A. Herman and T. Strohmer. High-resolution radar via compressed sensing. *IEEE Trans. Signal Process.*, 57(6):2275–2284, 2009.

[55] A. C. Gurbuz, J. H. McClellan, and W. R. Scott. Compressive sensing for subsurface imaging using ground penetrating radar. *J. Signal Process.*, 89(10):1959–1972, Oct. 2009.

[56] S. Bhattacharya, T. Blumensath, B. Mulgrew, and M. Davies. Fast encoding of synthetic aperture radar raw data using compressed sensing. In *Proc. IEEE Work. Stat. Signal Process.*, 2007.

[57] K. R. Varshney, M. Cetin, J. W. Fisher, and A. S. Willsky. Sparse representation in structured dictionaries with application to synthetic aperture radar. *IEEE Trans. Signal Process.*, 56(8):3548–3561, Aug. 2008.

[58] Y. Wang, G. Leus, and A. Pandharipande. Direction estimation using compressive sampling array processing. In *Proc. IEEE Work. Stat. Signal Process.*, 2009.

[59] J. H. G. Ender. On compressive sensing applied to radar. *J. Signal Process.*, 90(5):1402–1414, May 2010.

[60] Y. Yu, A. P. Petropulu, and H. V. Poor. MIMO radar using compressive sensing. *IEEE J. Sel. Topics Sig. Proc.*, 4(1):146–163, Feb. 2010.

[61] S. Shah, Y. Yu, and A. Petropulu. Step-frequency radar with compressive sampling SFR-CS. In *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Process. (ICASSP)*, 2010.

[62] J. J. Fuchs. The generalized likelihood ratio test and the sparse representations approach. In *Proc. Int. Conf. Image and Signal Process. (ICISP)*, 2010.

[63] L. C. Potter, E. Ertin, J. T. Parker, and M. Cetin. Sparsity and compressed sensing in radar imaging. *Proc. IEEE*, 98(6):1006–1020, Jun. 2010.

[64] A. C. Fannjiang, T. Strohmer, and P. Yan. Compressed remote sensing of sparse objects. *SIAM J. Imag. Sci.*, 3(3):595–618, Sept. 2010.

[65] C.R. Berger, S. Zhou, P. Willett, B. Demissie, and J. Heckenbach. Compressed sensing for OFDM/MIMO radar. In *Proc. Asilomar Conf. Signals, Systems, and Comput.*, Oct. 2008.

[66] I. Stojanovic, W. C. Karl, and M. Cetin. Compressed sensing of mono-static and multi-static SAR. In *Proc. SPIE Algorithms for Synthetic Aperture Radar Imagery XVI*, 2009.

[67] C. Y. Chen and P. P. Vaidyanathan. Compressed sensing in MIMO radar. In *Proc. Asilomar Conf. Signals, Systems, and Comput.*, 2008.

[68] A. Budillon, A. Evangelista, and G. Schirinzi. SAR tomography from sparse samples. In *Proc. IEEE Int. Symp. Geosci. and Remote Sens. (IGARSS)*, 2009.

[69] A. Budillon, A. Evangelista, and G. Schirinzi. Three-dimensional SAR focusing from multipass signals using compressive sampling. *IEEE Trans. Geosci. Remote Sens.*, 49(1):488–499, Jan. 2011.

[70] G. E. Smith, T. Diethe, Z. Hussain, J. Shawe-Taylor, and D. R. Hardoon. Compressive sampling for pulse doppler radar. In *Proc. IEEE Radar Conf.*, 2010.

[71] X. X. Zhu and R. Bamler. Tomographic SAR inversion by ℓ_1-norm regularization - the compressive sensing approach. *IEEE Trans. Geosci. Remote Sens.*, 48(10):3839–3846, Oct. 2010.

[72] X. X. Zhu and R. Bamler. Super-resolution for 4-D SAR tomography via compressive sensing. In *Proc. European SAR Conf. (EuSAR)*, 2010.

[73] N. S. Subotic, B. Thelen, K. Cooper, W. Buller, J. Parker, J. Browning, and H. Beyer. Distributed radar waveform design based on compressive sensing considerations. In *Proc. IEEE Radar Conf.*, 2008.

[74] N. S. Subotic J. Parker M. Stuff, B. Thelen and J. Browning. Optimization and waveforms for compressive sensing applications in the presence of interference. In *Proc. IEEE Int. Conf. Waveform Diversity and Design*, 2009.

[75] L. Anitori, M. Otten, and P. Hoogeboom. Compressive sensing for high resolution radar imaging. In *Proc. IEEE Asia-Pacific Microwave Conf. (APMC)*, 2010.

[76] L. Anitori, A. Maleki, M. Otten, R. Baraniuk, and P. Hoogeboom. CS radar imaging via adaptive CAMP. In *Proc. European SAR Conf. (EuSAR)*, 2012.

[77] L. Anitori, F. Le Chevalier, M. Otten, and P. Hoogeboom. Compressive sensing for high resolution profiles with enhanced doppler performance. *Proc. European Radar Conf. (EuRAD)*, 2012.

[78] B. Bubacarr and J. Tanner. Improved bounds for restricted isometry constants for Gaussian matrices. *SIAM J. Matrix Anal.*, 31(5):2882–2898, 2010.

[79] P. P. Gandhi and S.A. Kassam. Analysis of CFAR processors in homogeneous background. *IEEE Trans. Aerosp. Electron. Syst.*, 24(4):427–445, Jul. 1988.

[80] G. B. Goldstein. False alarm regulation in log-normal and Weibull clutter. *IEEE Trans. Aerosp. Electron. Syst.*, 9(1):84–92, Jan. 1973.

[81] E. Conte, M. Lops, and A. M. Tulino. Hybrid procedure for CFAR in non-Gaussian clutter. *IEE Proc. Radar, Sonar, and Navig.*, 144(6):361–369, Dec. 1997.

[82] R. Ravid and N. Levanon. Maximum-likelihood for Weibull background. *IEE Proc. Part F (London)*, 139(3):256–264, Jun. 1992.

[83] H. Rohling. Radar CFAR thresholding in clutter and multiple target situations. *IEEE Trans. Aerosp. Electron. Syst.*, 19(4):608–621, Jul. 1983.

[84] R. S. Raghavan. Analysis of CA-CFAR processors for linear-law detection. *IEEE Trans. Aerosp. Electron. Syst.*, 28(3):661–665, Jul. 1992.

[85] V. G. Hansen. Studies of logarithmic radar receiver using pulse-length discrimination. *IEEE Trans. Aerosp. Electron. Syst.*, 1(3):246–253, Dec. 1965.

[86] R. Nitzberg. Constant-false-alarm-rate signal processors for several types of interference. *IEEE Trans. Aerosp. Electron. Syst.*, 8(1):27–34, Jan. 1972.

[87] V. G. Hansen and H. R. Ward. Detection performance of cell averaging LOG/CFAR receiver. *IEEE Trans. Aerosp. Electron. Syst.*, 8(5):648–652, Sept. 1972.

[88] M. Sekine, T. Musha, Y. Tomita, and T. Irabu. Suppression of weibull-distributed clutters using a cell-averaging LOG/CFAR receiver. *IEEE Trans. Aerosp. Electron. Syst.*, 14(5):823–826, Sept. 1978.

[89] M. A. Khalighi and M. H. Bastani. Adaptive CFAR processor for nonhomogeneous environments. *IEEE Trans. Aerosp. Electron. Syst.*, 36(3):889–897, Jul. 2000.

[90] R. L. Mitchell. Importance sampling applied to simulation of false alarm statistics. *IEEE Trans. Aerosp. Electron. Syst.*, 17(1):15–24, Jan. 1981.

[91] K. R. Gerlach. New results in importance sampling. *IEEE Trans. Aerosp. Electron. Syst.*, 35(3):917–925, Jul. 1999.

[92] J. A. Bucklew. *Introduction to rare event simulation.* Springer, New York, 2004.

[93] R. Srinivasan. *Importance Sampling - Applications in Communications and Detection.* Springer-Verlag, Berlin, 2002.

[94] R. Srinivasan. Simulation of CFAR detection algorithm for arbitrary clutter distribution. *IEE Proc. Radar, Sonar, and Navig.*, 147(1):31–40, Feb. 2000.

[95] J. A. Bucklew. Conditional importance sampling estimators. *IEEE Trans. Inf. Theory*, 51(1):143–153, Jan. 2005.

[96] R. Srinivasan and M. Rangaswamy. Importance sampling for characterizing STAP detectors. *IEEE Trans. Aerosp. Electron. Syst.*, 43(1):273–285, Jan. 2007.

[97] B.O. Steenson. Detection performance of a mean-level threshold. *IEEE Trans. Aerosp. Electron. Syst.*, 4(4):529–534, Jul. 1968.

[98] S. R. Babu and R. Srinivasan. Analysis of envelope detected mean level CFAR processors using importance sampling. In *Proc. IEEE Radar Conf.*, 2000.

[99] K. R. Gerlach. New results in importance sampling. *IEEE Trans. Aerosp. Electron. Syst.*, 35(3):917–925, Jul. 1999.

[100] D. L. Stadelman, D. D. Weiner, and A. D. Keckler. Efficient determination of thresholds via importance sampling for monte carlo evaluation of radar performance in non-gaussian clutter. In *Proc. IEEE Radar Conf.*, 2002.

[101] R. Srinivasan and M. Rangaswamy. Fast estimation of false alarm probabilities of STAP detectors - the AMF. In *Proc. IEEE Radar Conf.*, 2005.

[102] M. Rangaswamy, F. C. Lin, and K. R. Gerlach. Robust adaptive signal processing methods for heterogeneous radar clutter scenarios. *J. Signal Process.*, 84(9):1653–1665, Sept. 2004.

[103] M. Rangaswamy and F. C. Lin. Normalized adaptive matched filter-a low rank approach. In *Proc. IEEE Symp. on Signal Process. and Inf. Technology (ISSPIT)*, 2003.

[104] S. M. Song, W. M. Kim, D. Park, and Y. Kim. Estimation theoretic approach for radar pulse compression processing and its optimal codes. *Electronic Letters*, 36(3):250–253, Feb. 2000.

[105] S. D. Blunt and K. Gerlach. Adaptive pulse compression via MMSE estimation. *IEEE Trans. Aerosp. Electron. Syst.*, 42(2):572–583, Apr. 2006.

[106] M. I. Skolnik. *Radar Handbook.* McGraw-Hill, New York, 1970.

[107] M. A. Richards. *Fundamentals of Radar Signal Processing.* McGraw-Hill, New York, 2005.

[108] J. V. DiFranco and W. L. Rubin. *Radar detection.* Artech House, Dedham, 1980.

[109] H. L. Van Trees. *Detection, estimation and Modulation theory: Part I.* John Wiley & Sons, New York, 2001.

[110] H. L. Van Trees. *Detection, estimation and Modulation theory: Part III.* John Wiley & Sons, New York, 2001.

[111] L. R. Varshney and D. Thomas. Sidelobe reduction for matched filter range processing. In *Proc. IEEE Radar Conf.*, 2003.

[112] S. M. Kay. *Fundamentals of Statistical Singnal Processing: Detection Theory.* Prentice-Hall, Upper Saddle River, 1998.

[113] P. Stoica, J. Li, and M. Xue. Transmit codes and receive filters for radar. *IEEE Signal Process. Mag.*, 25(6):94–109, Nov. 2008.

[114] F. J. Harris. On the use of windows for harmonic analysis with the discrete fourier transform. *Proc. IEEE*, 66(1):51–83, Jan. 1978.

[115] F. F. Jr. Kretschmer and K. Gerlach. Low sidelobe radar waveforms derived from orthogonal matrices. *IEEE Trans. Aerosp. Electron. Syst.*, 27(1):92–102, Jan. 1991.

[116] R. L. Frank. Polyphase codes with good nonperiodic correlation properties. *IEEE Trans. Inf. Theory*, 9(1):43–45, 1963.

[117] A. Divito, A. Farina, G. Fedele, G. Galati, and F. Studer. Synthesis and evaluation of phase codes for pulse compression radar. *Rivista Tecnica Selenia*, 9(2):12–24, 1985.

[118] Y. I. Abramovich and M. B. Sverdlik. Synthesis of a filter which maximizes the signal-to-noise radio under additional quadratic constraints. *Radio Eng. Electron. Phys.*, 15:1977–1984, Nov. 1970.

[119] M. H. Ackroyd and F. Ghani. Optimum mismatched filters for sidelobe suppression. *IEEE Trans. Aerosp. Electron. Syst.*, 9(2):214–218, Mar. 1973.

[120] S. Zoraster. Minimum peak range sidelobe filters for binary phase-coded waveforms. *IEEE Trans. Aerosp. Electron. Syst.*, 16(1):112–115, Jan. 1980.

[121] A. Zejak, E. Zentner, and P. Rapajic. Doppler optimised mismatched filters. *IET Electronics Letters*, 27(7):558–560, Mar. 1991.

[122] C. Candan. On the design of mismatched filters with an adjustable matched filtering loss. In *Proc. IEEE Radar Conf.*, 2010.

[123] A. De Maio, Y. Huang, M. Piezzo, S. Zhang, and A. Farina. Design of radar receive filters optimized according to ℓ_p-norm based criteria. *IEEE Trans. Signal Process.*, 59(8):40234029, Aug. 2011.

[124] S. Verdú. Computational complexity of optimum multiuser detection. *Algorithmica*, 4:303–312, 1989.

[125] H. M. Finn and R. S Johnson. Adaptive detection mode with threshold control as a function of spatially sampled clutter-level estimates. *RCA Review*, 29:414–464, Sept. 1968.

[126] G. M. Dillard and C. E. Antoniak. A practical distribution-free detection procedure for multiple-range-bin radar. *IEEE Trans. Aerosp. Electron. Syst.*, 6(5):629–635, Sept. 1970.

[127] A. Di Vito and G. Moretti. Probability of false alarm in CA-CFAR device downstream from linear-law detector. *IET Electronics Letters*, 25(25):1692–1693, Dec. 1989.

[128] http://dsp.rice.edu/cs.

[129] J. Ender. Sparse representation and autofocus for ISAR and spotlight SAR imaging. In *Challenges in SAR Work., IPAM,UCLA*, Feb. 2012.

[130] L. Anitori, M. Otten, and P. Hoogeboom. Detection performance of compressive sensing applied to radar. In *Proc. IEEE Radar Conf.*, 2011.

[131] F. Belfiori, L. Anitori, W. van Rossum, M. Otten, and P. Hoogeboom. Digital beam forming and compressive sensing based DOA estimation in MIMO arrays. *Proc. European Radar Conf. (EuRAD)*, 2011.

[132] V. Cevher, M. F. Duarte, and R. Baraniuk. Distributed target localization via spatial sparsity. In *Proc. European Sig. Pro. Conf. (EUSIPCO)*, 2008.

[133] C. Feng, S. Valaee, and Z. Tan. Multiple target localization using compressive sensing. In *Proc. IEEE Global Telecom. Conf. (GLOBECOM)*, 2009.

[134] A. C. Gurbuz, J. H. McClellan, and W. R. Scott. A compressive sensing data acquisition and imaging method for stepped frequency GPRs. *IEEE Trans. Signal Process.*, 57(7):26402650, Jul. 2009.

[135] J. A. Tropp, J.N. Laska, M. F. Duarte, J.K.Romberg, and R. Baraniuk. Beyond nyquist: Efficient sampling of sparse bandlimited signals. *IEEE Trans. Inf. Theory*, 56(1):520–544, Jan. 2010.

[136] L. Anitori, A. Maleki, W. van Rossum, R. Baraniuk, and M. Otten. Compressive CFAR radar detection. In *Proc. IEEE Radar Conf.*, 2012.

[137] D. L. Donoho, A. Maleki, and A. Montanari. Message passing algorithms for compressed sensing. *Proc. Natl. Acad. Sci.*, 106(45):18914–18919, 2009.

[138] D. L. Donoho. For most underdetermined systems of linear equations, the minimal ℓ_1-norm near-solution approximates the sparsest near-solution. *Comm. Pure Appl. Math*, 59:797–829, 2004.

[139] S. Boyd and L. Vandenberghe. *Convex optimization.* Cambridge University Press, New York, 2004.

[140] A. Maleki and D. L. Donoho. Optimally tuned iterative thresholding algorithm for compressed sensing. *IEEE J. Sel. Topics Sig. Proc.*, 4(2):330–341, Apr. 2010.

[141] O. Batu and M. Cetin. Hyper-parameter selection in non-quadratic regularization-based radar image formation. In *Proc. SPIE*, 2008.

[142] O. Batu and M. Cetin. Parameter selection in sparsity-driven SAR imaging. *IEEE Trans. Aerosp. Electron. Syst.*, 47(4):3040–3050, Oct. 2011.

[143] P. Boufounos, M. F. Duarte, and R. Baraniuk. Sparse signal reconstruction from noisy compressive measurements using cross validation. In *Proc. IEEE Work. Stat. Signal Process.*, 2007.

[144] C. H. Seng, A. Bouzerdoum, S. L. Phung, and M. Amin. Automatic parameter selection for feature-enhanced radar image restoration. In *Proc. IEEE Radar Conf.*, 2010.

[145] E. Candès and T. Tao. The Dantzig selector: statistical estimation when p is much larger than n. *Ann. Stat.*, 35:2392–2404, 2007.

[146] M. Elad D. L. Donoho and V. Temlyakov. Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Trans. Inf. Theory*, 52(1):6–18, Jan. 2006.

[147] M. F. Duarte, M. A. Davenport, M. B. Wakin, and R. Baraniuk. Sparse signal detection from incoherent projections. In *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Process. (ICASSP)*, 2006.

[148] J. Haupt and R. Nowak. Compressive sampling for signal detection. In *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Process. (ICASSP)*, 2007.

[149] M. A. Davenport, M. B. Wakin, and R. Baraniuk. The compressive matched filter. Technical Report TREE-0610, Rice Univ., ECE Dept., Nov. 2006.

[150] M. A. Davenport, P. T. Boufounos, M. B. Wakin, and R. Baraniuk. Signal processing with compressive measurements. *IEEE J. Sel. Topics Sig. Proc.*, 4(2):445–460, Apr. 2010.

[151] Z. Wang, G. R. Arce, and B. M. Sadler. Subspace compressive detection for sparse signals. In *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Process. (ICASSP)*, 2008.

[152] J. L. Paredes, Z. Wang, G. R. Arce, and B. M. Sadler. Compressive matched subspace detection. In *Proc. European Sig. Pro. Conf. (EUSIPCO)*, 2009.

[153] B. Pollock and N. A. Goodman. Detection performance of compressively sampled radar signals. In *Proc. IEEE Radar Conf.*, 2011.

[154] L. Anitori, M. Otten, and P. Hoogeboom. False alarm probability estimation for compressive sensing radar. In *Proc. IEEE Radar Conf.*, 2011.

[155] Y. Chi, A. Pezeshki, L. Scharf, and R. Calderbank. Sensitivity to basis mismatch in compressed sensing. *IEEE Trans. Signal Process.*, 59(5):2182–2195, May 2011.

[156] H. Zhu, G. Leus, and G. B. Giannakis. Sparse regularized total least-squares for sensing applications. In *Proc. of Work. on Sig. Pro. Advances in Wireless Comm.(SPAWC)*, 2010.

[157] D. L. Donoho, A. Maleki, and A. Montanari. Message passing algorithms for compressed sensing: I. motivation and construction. In *IEEE Proc. Inform. Theory Work. (ITW)*, 2010.

[158] A. Maleki and A. Montanari. Analysis of approximate message passing algorithm. In *Proc. IEEE Conf. Inform. Science and Systems (CISS)*, 2010.

[159] D. L. Donoho, A. Maleki, and A. Montanari. The noise-sensitivity phase transition in compressed sensing. *IEEE Trans. Inf. Theory*, 57(10):6920–6941, Oct. 2011.

[160] A. Maleki. *Approximate message passing algorithm for compressed sensing.* PhD thesis, Stanford University, 2010.

[161] M. Bayati and A. Montanari. The LASSO risk for Gaussian matrices. *IEEE Trans. Inf. Theory*, 58(4):1997–2017, Apr. 2012.

[162] M. Bayati and A. Montanari. The dynamics of message passing on dense graphs, with applications to compressed sensing. *IEEE Trans. Inf. Theory*, 57(2):764–785, Feb. 2011.

[163] F. J. Massey. The Kolmogorov-Smirnov test for goodness of fit. *J. Amer. Statist. Assoc.*, 46(253):68–78, 1951.

[164] D. P. Meyer and H. A. Mayer. *Radar target detection: handbook of theory and practice.* Academic Press Inc., New York, NY, 1973.

[165] L. Anitori, R. Srinivasan, and M. Rangaswamy. Importance sampling for NMF class of stap detectors. In *Proc. IEEE Int. Radar Conf. (CIE)*, 2006.

[166] R. Klemm. *Principles of space-time adaptive processing.* Institution of Engineering and Technology, Berlin, 2002.

[167] I. S. Reed, J. D. Mallett, and L. E. Brennan. Rapid convergence rate in adaptive arrays. *IEEE Trans. Aerosp. Electron. Syst.*, 10(6):853–863, Nov. 1974.

[168] E. J. Kelly. An adaptive detection algorithm. *IEEE Trans. Aerosp. Electron. Syst.*, 22(1):115–127, Mar. 1986.

[169] F. C. Robey, D. R. Fuhrmann, E. J. Kelly, and R. Nitzberg. A CFAR adaptive matched filter detector. *IEEE Trans. Aerosp. Electron. Syst.*, 28(1):208–216, Jan. 1992.

[170] M. Rangaswamy. An overview of space-time adaptive processing for radar. In *Proc. IEEE Radar Conf.*, 2003.

[171] J. H. Michels, M. Rangaswamy, and B. Himed. Performance of parametric and covariance based STAP tests in compound-gaussian clutter. *J. Digital Signal Process.*, 12(2-3):307–328, Nov. 2002.

[172] D. Schleher. Harbor surveillance radar detection performance. *IEEE J. Ocean. Eng.*, 2(4):318–325, Oct. 1977.

[173] L. Anitori, R. Srinivasan, and M. Rangaswamy. Envelope-law and geometric-mean stap detection. *IEEE Trans. Aerosp. Electron. Syst.*, 46(1):184–192, Jan. 2010.

[174] L. Anitori and R. Srinivasan. Advanced robust STAP algorithms and fast performance evaluation techniques based on rare event theory. Technical report, Twente Univ., ECE Dept., Mar. 2007.

[175] I. Kirsteins and D. Tufts. Adaptive detection using a low rank approximation to a data matrix. *IEEE Trans. Aerosp. Electron. Syst.*, 30(1):55–67, Jan. 1994.

[176] P. Chen, M. C. Wicks, and R. S. Adve. Development os a statistical procedure for detecting the number of signals in a radar measurement. *IEE Proc. Radar, Sonar, and Navig.*, 148(4):219–226, Aug. 2001.

[177] M. S. Srivastava and C. G. Khatri. *An introduction to multivariate statistics.* Elsevier North Holland, New York, 1979.

[178] M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* National Bureau of Standards, U.S. Dept. of Commerce, 1972.

[179] I. S. Gradshteyn and I. M. Ryzhik. *Table of Integrals, Series, and Products. Fifth Edition.* Academic Press, London, 1994.

[180] D. Maiwald and D. Kraus. Calculation of moments of complex Wishart and complex inverse Wishart distributed matrices. *IEE Proc. Radar, Sonar, and Navig.*, 147(4):162–168, Aug. 2000.

Summary

Compressive Sensing and Fast Simulations: applications to radar detection

In most modern high-resolution multi-channel radar systems one of the major problems to deal with is the huge amount of data to be acquired, processed and/or stored. But why do we need all these data? According to the well known Nyquist-Shannon sampling theorem, real signals have to be sampled at at least twice the signal bandwidth to prevent ambiguities. Therefore, sampling of very wide bandwidths may require Analog to Digital Converter (ADC) hardware that is unavailable or very expensive; especially in multi-channel systems, the cost and power consumption can become critical factors. In applications involving interleaving of radar modes in time or space (antenna aperture), multi-function operation often leads to conflicting requirements on sampling rates in both time and spatial domains. So while, on one hand, the increased number of degrees of freedom improves the system performance, on the other hand it puts a significant burden on both the off-line analysis and performance evaluation of sophisticated detectors, and on the real time acquisition and processing. For example, space-time adaptive processing algorithms significantly enhance the detection of targets buried in noise, clutter and jamming. However, evaluating the optimal filter weights is an immense computational load when simulating such detectors in the design phase as well as in real time implementation. In some cases, measurement time may also be a constraint, as in 3D radar imaging for airport security inspection of passengers. Conventional acquisition of a full 3D high resolution image requires a measurement time that can be unacceptable in this situation.

In this thesis we investigate sampling methods that can deal with the problems of processing complexity as well as analysis (or performance evaluation) extremely efficiently by reducing the required amount of samples. By cleverly using properties of the signals or random variables involved, the considered techniques, namely Compressive Sensing (CS) and Importance Sampling (IS), both alleviate the burden related to data handling in complex radar detectors. These methods, although very different in nature, provide an alternative to classical sampling techniques. The first, compressive sensing, is based on a revolutionary acquisition and processing theory that enables reconstruction of sparse signals from a set of measurements sampled at a much lower rate than required by the Nyquist-Shannon theorem. This results in both shorter acquisition time and reduced amount of data. The second, importance sampling, has roots in statistical physics and represents a fast and effective method for the design and analysis of detectors whose performance have to be evaluated by simulations. By efficiently sampling the underlying probability density function, importance sampling provides a very fast alternative to conventional Monte Carlo simulation.

The first part of the thesis deals with the design and analysis of adaptive detectors for compressive sensing based radars. In systems using compressive sensing, the target signal, which is assumed to be sparse, is estimated from the noisy, undersampled measurements via ℓ_1-norm minimization algorithms. CS recovery algorithms require proper setting of parameters (thresholds) and are therefore not inherently adaptive. Classical radar systems employ a matched filter, matched to the transmitted waveform, followed by a Constant False Alarm Rate (CFAR) processor for the detection of targets embedded in unknown background clutter and noise. However, the non-linearity introduced by a CS recovery algorithm does not allow straightforward application of conventional adaptive detector design methodology. In the work reported here, by making use of the properties of the Complex Approximate Message Passing algorithm, we are able to propose an adaptive non-linear recovery stage combined with classical CFAR processing, and derive a novel adaptive CS detector. Additionally, our theoretical findings are also demonstrated via both simulated and experimental results. Furthermore, we provide a methodology to predict the performance of the proposed detectors that can be used to evaluate how transmitted power can be traded against undersampling, making it possible to incorporate CS-based sampling and detection in radar system design.

The second part of this thesis focuses on deriving methods of importance sampling for fast simulation of rare events especially applicable to Space Time Adaptive Processing (STAP) radar detectors. These type of methods are, however, of much wider applicability. They can and have been used in many other situations that require intensive and time-consuming Monte Carlo simulations. In conducting rare event simulations of systems that involve signal processing operations that are mathematically complex, there are two principal issues that contribute to simulation time. The first issue concerns the rare event itself whose probability is being sought. The second concerns the computational intensity that accompanies the signal processing. It is a daunting task to conduct conventional Monte Carlo simulations that involve several millions of trials to estimate low false alarm probabilities, with as many matrix inversions, as required in STAP. We demonstrate how fast simulation schemes can deal with these aspects, and devise tailored importance sampling biasing schemes

for evaluating the performance of STAP detectors which are analytically difficult or impossible to analyze, such as low rank STAP detectors. By comparing our results with traditional Monte Carlo methods, we show that importance sampling can achieve tremendous gain in terms of computational time.

Samenvatting

Compressive Sensing and fast simulations: applications to radar detection

In de meeste moderne hoge resolutie- , meerkanaals radarsystemen is het omgaan met grote hoeveelheden data die vergaard en verwerkt moeten worden een probleem van betekenis. Waarom hebben we zoveel data nodig? Welnu, volgens het algemeen bekende Shannon-Nyquist bemonsteringstheorema moeten, om dubbelzinnigheden te voorkomen, rele signalen bemonsterd worden met een frequentie die minstens tweemaal de bandbreedte van het gemeten signaal is. Daardoor vraagt het bemonsteren van breedbandige signalen om vaak dure of zelfs niet beschikbare Analoog-Digitaal omzetters. Vooral in meerkanaals radarsystemen kunnen de kosten en stroomverbruik kritieke beperkende factoren vormen. In bepaalde toepassingen, zoals het verweven van verschillende radar modes in tijd of spatieel (antenne-)domein, leidt multi-functionele operatie soms tot onderling conflicterende eisen ten aanzien van bemonstering in tijd en/of ruimte. Terwijl enerzijds het toegenomen aantal vrijheidsgraden de prestaties van een meerkanaalssysteem kunnen verbeteren, legt het anderzijds een zware last op zowel de acquisitie en verwerking van meetgegevens in 'real-time' als op de analyse van prestaties van geavanceerde detectiemechanismen. Bijvoorbeeld: Space-Time Adaptive Processing (STAP) algoritmen kunnen de detectie van doelen die verborgen zijn in ruis, achtergrondreflecties en storing aanzienlijk verbeteren. Het berekenen van optimale filtercofficienten is echter een enorme rekentaak, zowel bij het simuleren van de detectieprestaties als bij het daadwerkelijk implementeren in real-time. In sommige toepassingen is meettijd ook een kritieke beperking, zoals bijvoorbeeld bij het 3D-scannen van personen voor luchthavenbeveiliging; acquisitie van een volledige 3D beeld op conventionele wijze duurt onacceptabel lang in dat geval.

Voor dit proefschrift zijn bemonsteringsmethoden onderzocht die bovenstaande problemen ten aanzien van processingcomplexiteit en datahoeveelheid efficient kunnen oplossen of verminderen, door het aantal benodigde signaalbemonsteringen ('samples') te reduceren. Door slim gebruik te maken van de signaaleigenschappen of van

de eigenschappen van stochastische variabelen, kunnen de onderzochte methoden, te weten Compressive Sampling en Importance Sampling, de last van dataverwerking in complexe radardetectoren aanzienlijk verlichten. Deze twee methoden, die onderling zeer verschillend van aard zijn, bieden een alternatieve aanpak voor klassieke bemonsteringsmethoden. De eerst methode, Compressive Sensing, afgekort CS, is gebaseerd op een revolutionaire theorie over data-aquisitie- en verwerking, die het mogelijk maakt om de reconstructie van zogenaamde 'sparse' signalen te bewerkstelligen uit een set van meetgegevens die vergaard is met een veel lagere bemonsteringsfrequentie dan volgens het Shannon-Nyquist theorema nodig is. Dit leidt tot een kortere meettijd en tot een kleinere hoeveelheid data. De tweede methode, Importance Sampling, afgekort IS, komt oorspronkelijk uit de Statistische Fysica en is een snelle en efficinte methode voor het door middel van simulatie analyseren en evalueren van detectoren, met name wanneer een dergelijke evaluatie niet met analytische methoden kan worden uitgevoerd. Door de onderliggende kansdichtheidsfuncties slim te bemonsteren, biedt IS een veel sneller alternatief voor zware conventionele Monte Carlo simulaties.

Het eerste deel ven het proefschift behandelt de de analyse en ontwerp van adaptieve radardetectoren gebaseerd op Compressive Sampling. In systemen die op CS gebaseerd zijn, wordt het doelsignaal, waarvan we aannemen dat het 'sparse' is, geschat uit een onderbemonsterde set van ruizige meetgegevens door middel van l1-minimalisatiealgoritmen. Deze reconstructiealgoritmen bevatten parameters die op een zeer specifieke wijze moeten worden afgesteld, waardoor dit proces niet van zichzelf adaptief is. Conventionele radarsystemen maken gebruik van een zogenaamd Matched Filter, dat is aangepast aan de uitgezonden golfvorm, gevolgd door een adaptieve Constant False Alarm Rate (CFAR) detector, om detectie van doelen in onbekende ruis- en interferentiecondities mogelijk te maken. CS-reconstructie-algoritmen introduceren echter niet-lineair gedrag waardoor de standaard CFAR ontwerpmethodologie niet zonder meer kan worden toegepast. In het hier gepresenteerde werk wordt gebruik gemaakt van de eigenschappen van het Complex Message Passing Algoritme, waardoor het mogelijk is om een niet-lineair reconstructie algoritme toch te combineren met een klassieke CFAR detector en zodoende een nieuwe adaptieve CS detector te realiseren. De theoretische bevindingen zijn bovendien bevestigd geworden door gesimuleerde- en experimentele resultaten. Daarboven wordt een ontwerpmethodologie aangereikt waarbij de detectieprestaties kunnen worden voorspeld en waarbij het mogelijk wordt om afwegingen te maken ten aanzien van elektrisch vermogen versus onderbemonstering. Hiermee is het mogelijk geworden om CS-methoden direct te betrekken in het ontwerp van radarsystemen en -detectoren.

Het tweede deel van dit proefschrift is toegespitst op het afleiden van IS methoden voor het snel simuleren van 'events' zoals valse detecties, die met zeer lage waarschijnlijkheid en dus zeer weinig voorkomen en in het bijzonder de toepassing daarvan op STAP detectoren. IS methoden zijn echter veel breder toepasbaar. Dit type van methoden is reeds toegepast in vele andere toepassingen waarbij intensieve en tijdrovende Monte-Carlo simulaties benodigd zijn. Bij het simuleren van onwaarschijnlijke 'events' in signaalverwerkingssystemen die mathematisch complex zijn, zijn er twee aspecten die de simulatietijd in hoge mate bepalen. Ten eerste is het de lage waarschijnlijkheid van de event zelf. Ten tweede is het de complexiteit van (reken-)bewerkingen die voor de signaalverwerking nodig zijn. Daardoor is het een lastige taak om Monte Carlo simulaties uit te voeren bestaande uit miljoenen simulatie-runs om vals-alarmwaarschijnlijkheid te bepalen van STAP detectoren, waarbij o.a. evenzovele matrixinversies moeten worden uitgevoerd. We laten zien hoe dit met snelle simulatiemethoden kan worden aangepakt en leiden IS 'biasing' methoden af, die zijn toegesneden op de evaluatie van STAP detectoren die moeilijk via analytische weg te evalueren zijn, zoals Low-Rank STAP detectoren. Door de resultaten te vergelijken met traditionele Monte Carlo simulaties wordt aangetoond dat met IS zeer grote winst in rekentijd kan worden geboekt.

About the author

Laura Anitori was born in Ronciglione, Italy, on May 18, 1980. In September 2000 she started her Telecommunication Engineering studies at the University of Pisa, Italy. She received the Master of Science degree in Telecommunication Engineering from the same university in April 2005 (summa cum laude). She graduated with a thesis on high resolution ISAR imaging entitled 'Development of a Clean Technique for high resolution Inverse Synthetic Aperture Radar (ISAR) imaging', under the supervision of Prof. E. Dalle Mese, Prof. F. Berizzi, and Dr. M. Martorella.

From August 2005 to April 2007 she worked as a research assistant in the telecommunication engineering group at the Technical University of Twente, Enschede, The Netherlands. During this period she worked under the supervision of Prof. R. Srinivasan on the development and application of adaptive importance sampling techniques for estimating false alarm probabilities of radar detectors that use space-time adaptive processing algorithms.

In the period May 2007-December 2009 she has been with the Radar group at TNO Defence, Security and Safety, The Hague, The Netherlands. Here she continued working on radar signal processing in both national and international projects.

Since 2010, in a joint cooperation between TNO and the Technical University of Delft, she started working on her Ph.D. degree on the subject of compressive sensing based radar detection. During her Ph.D. work, in April 2011, she visited the digital signal processing group at Rice University, Houston, Texas, where she worked in cooperation with Prof. R. Baraniuk and Dr. A. Maleki.

She is a student member of IEEE and serves as reviewer for IEEE journals (Aerospace and Electronic Systems and Signal Processing), IET and Elsevier journal of Signal Processing. She was the first author of the paper that received the second prize at the best student paper award at the 2012 IEEE Radar Conference, Atlanta.

Author's publications

Journal Papers

J1 L. Anitori, A. Maleki, M. Otten, R. Baraniuk, and P. Hoogeboom. Design and analysis of compressive sensing radar detectors. To appear on *IEEE Trans. Signal Process.*, Sept. 2012.

J2 A. Maleki, L. Anitori, Z. Yang, and R. Baraniuk. Asymptotic analysis of complex LASSO via Complex Approximate Message Passing (CAMP). Submitted to *IEEE Trans. Inf. Theory*, Aug. 2011. Available at `http://arxiv.org/abs/1108.0477`.

J3 L. Anitori, R. Srinivasan, and M. Rangaswamy. Envelope-law and geometric-mean STAP detection. *IEEE Trans. Aerosp. Electron. Syst.*, 46(1):184-192, Jan. 2010.

Conference Papers

C1 L. Anitori, F. Le Chevalier, M. Otten, and P. Hoogeboom. Compressive sensing for high resolution profiles with enhanced Doppler performance. In *Proc. European Radar Conf. (EuRAD)*, 2012.

C2 L. Anitori, A. Maleki, M. Otten, R. Baraniuk, and P. Hoogeboom. Compressive CFAR radar detectors. In *Proc. Compressive Sensing Radar Work. (CoSeRa)*, 2012.

C3 L. Anitori, A. Maleki, M. Otten, W. van Rossum, and R. Baraniuk. Compressive CFAR radar detection. In *Proc. IEEE Radar Conf.*, 2012. (**Awarded with the second prize at the best student paper competition.**)

C4 L. Anitori, A. Maleki, M. Otten, R. Baraniuk, and P. Hoogeboom. CS radar imaging via adaptive CAMP. In *Proc. European SAR Conf. (EUSAR)*, 2012.

C5 L. Anitori, M. Otten, and P. Hoogeboom. Detection performance of compressive sensing applied to radar. In *Proc. IEEE Radar Conf.*, 2011.

C6 L. Anitori, M. Otten, and P. Hoogeboom. False alarm probability estimation for compressive sensing radar. In *Proc. IEEE Radar Conf.*, 2011.

C7 F. Belfiori, L. Anitori, W. van Rossum, M. Otten and P. Hoogeboom. Digital beam forming and compressive sensing based DOA estimation in MIMO arrays. In *Proc. European Radar Conf. (EuRAD)*, 2011.

C8 J. J. M. de Wit, L. Anitori, W. L. van Rossum, and R. G. Tan. Radar mapping of buildings using sparse reconstruction with an overcomplete dictionary. In *Proc. European Radar Conf. (EuRAD)*, 2011.

C9 L. Anitori, M. Otten, and P. Hoogeboom. Compressive sensing for high resolution radar imaging. In *Proc. IEEE Asia-Pacific Microwave Conf. (APMC)*, 2010.

C10 M. Otten, N. Maas, R. Bolt, and L. Anitori. Light weight digital array SAR. In *Proc. IEEE Int. Symp. Phased Array Systems and Technology (ARRAY)*, 2010.

C11 L. Anitori, M. Otten, and P. Hoogeboom. Radar detection via compressive sensing using partial Fourier coefficients. In *Proc. Radar Methods and Systems Work. (RMSW)*, 2010.

C12 L. Anitori, D. Deiana, and J. S. F. Kerkkamp. Detection and classification of low metal antipersonnel landmines using ground penetrating radar. In *Proc. ANTEM/AMEREM*, 2010.

C13 D. Deiana, and L. Anitori. Detection and classification of landmines using AR modeling of GPR data. In *Proc. IEEE Int. Conf. Ground Penetrating Radar (GPR)*, 2010.

C14 L. Anitori, and A. J. E. Smith. Satellite automatic identification system SAT-AIS for vessel traffic monitoring in the North sea. In *Proc. MAST Conf.*, 2009.

C15 B. vd Broek, G. Burghouts, S. vd Broek, R. Hagen, L. Anitori, W. van Rossum, and A. J. E. Smith. Automatic detection of hostile behaviour. In *Proc. SPIE Unmanned/Unattended Sensors and Sensor Networks VI*, 2009.

C16 L. Anitori, A. de Jong and F. Nennie. FMCW radar for life-sign detection. In *Proc. IEEE Radar Conf.*, 2009.

C17 A. Gabban, H. Greidanus, A. J. E. Smith, L. Anitori, F. X. Thoorens and J. Mallorqui. Ship Surveillance with TERRASAR-X Scansar. In *Proc. TerraSAR-X Science Team Meeting*, 2008.

C18 A. J. E. Smith, L. Anitori, A. Gabban, H. Greidanus and J. Mallorqui. Inference of vessel behaviour in the Dutch coastal North sea. In *Proc. Fraunhofer Symp. Future Security*, 2008.

C19 L. Anitori, R. Srinivasan and M. Rangaswamy. Performance of low-rank STAP detectors. In *Proc. IEEE Radar Conf.*, 2008.

C20 A. J. de Jong and L. Anitori. Life-sign detection with FMCW radar. In *Proc. NATO Symp. Sensors and Technology for Defence Against Terrorism*, 2008.

C21 L. Anitori, R. Srinivasan and M. Rangaswamy. Importance sampling for NMF class of STAP detectors. In *Proc. IEEE Int. Radar Conf. (CIE)*, 2006.

C22 L. Anitori, and R. Srinivasan. Importance Sampling for STAP detection. *URSI Benelux Meeting*, 2006.

Technical Reports

R1 L. Anitori, and R. Srinivasan. Advanced Robust STAP Algorithms and Fast Performance Evaluation Techniques based on Rare Event Theory. *Technical Report, Twente University, EEMCS Dept., The Netherlands*, Mar. 2007. `www.dtic.mil/dtic/tr/fulltext/u2/a494889.pdf`

R2 A. J. Schoolderman, J. B. Rhebergen, L. Anitori, and A. Ranada Shaw. Effects of the soil on the performance of metal detectors and ground penetrating radars for landmine detection. *Technical Report, TNO The Hague, The Netherlands*, Jul. 2009.

R3 M. P. G. Otten, A. P. M. Maas, R. J. Bolt, L. Anitori, and W. L. van Rossum. Light weight FMCW SAR. *Technical Report, TNO The Hague, The Netherlands*, Nov. 2011.

Acknowledgements

A PhD is like a travel around the world. You meet people from different places, and everyone of them has his/her own cultural background, his/her own opinion and something to share. Like when you travel to a new country, you look at how people live there, what are their habits and costumes, what do they eat. In my PhD trip I have met many different people and have learned something or a lot from each one of them. I had many opportunities to express my opinion, listen to theirs and bring something back home. There are many people who have contributed to my PhD adventure, all in different ways, and I would like to take this opportunity to thank them all for helping me making this happen.

Following a chronological order, I should first thank my first promotor, Prof. Rajan Srinivasan. Rajan, you have been truly a master for me. Demanding and firm, but at the same time rewarding when deserved. I have learned from you what research is all about, 'the fire in the belly' as you once told me. Most of all, I have learned from you about scientific ethic and moral. I will take your teachings with me throughout my scientific career.

Many thanks also to Dr. Rangaswamy, a scientist I admire for his vast knowledge and his modesty. You always had good words for me and motivated me to pursue my PhD, even when I was very discouraged from 'external events'...Thank you.

After my years at Twente University I came at TNO, The Hague. Here, not only I met a lot of good scientists, but also made very good friends. There are quite a few people that I have to thank for making this PhD possible. Without your support, I could never have finished it. Many thanks to Kemo Agovic, Frank van den Bogaart, Lucas van Ewijk, Dr. Arnold Schoolderman, and Dr. Arthur Smith, who have financially supported the second part of this work. Thanks also to my promotor, Prof. Peter Hoogeboom, for his support in helping me reach the end of this trip. Furthermore, I would like to thank Prof. Frank van Vliet and, again, Dr. Arthur Smith for proof reading the first draft of my thesis.

Arthur, I would like to especially thank you not only for the financial support, but also for all the troubles you went through to arrange all what was necessary. A

big part of this work is because of the effort you put in it, and not just for the money. As you always say, 'you need to be able and willing'! You have been and are not only a colleague, but a very good friend. You helped me when I needed it, listened to me when I was upset, and gave me advise when I was uncertain about what to do. Thanks for everything.

Many thanks also to Matern Otten, who has been my daily supervisor, not just during my PhD, but ever since I entered TNO. I hope to be one day as good as you are. Thanks for listening, helping, discussing and encouraging me. It was great working with you all these years. Many thanks also to Wim van Rossum, who has always been available to answer my million questions and helped me when I was stuck, especially with some phase terms.... Thanks also to all my colleagues of the radar group at TNO. I have learned from you all, personally and scientifically. Many warm thanks also to my friends of the 'Italian gang': Annalisa, Daniele, Giampiero, Lorenzo, Stefania and my paranynphs Daniela and Francesco. For the things that I miss of Italy, especially family, language and coffee, you make me feel back home, even if we are all in The Netherlands.

I would also like to thank Dr. Geert Leus, Prof. Frans Groen, and Prof. Francois Le Chevalier for the many conversations and support, especially at times when I felt lonely and stuck with my research.

Many thanks to Prof. Joachim Ender, who has also supervised me during these last two years. He welcomed me several times at Fraunhofer FHR, and found time to have meetings with me in his extremely busy agenda. Thanks for always giving me positive feedbacks. They have helped me to keep going. Also, thank you for allowing me to perform the first CS radar measurements with the Fraunhofer Labrador system.

I am also very grateful to Prof. Richard Baraniuk, who welcomed me in his DSP group at Rice University. I am very much in debt to you, as the main results on CS and CAMP I have developed at Rice University in cooperation with Dr. Arian Maleki. Arian, thanks a lot for the wonderful collaboration and for all the patience you had with me, when I asked you the same questions on AMP over and over again. I keep doing that even now, and you always take time to explain me. You are one of the best persons I have ever worked with. Not only you are very knowledgeable and perfectionist in your work, but you are also always available and open to others.

Finally, I would like to thank my family, mamma e papà, and Jurriaan, to whom I dedicate this work.

Jurriaan, I met you as soon as I moved to The Hague and since then we have always been together. Even in my best dreams, I could never imagine to meet a person like you to share my life with. I apologize for the last months when I was writing my thesis, and I was totally absent and always concentrated on my own work. Not only you understood and let it be, but even made sure that I would find something to eat when I came home very late from work! Thanks for making my life beautiful, every single day.

And finally, last in the list, but first in my heart, I want to thank my parents. Cari mamma e papà, grazie per avermi insegnato a non mollare mai, a combattere ed essere forte, a non arrendermi alla prima difficolta'. Grazie per avermi lasciato libera di scegliere e di sbagliare, e per aver perdonato i miei sbagli. Grazie per essermi sempre vicini, anche se siamo lontani. Se non fosse stato per voi, non avrei mai raggiunto questo traguardo. Vi voglio immensamente bene.