TARGET IDENTIFICATION USING LATE-TIME RADAR RETURNS

RESONANCE-BASED TARGET IDENTIFICATION USING LATE-TIME RADAR RETURNS

By MIHAIL S. GEORGIEV, M.Sc.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy

McMaster University

DOCTOR OF PHILOSOPHY (2025)

Hamilton, Ontario, Canada (Electrical and Computer Engineering)

TITLE:	Resonance-Based Target Identification Using Late-Time Radar Returns
AUTHOR:	Mihail S. Georgiev M.Sc.(Hons.) (Computational Science and Engineering), Technische Universität München, Munich, Germany B.Eng. (Electrical Engineering), McMaster University, Hamilton, Canada
SUPERVISOR:	Timothy N. Davidson
NUMBER OF PAGES:	xv, 157

Lay Abstract

A radar can cause a conductive target to resonate, and it can record these resonances. Much like musical instruments can be identified by their sound, targets can be identified by their resonances. Simple methods to estimate resonances are robust when the resonances are much stronger than the radio noise, but that is rarely the case in practice. Fortunately, in many applications, the measurements can be repeated because the resonant behaviour is consistent. Estimates from each measurement are unreliable, but there is potential for statistical analysis of the repeated measurements to yield reliable aggregate estimates. This thesis develops three methods for doing so and demonstrates their performance. A method for jointly processing multiple measurements is developed in Chapter 2 and demonstrated on synthetic signals. In addition to noise, there can be environmental clutter resonating at the same time as a target. A method for estimating the target's resonances separately from those of the clutter is developed in Chapter 3. A method for classifying (identifying) a target that is embedded in clutter is developed in Chapter 4. The performance of the latter two methods is demonstrated on measurements of handheld weapons.

Abstract

In time-domain radar, the resonant features of a target are contained in its late-time response (LTR). The frequencies and attenuation rates (i.e., *complex frequencies*) of these resonances can be viewed as inherent features of the target. Thus, they could, at least in principle, be used to describe and identify it. Unfortunately, methods for estimating these complex frequencies from a single observation window of the LTR are not robust to noise. That is a concern because the LTR is an attenuating phenomenon that is quite weak to begin with, and hence, the effective signal-to-noise ratio (SNR) over an LTR observation window is low. As a result, the potential for target identification using the resonances in the LTR has yet to be robustly realized in practice in the nearly 35 years since it was identified. This thesis suggests new approaches to processing the LTR that provide more robust realizations of the underlying principles, and demonstrates their performance in physical experiments on an indoor ultra-wideband radar.

The premise for the thesis begins with the observation that it is now possible to design time-domain radars that have a rather high pulse repetition rate, allowing them to quickly capture many measurement shots. Averaging to improve the effective SNR can be employed, but this alone is insufficient for robustness. Statistical analysis of the estimates of the complex frequencies is more effective, but it is hampered because

there is no known expression for the distribution of these frequencies.

In Chapter 2, a technique is developed for estimating the complex frequencies via a related distribution of the z-transform roots of the observed LTRs, which has a known expression when the noise is Gaussian. The technique is demonstrated to be effective in Gaussian noise and in the presence of non-Gaussian uncertainties such as sampling jitter, provided the uncertainty is "approximately Gaussian."

In has been shown that, for Gaussian noise, the maximum-likelihood estimator for the complex frequencies requires the solution of a nonlinear least-squares problem. Since finding the globally optimal solution to that problem is inherently difficult, Chapter 3 presents an efficient method for finding good estimates of the complex frequencies that is based on their empirical distribution. The method is applied to measurements of handheld weapons in the presence of environmental clutter.

Chapter 4 tackles the problem of using LTR measurements to classify targets. The proposed method employs empirical distributions of the estimates of the targets' complex frequencies, rather than employing specific estimates. The method is again applied to measurements of handheld weapons in the presence of environmental clutter and shown to be highly effective.

For Dr. Jian-Kang Zhang

Acknowledgements

I would like to thank my first supervisor, Dr. Jian-Kang Zhang, who supported me until his passing, and my current supervisor, Dr. Timothy N. Davidson, who stepped in and has supported me unwaveringly since. I can only be grateful for the time and effort he put in to enable my research even though it lies somewhat outside his area of expertise.

I would like to thank the team at the Electromagnetic Vision Research Laboratory and their industrial partners, without whose support, the practical aspects of my research would not have been possible. In particular, Dr. Natalia K. Nikolova allowed me to work with her team, and gave me access to access to a prototype radar without which my research would not have been possible. I would thus like to single out Aaron D. Pitcher and Charl Baard, the main designers of the radar, and thank them for both implementing the radar itself and for helping me with my experiments.

Finally, I would like to thank my friends and family for their support.

Table of Contents

List of Figures

List of Tables

Declaration of Academic Achievement

The contributions to knowledge described in this thesis include both contributions to the theory and the set of available algorithms that are available for the processing of late-time radar returns, and demonstrations of the performance of those algorithms in physical experiments.

The development of the theory and algorithms was inspired by the observation made by my former supervisor, Dr. Jian-Kang Zhang, that the insights of Barone [8] might be applicable to the joint processing of multiple late-time radar returns. That observation formed the backbone of the theoretical and algorithmic contributions reported in this thesis. Unfortunately, Dr. Zhang passed away before he could see much of that work realized.

The practical aspects of this work have been enabled by a unique ultra-wideband (UWB) equivalent-time-sampling radar with a high pulse repetition rate that was designed and built in the Electromagnetic Vision (EMVi) Laboratory at McMaster University. At the heart of that radar is a control board that utilizes a field-programmable gate array (FPGA) to control a custom RF front-end. The FPGA is responsible for triggering a pair of UWB transmitters and a dual-channel equivalent-time sampling

receiver. Integrated into this FPGA is a central processing unit (CPU), which is responsible for offloading data from the radar to a PC. Aaron D. Pitcher developed most of the hardware, wrote the Verilog code for the FPGA design, and the programmed the CPU to enable access to the sampled data. Pitcher and I collaborated in creating high-performance data-offloading software. Then, in consultation with Pitcher, I designed and implemented a software suite [70] for real-time digital signal processing (DSP) and analysis. This, in turn, enabled him to debug the hardware design of the radar receiver. Due to our extensive collaboration in the design of the radar, Pitcher is a co-author on all of my publications that involve the use of the radar, and I am a co-author on his publications involving its receiver.

This thesis is presented in a "sandwich-like" format. The key academic achievements are captured in reproductions of papers that have been published or in manuscripts that are currently under review.

The first academic achievement reported in this thesis is the theoretical and algorithmic development reported in Chapter 2. This work is on effective methods for estimating the resonances of a radar target from joint processing of multiple measurement shots of the target, the effectiveness of which is demonstrated through experiments on synthetic data. It has been published as:

> M. Georgiev and T. N. Davidson, "Estimating resonances in low-SNR late-time radar returns with sampling jitter," *IEEE Trans. Signal Process.*, vol. 72, pp. 4651–4665, 2024.

It is reproduced here, in a different format, with the implicit permission that is integral to the IEEE's copyright policy. That publication also includes 1 page of supplementary material that is published on the IEEE's online library known as

IEEE Xplore. That supplementary material has also been included in Chapter 2. I conceived the main ideas in this chapter, refined those ideas in consultation with Dr. Davidson, and I wrote the first draft of the manuscript. Dr. Davidson and I had intended to include Dr. Zhang as a co-author of this paper to recognize his initial observations, described above, but he had passed away before the paper was written, and hence this was not permitted by the IEEE's policies. We formally acknowledged Dr. Zhang's contributions in the published work, and that acknowledgement has been retained within the chapter.

The second achievement is reported in Chapter 3. It concerns the development of a technique for estimating a target's resonances in the presence of environmental clutter, the effectiveness of which is demonstrated by physical experiments. In December 2024, this chapter was submitted for possible publication in the *IEEE Transactions on Radar Systems*, with authors M. S. Georgiev, A. D. Pitcher and T. N. Davidson, and the title "Multi-Shot Estimation of Resonance Parameters of Late-Time Radar Returns in Clutter." I envisioned the architecture for the algorithm and refined its structure in consultation with Dr. Davidson. I performed all the experiments on the custom-designed radar that was built with the help of Aaron Pitcher, as described above, and I implemented all the algorithms for processing the signals received by the radar. I also wrote the first draft of the manuscript.

The third achievement is reported in Chapter 4. It pertains to the development of an algorithm, as well as some requisite theory, for the classification of radar targets based on the empirical distribution of resonances. Using physical experiments, this algorithm is shown to be effective even in the presence of environmental clutter. I envisioned the overall theoretical framework and the architecture for the algorithm,

and refined its structure in consultation with Dr. Davidson. I performed all the experiments on the custom-designed radar that was built with the help of Aaron Pitcher, as described above, and I implemented all the algorithms for processing the signals received by the radar. I also wrote the first draft of the manuscript. This chapter has been published as:

> M. S. Georgiev et al., "Classification of radar targets via distribution matching of late-time resonance parameters," *IEEE Trans. Radar Syst.*, vol. 3, pp. 645–655, 2025.

Beyond the academic achievements described in this thesis, my contributions to the development of the software components of the custom-built ultra-wideband radar have been recognized in the following publications:

- M. Georgiev et al., "Novel noise-tolerant method for extracting target resonances using pulse radar," in *Proc. 17th European Radar Conf.*, 2021, pp. 258–261
- A. D. Pitcher et al., "Correcting timebase errors in ultra-wideband equivalent-time sampling receivers," in *Proc. 21st European Radar Conf.*, Paris, 2024, pp. 47–50
- A. D. Pitcher et al., "Parallelized field-programmable gate array data processing for high-throughput pulsed-radar systems," *Sensors*, vol. 25, no. 1, p. 239, 2025

and the following technical reports (excluding those subsumed by the above publications):

- C. Baard and M. Georgiev, "Methods for detecting target in range and system-level noise estimation," *Electromagn. Vision Res. Lab.*, McMaster University, Hamilton, ON, Tech. Rep. EMVi-R-93, Apr. 2020

- M. Georgiev, "Interference filtering and correction of systemic timebase distortion in an equivalent-sampling receiver," *Electromagn. Vision Res. Lab.*, McMaster University, Hamilton, ON, Tech. Rep. EMVi-R-96, Jun. 2020

- M. Georgiev and J. Nguyen, "System noise and signal-to-noise ratio analysis of the UWB equivalent-time-sampling radar," *Electromagn. Vision Res. Lab.*, McMaster University, Hamilton, ON, Tech. Rep. EMVi-R-104, Nov. 2020

- J. Nguyen et al., "Stage 3 technical progress report to DASA," *Electromagn. Vision Res. Lab.*, McMaster University, Hamilton, ON, Tech. Rep. EMVi-R-108, Apr. 2021

- A. D. Pitcher et al., "Unobtrusive inspection for on-body threats concealed under clothing: Concealed weapon detection using polarization energy ratios," *Electromagn. Vision Res. Lab.*, McMaster University, Hamilton, ON, Tech. Rep. EMVi-R-110, Sep. 2021

- A. D. Pitcher and M. Georgiev, "The digital signal processing pipelines (DSPPL) suite," *Electromagn. Vision Res. Lab.*, McMaster University, Hamilton, ON, Tech. Rep. EMVi-R-121, Sep. 2024

- A. D. Pitcher and M. Georgiev, "Equivalent-time sampling ultra-wideband pulsed-radar receiver: RF front-end design," *Electromagn. Vision Res. Lab.*, McMaster University, Hamilton, ON, Tech. Rep. EMVi-R-122, Sep. 2024

Chapter 1

Introduction

1.1 Background on Radar-Based Detection and Estimation

Radar technology was conceived for the detection and ranging of targets using radio [81]. A signal is transmitted, transformed by the environment, and then received. By comparing the received signal to the transmitted one, one can learn about this transformation, and through analysis of the transformation, one can learn about the environment. As a simple example, if a conductive target is placed in the environment a certain distance from the radar, and the radar transmits a pulse[1], some portion of that pulse will reflect off the target and be received by the radar. If the target is moved farther away, as a general trend, the received pulse will be delayed, and in this way, the target can be ranged.

However, the restriction to ranging overlooks most of the information contained

[1]That is, the radar stimulates its transmitting antenna in such a way as to produce a spatially concentrated radio wave such as a plane wave.

in the received signal. For example, one can estimate the cross-sectional size of a target from the energy of the received pulse because larger targets reflect more of the pulse [18], and one can estimate its depth from how much the pulse is smeared because only a flat target will reflect all of the pulse at once [1]. It stands to reason that, through more sophisticated analysis of the received signal, one can glean more information about the environment than just the distance to a target of interest; e.g., [15]. In particular, it is possible to infer knowledge about some salient features of such a target and thus identify it, either by classifying it as a particular type of object, or else by describing it in terms of its identifiable features.

Such identification schemes can be broadly classified as imaging and non-imaging, depending on whether or not they generate something akin to a photograph or holograph of the target of interest. The aforementioned ranging application can be categorized as a non-imaging technique, whereas airport body scanners utilize imaging techniques [2].

Imaging techniques (e.g., [4]) necessitate multiple measurements and typically operate at millimetre-wave frequencies (approximately 30 GHz to 300 GHz); e.g., the system described in [78] operates in the band from 27 GHz to 33 GHz. The aforementioned body scanners consist of arrays of mobile antennas, and thus are able to acquire the many measurements needed to generate the scanned body's image [78]. Such a generated image can be of fairly high fidelity [78]. However, imaging of this nature is considered invasive by many [2] and thus typically controlled by various laws and regulations [61].

Non-imaging techniques tend to be somewhat simpler and can operate effectively at much lower frequencies than imaging techniques (e.g., the designs presented here

operate predominantly in the low-GHz band). They require only one transmitter and one receiver, at a minimum, but there is a benefit to having greater diversity than that. An intuitive way to understand how a non-imaging technique can be used to identify a target is through analogy to listening to music. For example, one can identify a guitar as such because plucking of one of its strings causes a distinct twang and the subsequent ringing (vibrating, resonating) of the string has distinctive qualities. In radar, the resonances occur due to currents induced in the target, which form standing waves that depend on the target's shape and material properties, and attenuate over time due to the target's resistivity; that is, the standing waves are longitudinal rather than transverse as in the guitar string. The twang corresponds to the *early-time return* or *early-time response (ETR)* of a target and the subsequent ringing that occurs corresponds to its *late-time return* or *late-time response (LTR)* [82]. The only significant breakdown of this analogy is that it is the radar that stimulates the target from afar, rather than a musician plucking a string directly. Instead, one could potentially imagine a shockwave from a bomb causing the guitar to resonate.

For the purposes of detection and ranging, the ETR is altogether sufficient, and the LTR is either irrelevant or outright detrimental, such as in cases where a strong LTR from one target masks a much weaker return froI haven't seen SAC in quite a while, but im another. However, when the end goal is target identification, the LTR becomes useful. Whereas the ETR is highly dependent on the stimulus, the frequency and attenuation rate of the LTR are dependent on the target itself. For example, a thin conductive rod of length 15 cm will resonate primarily at 1 GHz,[2] and if that rod is made of copper, the resonance will attenuate more slowly than if it were made

[2]Using a thin-wire approximation (e.g., [44]), the wavelength of the fundamental resonance is $\lambda = 30$ cm, and the corresponding frequency is $f = c/\lambda \approx 1$ GHz, where the speed of light is $c \approx 30$ cm/ns. Overtones of this fundamental frequency are present as well.

of iron, which is more resistive [44]. More complicated targets would naturally have more resonant components.

With this background knowledge, one can envision how a non-imaging target identification technique might work. First, by analyzing the ETR, target detection and ranging could be performed. Then, assuming a target has been detected, the resonant frequencies and attenuation rates of the target could be estimated from the LTR. Through the estimation of these *resonance parameters* or *resonance features*, one could infer salient information about the target. If a database of such resonance parameters is available, then the targets could even be identified, just like one can identify a guitar in a musical recording by knowing what it sounds like.

The general idea is thus quite promising and has been advocated for by Carl Baum [12] and others [21], [47]. However, as discussed in more detail in the following section, the practical implementation of such a system is very challenging. Several methods have shown promise in simulation experiments on electromagnetically simple targets in idealized environments at somewhat optimistic SNRs [28], [62]. However, it appears that it has been difficult to realize that promise in physical experiments, even in a well-controlled environment. The goal of this thesis is to take a fresh look at the theory, algorithms and practical implementation of target identification from measurements of the LTR, inspired, in part, by the capabilities of a custom-designed picosecond-pulse radar with high pulse repetition rate that has been developed at McMaster University. This time-domain radar operates between approximately 0.5 GHz and 5 GHz, which corresponds to target features sized between approximately 3 cm

and 30 cm.[3] This choice is practical in terms of experimental feasibility since countless common handheld metallic items are of such a size, and it is relatively easy to manufacture uncommon items of such a size by hand. In terms of application, such a system would be useful in the detection of visually concealed items carried by a person (e.g., contraband, weapons), which are quite likely to be of such a size. Such a design faces a number of theoretical and practical challenges. As outlined in the following sections, the work presented in this thesis addresses many of those challenges at the theoretical, algorithmic and practical implementation levels, and demonstrates that by employing the proposed approaches, promising performance can indeed be achieved in practice.

1.2 Main Challenges

A noise-free LTR signal can be modelled as a superposition of damped sinusoids (cf. [12], [50]). Each such component has an initial amplitude and phase, which can be combined into a *complex amplitude*, and a resonant frequency and attenuation rate, which can be combined into a *complex frequency*. The complex amplitude inherently depends on the initial point of the observation window within the LTR. In contrast, any observation window within the LTR should contain the same complex frequencies. Thus, the complex frequencies are features that are time invariant for the duration of the LTR, making them a natural choice for characterizing the target.

In [50], Hua et al. make use of this fact by sliding such an observation window over the LTR and comparing the signal portions contained therein. In particular,

[3] In principle, overtones of features larger than this can be captured, but they are significantly weaker than the fundamental resonance.

they do so over the discretely sampled LTR and arrange the subsequent observation windows into a pair of Hankel matrices referred to as a *matrix pencil*. Finally, they show that the complex frequencies are directly related to the generalized eigenvalues of this matrix pencil.[4] Provided certain constraints are met, this *matrix-pencil method (MPM)* yields a unique set of estimates of the complex frequencies for a given signal.

However, as Rellich outlines in Chapter 1, Section 1 of [73], entitled "A Small Perturbation Parameter does not mean a Small Perturbation," a small perturbation to the matrix pencil does not necessarily result in a small perturbation to the eigenvalues. Therefore, when noise randomly perturbs the LTR signal, even slightly, it randomly perturbs the estimates of the sought-after resonance parameters, and those perturbations are not necessarily small. This is the fundamental problem that makes this sort of estimation difficult. That difficulty is only exacerbated by the fact that due to its very nature, the LTR attenuates, thus a longer observation window will lead to a lower average signal-to-noise ratio (SNR) over the window regardless of the initial signal amplitude. In practice, the start of LTR is typically some 10 dB weaker than the ETR, so even this initial amplitude is quite low.

Strategies to mitigate the impact of these noise-induced perturbations include filtering, as is implicitly done in a significant improvement to the MPM known as the generalized pencil-of-functions (GPoF) method [49], and gathering more samples, either through longer observation windows or multiple measurement shots. These must be evaluated in the context of what can be measured experimentally. A major issue in such experiments stems from multipath interference. Techniques such as the MPM are rather sensitive to model mismatch, thus the observation window of the LTR needs to be purely a superposition of damped sinusoids (and noise). This

[4]Each complex frequency $\tilde{\omega}_k$ corresponds to two eigenvalues, $e^{+j\tilde{\omega}_k}$ and $e^{-j\tilde{\omega}_k^*}$.

means that the longest possible observation window for the LTR is between the end of the ETR of the initial response and the start of the ETR of the first multi-path reflection. In many applications, including the experiments performed in this thesis, that precludes the use of techniques such as the one in [37], which offers a significant improvement to the MPM but only for very long observation windows. As pointed out in Chapter 2, the filtering in the GPoF reduces the variance of the estimates of the resonance parameters, as indicated in [49], but that does not necessarily improve the average estimation performance. This leaves gathering multiple measurement shots as the remaining alternative.

The key problems with working with many measurement shots is that the shots must be acquired and then jointly processed in a manner that captures both the nature of the acquisition and the nature of the environment. Consider an idealized scenario in which multiple measurement shots can be acquired so rapidly that the target and environment are effectively static, and there are no irregular sampling artifacts in the observation windows of the LTRs. Simple averaging of the measurements shots is a viable option, and doing so improves the effective SNR [55]. However, even in this idealized scenario, such averaging does not guarantee a small perturbation in the estimates of the complex frequencies.

That observation suggests that more sophisticated statistical signal-processing techniques might be needed. Unfortunately, that leads to another problem: Although there are a few standard ways[5] to estimate complex frequencies from LTR-like measurements of signals, there is no known closed-form expression for the probability distribution of such estimates from signals that include additive noise for any of these

[5]These include the MPM and GPoF, Prony's method [86] and finding the poles of Padé approximants of the z-transform [39].

methods. This makes statistical analysis challenging. An empirical observation is that the estimates of the complex frequencies tend to cluster near to the complex frequencies of the corresponding noise-free signal. Methods such as the one in [8] seek to exploit that observation by identifying clusters and using their "centres" as estimates of the complex frequencies. However, these clusters often overlap and differ greatly in spread. This makes the method in [8], which employs of averaging of the estimates in a cluster to find its "centre," awkward to use since it is difficult to determine to which cluster an estimate belongs. In this thesis, this shortcoming is partly overcome by using the estimates to create a smooth empirical distribution of the complex frequencies, and analyzing this distribution instead of the estimates themselves.

1.3 First Major Contribution – Estimating Complex Frequencies from Multiple Measurement Shots

As outlined above, the GPoF [49] is a promising approach to estimating the complex frequencies when only a single measurement shot is available. It is computationally efficient and has the flexibility that it does not require a model for the additive noise. However, the sensitivity of the underlying eigenvalue problem to perturbations limits its reliability, especially in scenarios with low SNR. In light of this, Barone suggested using the GPoF to generate estimates of the complex frequencies on each of many measurement shots [8]. He found that at sufficiently high SNRs these estimates form

distinct clusters and that the clusters' centres are good estimates of the complex frequencies of the LTR. In essence, the computation of the cluster centres both averages the "reliable" individual estimates and excludes outliers.

When the SNR is lower, the clusters are no longer as well defined, and thus Barone's method collapses. In an effort to extend the underlying principles of Barone's approach to the low-SNR regime, a slight modification is proposed. Rather than taking a clustering approach, the estimates from each measurement shot are used to construct an empirical distribution using kernel density estimation (KDE) [83]. Instead of cluster centres, local maxima (peaks) of the KDE serve as estimates of the complex frequencies. At high SNRs, these peaks are indeed very close to Barone's cluster centres, and thus close to the complex frequencies that are being estimated. As the SNR is lowered, Barone's original method collapses, whereas the KDE peak-finding modification continues to yield reasonable estimates. However, those estimates drift as the SNR decreases, and thus further refinement is needed.

A promising candidate for such a refinement technique lies in the work of Umesh and Tufts, who tackle the problem of estimating the complex frequencies from a single measurement shot by deriving an associated nonlinear least-squares problem [84]. In the case of white Gaussian noise, their proposed estimator is the maximum-likelihood estimator. It is shown that the extension of this work to the case of multiple measurement shots is straightforward in an algebraic sense, but that it requires careful consideration of the model for the relationships between the complex amplitudes in different measurement shots.[6] Furthermore, an important issue with this overall approach is that the nonlinear least-squares problem has a very challenging optimization

[6]The complex amplitudes are "nuisance parameters" in our setting in the sense that we wish to estimate the complex frequencies, but in order to do so we need to simultaneously model and (at least implicitly) estimate the complex amplitudes.

landscape. For the approach to be effective on a reasonable time scale with current computing technology, a starting point that is reasonably close to the optimal estimate is needed. One of the insights that underlies each of the contributions in this thesis is that the aforementioned modification to Barone's method yields such a starting point.

Unfortunately, the sampling jitter and synchronization[7] issues that were present in the prototype radar at the time of this work suggested that the nonlinear least-squares estimator might not perform adequately in practice. This is because the implicit assumption that underlies its good performance, namely that the dominant form of signal degradation is additive Gaussian noise, would likely not hold well. As a result, although the nonlinear least-squares estimates constitute a refinement of the estimates from the proposed modification to Barone's method, they themselves might need refinement in practice.

One approach that could be envisioned for refining the least-squares estimates would be to first obtain an analytical expression for the distribution of the estimates of the complex frequencies of an LTR that is parameterized in terms of the true complex frequencies and the true complex amplitudes. Then, one could optimize the parameters so that the analytic distribution "matches" the empirical distribution that was constructed using KDE techniques in the proposed modification of Barone's technique. While this approach has considerable potential, unfortunately there is no known analytical expression for the distribution of the estimates of the complex frequencies, even in the simple case of additive Gaussian noise. However, Hammersley derived an analytical expression for the distribution of the roots of a polynomial whose coefficients have an arbitrary Gaussian distribution in [40]. The z-transform

[7]Jitter of the entire signal, rather than individual samples.

of a finite-length window of an LTR signal in additive correlated Gaussian noise is one example of such a polynomial. This may not immediately seem relevant, but the MPM eigenvalues are also the poles of a particular Padé approximant of the z-transform, and work such as that of Froissart [32], where it is observed that a pole and a zero can often interact as a "doublet," suggests that this root distribution likely relates closely to the eigenvalue/pole distribution. Furthermore, the peaks of the root distribution drift as the SNR is lowered, which aligns with the observations made regarding the peaks of the KDE.

Thus, a method of using the Hammersley distribution as a model was conceived, and a means to compare it to the KDE of the z-transform's roots, rather than the complex frequencies, was devised. In this method, the mean of the distribution of the coefficients of the polynomial is the idealized LTR, and thus it is parametrized in terms of the complex amplitudes and complex frequencies. The covariance of the coefficients is the noise covariance and is assumed to be known. By optimizing over the parameters of the mean, with the objective to minimize the difference between the empirical KDE of the roots and Hammersley's analytical expression for the distribution of the roots, the complex frequencies can be estimated. In the case of an LTR in zero-mean, additive, possibly correlated, Gaussian noise, the covariance of the coefficients of the polynomial in Hammersley's model is simply the covariance of the additive noise. In scenarios with significant jitter or synchronization errors, the distribution of the coefficients of the polynomial is not Gaussian, but it is shown that an effective Gaussian approximation can be constructed.

When combined with some well-chosen constraints, this method based on Hammersley's distribution of the roots proved to be the most effective in numerical experiments with synthetic signals, even in the case of Gaussian noise at very low SNR. It outperformed the proposed modification of Barone's method, and the nonlinear least-squares method derived from the work of Umesh and Tufts, especially in the cases of jitter and synchronization error. In particular, the demonstrated performance was quite reasonable even for levels of jitter that are comparable to the sampling interval, which is unreasonably high inasmuch as this degree of jitter would make for a rather poor sampling receiver.

This work was published in the *IEEE Transactions on Signal Processing* as [34] and appears as Chapter 2 of this thesis.

1.4 Second Major Contribution – Estimating Complex Frequencies in a Cluttered Environment

The approach of matching a KDE of the roots of the z-transform of the observed window on the LTR to a Hammersley root distribution described in Chapter 2 may be an effective method to estimate the complex frequencies of an LTR, but it is also very computationally demanding. In contrast, the extension of Umesh and Tufts' nonlinear least-squares approach to direct estimation of the complex frequencies that was also proposed in Chapter 2 performs adequately in many cases and converges quite quickly when sufficiently well initialized. This leaves the robustness to non-Gaussian uncertainty as the main advantage of the Hammersley-inspired method.

However, after these methods were developed, the sampling issues of the custom-designed prototype radar were resolved, and the jitter in the re-designed receiver was measured to be low enough to be neglected. Thus, a pragmatic decision was made to base the further development of LTR analysis techniques on the extensions of the nonlinear least-squares approach, initialized, as before, by a KDE-based modification of Barone's method.

An arguably more critical problem in LTR analysis is that of environment clutter, which is unavoidable in practice as even the radar system itself contributes to such clutter. In general, clutter can cause multi-path interference, but in the envisioned application, multi-path interference simply curtails the available LTR window. (The observation window is truncated so that it closes before the first multi-path reflection arrives.) The important role that clutter plays in the envisioned application is that other objects in the environment and the radar itself resonate, and it is necessary to distinguish these resonances from those of the target.

To that end, a method is proposed in which the complex frequencies of the resonances of the environment are estimated first. These resonances are then assumed to also be present once a target is introduced, with the same frequencies but with different complex amplitudes than prior to the target being introduced.[8] As part of this method, the extension of the nonlinear least-squares problem and its Barone-based initialization are adjusted to account for these now-fixed complex frequencies.

With the problem of clutter potentially resolved, this method becomes a viable option for analyzing LTRs obtained from physical experiments with the prototype radar. The results of those experiments show that the proposed method is reasonably

[8]For example, consider that the target may partly block a resonating feature of the environment, thereby reducing its amplitude.

capable at estimating target resonances in the presence of clutter, even when that clutter heavily influences the received signal.

1.5 Third Major Contribution – Classifying Targets Using Distributions of Complex Frequencies

At this point, two promising methods for estimating resonance parameters have been demonstrated, but the underlying problem of eigenvalue perturbation remains: one can only have so much confidence that a particular estimate is accurate. However, if the ultimate goal is to classify unknown targets, specific estimates of the resonance parameters are not necessarily needed. If empirical distributions of the estimates of the complex frequencies of a particular target are stored in a database, and then the distribution for a new, unknown target is measured, it follows that this new distribution can be compared to each of those in the database. The classification problem thus becomes a multiple-hypothesis test; i.e., it becomes the problem of determining which database target this new target most likely to be.

Were only one or a handful of measurement shots available, one might conduct a classical multiple-hypothesis test of each estimate of a resonance frequency that is available, then combine these into a compound result. When a significant number of measurement shots are available, it becomes possible to construct an empirical distribution of the estimates of the complex frequencies. Furthermore, the notion of hypothesis testing naturally extends to measuring the Kullback-Leibler divergence (relative entropy) between the distribution associated with the unknown target and

each distribution in the database [58]. By combining these observations, an approach to the classification of unknown targets using LTR measurements is developed. That approach falls into a category of classification schemes that involve distribution matching.

Some care is needed in order to conduct the distribution matching effectively in this setting. First, due to its construction, a KDE is a good approximation to a distribution in subdomains where that distribution is quite dense (in the sense that its probability density function has a high value), and can be relatively poor in subdomains where it is sparse because there are only a few outlier samples on which to base the estimate. Thus, if two KDEs are to be matched, it should be done in these dense subdomains. In the proposed application, such refinement can be guided by practical insights. In particular, only the subdomain that corresponds to the frequency band in which the radar operates is considered. Similarly, analysis of the physics of the system yields sensible bounds that can be placed on the attenuation rates. Second, in a KDE constructed from measurement shots of a target embedded in clutter, the clutter, which includes the radar system itself, has a substantial impact. When a KDE constructed from this clutter, or even just some of it, is available, it is sensible to use it to identify subdomains where the "target-in-clutter" measurement is most similar to the clutter measurement. These subdomains can then be excluded from the matching against database targets. This is especially useful in scenarios in which the clutter that is present when the unknown target is measured is different from the cutter that was present when the database of known targets was constructed, but it remains helpful even when the clutter environments are very similar.

The proposed approach has many advantages, but a key one is that it does not

rely on very many ad-hoc parameters, and for those that are present there are sensible ways to choose them. An upper bound on the number of complex frequencies must be set, but it can be quite a loose upper bound.[9] The other parameter has to do with how smooth the empirical distribution should be, and this choice can be largely guided by the number of measurement shots available. With infinite measurement shots, no added smoothing is necessary at all, but with very few shots, the divergence between empirical distributions can be quite large simply because regions with some samples in one distribution are compared to regions with no samples in the other, and added smoothing helps reduce this effect.

All of this is unlike methods for estimating the complex frequencies themselves, in which at least the number of complex frequencies needs to be specified, even though it is generally unknown. Beyond that, the optimization algorithms used in the estimation process have to be chosen carefully, and their parameters, including starting points for iterative algorithms have to be selected. These difficulties are avoided in the proposed distribution-based approach, and perhaps that has much to do with the ultimate efficacy of the technique in physical experiments involving the identification of handheld weapons. In particular, targets are demonstrated to be classified correctly, even when partly obscured by resonating clutter, and the method was still largely effective when the targets were in orientations different to those that were used when the database was constructed.

The proposed approach to classification of an unknown target is likely to be successful if the distribution of the unknown target diverges only a small amount from

[9] In fact, the choice made has to do more with the GPoF, for which there is a rule of thumb that, for a signal of length N, one should estimate at most about $N/4$ complex frequencies for good results, even though the theoretical maximum is $N/2$.

one of the elements of the database and diverges from the other elements by a significantly larger amount. In a scenario in which the distribution of the unknown target diverges significantly from all elements in the database, it becomes probable that the unknown target is not one of the items from which the database was constructed. In that case, it makes sense to fall back onto the methods of the second contribution, which characterize the target rather than classify it. That is to say, the proposed classification scheme complements rather than supplants the previously developed method.

Chapter 2

Estimating Resonances in Low-SNR Late-Time Radar Returns with Sampling Jitter

This work is published in the *IEEE Transactions on Signal Processing* as [34]. Its reproduction here is in accordance with the IEEE's Copyright Policy.

Abstract

The frequency and attenuation rate of a resonance in the late-time return of a radar signal are indicative of a target's geometry and conductivity, and hence they can be used as features in a variety of filtering and classification applications. However, late-time returns are typically observed over short windows at low signal-to-noise ratios (SNRs, averaged over the window), and often in the presence of sampling jitter. This can make the estimation of these parameters difficult, even when multiple measurement shots are available. In this chapter, we develop a new multi-shot estimation

method that is based on models for the distribution of the roots of the z-transform of the received signal. Under an additive-Gaussian-noise model, we have a closed-form expression for the root distribution in terms of the resonance parameters, and the parameters are estimated by matching the model distribution to the empirical distribution. The root distribution has a strong dependence on the frequency and attenuation rate, and leads to significantly better estimates than existing techniques at low SNRs. By developing approximate models, we extend these performance advantages to scenarios with significant sampling jitter and synchronization offsets.

2.1 Introduction

A common non-imaging technique for classifying radar targets involves the identification of resonant modes, which appear as damped sinusoids in the radar return [11], [12]. In impulse radar, the frequency of such a mode is indicative of the size and shape of an object, whereas the attenuation (damping) rate is indicative of the overall loss of the object, which is dependent on its conductivity and its geometry [21], [80]. For example, a 15-cm thin metallic rod has a dominant resonant mode with a 30-cm wavelength and thus a 1-GHz frequency. The more resistive the rod, the faster the resonance attenuates. More complex objects have multiple detectable resonances, whose frequencies and attenuation rates can be treated as classifiable features.

Such resonance classification techniques have numerous applications, including the detection of targets at close-range, such as landmines, which prompted the development of the original singularity expansion method (SEM) that this work aims to improve upon [12], or on-body concealed weapons [3]. In on-body applications, these techniques have certain advantages over imaging methods, such as millimeter-wave

scanning [60], as they can operate in a monostatic regime with a single stationary antenna. Polarization-diversity systems may employ as many as four antennas, but they remain stationary, whereas an imaging radar requires some combination of multiple antennas and mechanical movement to form a synthetic aperture. Furthermore, non-imaging methods do not produce an image of the target, which may be a person or the surrounding environment. As such, they are perceived as less invasive [5] and more suitable in applications where privacy is a concern.

For target resonances to be measured, an ultra-wideband pulse is transmitted, and the signal reflected from the target is received and sampled in the time domain [11]. The spectral span of such pulses is designed to include all expected resonant frequencies. The received signal includes an *early-time response (ETR)*, which has a complex structure [46], but is dominated by the direct reflection of the pulse from the target. If resonance is excited within the target by the pulse, the ETR is followed by ringing, which contains the resonant modes. This ringing is termed the *late-time response (LTR)* [42]. An explanation of how such resonances evolve in a complex target is given in [45], and the complexities of this evolution are demonstrated via simulation in [57]. Thus, the identification of the onset of the LTR requires some sophisticated analysis [42], [76], but once it has been identified, the goal of the receiver is to estimate the resonant frequencies and attenuation rates. As we describe in more detail below, the estimation task faces three major challenges: (i) the usable LTR duration is typically short, often on the order of 10s of samples, (ii) the average SNR over the usable duration of the LTR is typically low, often below 10 dB, and (iii) it is difficult to construct a sampler with sufficiently low jitter at the requisite sampling rates.

There are two factors which limit the usable LTR duration. Firstly, due to the lossy nature of most radar targets, the LTR is a rapidly attenuating signal. Secondly, it is necessary to reject multipath interference and clutter. If there were only one reflecting object in the environment, there would only be one ETR followed by one LTR. In the presence of clutter and multipath propagation, the observation window of the LTR must end before the next response arrives. That response might arise from a clutter reflection or may be a response from the desired target arriving along a reflected path. For example, if a monostatic radar and its target are 1 m above a reflective floor and 2 m apart, then the direct reflection has to travel 2 m to the radar and multi-path reflection off the floor has to travel $2\sqrt{2}$ m to the radar. This is 83 cm longer, a distance that light travels in 2.76 ns. Furthermore, the LTR window can only start once the ETR has decayed sufficiently. The ETR has an appreciable duration, due to both the non-ideal pulse shape and the depth of the target object. For example, if the pulse width is 0.1 ns and the target is 5 cm deep, then the last part of the ETR arrives $(0.1 + 0.33)$ ns after the first. Taking the reflection, pulse width and target depth into account, a realistic LTR window for this example is around 2.4 ns. At a sampling frequency of 20 GHz, a 2.4-ns window contains 48 samples.

With regard to the SNR, we observe that the start of the LTR is typically more than 10 dB weaker than the ETR [42]. By its very nature, the LTR only decays from that point in time. Ambient noise and active interference limit the effective sensitivity of the receiver. Furthermore, even if the radar and target are isolated from the ambient environment in order to eliminate the regulatory concerns, the power of the transmitter is limited. Factoring all this in, the average SNR over the window is typically 10 dB or less. This average SNR is the ratio of the signal energy

over the usable LTR duration to the noise energy over that duration, and is formally defined in Section 2.2.1. As such, the average SNR depends on the attenuation rate of each resonant component. We will assume relatively low attenuations, as is typical in the case of metallic radar targets; see [87] for a method suited to high-attenuation scenarios.

With regard to sampling jitter, real-time sampling oscilloscopes that operate at 20 gigasamples/s or faster with low jitter exist, but they are very costly. A more affordable alternative is equivalent sampling, which operates at a lower physical sampling frequency and repeats the same measurement multiple times with different offsets. For example, sampling at 200 megasamples/s one hundred times at offsets of 0 ps, 50 ps, ..., 4950 ps creates a system with an equivalent sampling period of 50 ps (a rate of 20 gigasamples/s). A problem arises in that low jitter compared to the physical sampling period (on this example, 5 ns) may be quite high compared to the equivalent period (in this example, 50 ps).

The goal of this chapter is to develop an approach to estimating the frequencies, $\breve{\omega}_k$, and attenuation rates, $\breve{\alpha}_k$, of the resonances in an LTR that can provide reliable performance in the difficult conditions described above. For ease of exposition, we will call $\breve{s}_k = -\breve{\alpha}_k + j\breve{\omega}_k$ the *complex frequency* of the resonance. As we outline in Section 2.2, the established approaches for estimating the complex frequencies include:

- the nonlinear least-squares technique in [84], which is the maximum likelihood (ML) estimator in additive white Gaussian noise [20];
- the matrix-pencil method (MPM) [50] and generalized pencil-of-functions (GPoF) method [49], which are inspired by the Kumaresan-Tufts method [51], [56], [72] and are based on observations regarding the structure of the LTRs

that enable estimation of the complex frequencies through the solution of a generalized eigenvalue problem; and

- Padé approximation methods [7], [14], [25], [39], which are based on approximating the z-transform of the finite-length received signal by a rational function whose poles are the estimates of the complex frequencies.

These existing *single-shot methods* perform well when applied to sufficiently long observation windows with sufficiently high SNRs using samples with sufficiently low jitter. However, even in the absence of jitter, they can be quite sensitive to certain noise realizations. For example, in the GPoF method, the noise realization perturbs both the matrices in the matrix pencil, and it is well known that small perturbations in those may result in significant perturbations in the generalized eigenvalues [73]. In fact, these single-shot methods are so susceptible to noise [9] that it is difficult to make use of them in the applications we have envisioned, such as on-body target detection.

In these applications, one strategy to mitigate the sensitivity to noise is to take multiple measurements of the LTR. (In equivalent sampling systems, this corresponds to multiple measurements at the high sampling rate.) In the envisioned applications, this is facilitated by the fact the range of the system is small enough that measurements can be repeated on a time scale of hundreds of microseconds. As we explain in Section 2.4.1, if the measurements are perfectly synchronized and free from jitter, and if the additive noise has zero mean, then simple, if somewhat naive, estimates of the complex frequencies can be constructed by applying the established single-shot techniques to the average of the received signals. However, these naive multi-shot estimators retain the inherent sensitivities of the underlying single-shot estimators.

Furthermore, they are inherently sensitive to synchronization offsets and jitter.

A more sophisticated strategy would be to construct a model for the timing errors (i.e., the synchronization errors and the jitter) and derive the ML estimator for the complex frequencies. Unfortunately, when the timing errors are modeled as non-random parameters, the number of "nuisance" parameters in the ML estimator grows linearly with the number of shots, and when the timing errors are modeled as random parameters with a known distribution, the expression for the likelihood, which the estimator must maximize, involves multidimensional integrals and is computationally unwieldy.

An interesting alternative strategy, inspired by the work of Barone [8], is to perform independent single-shot estimation on each measurement. The estimated complex frequencies from each of the shots are scattered randomly, but form clusters. An estimate of each complex frequency can be obtained by taking some measure of the center of the corresponding cluster, such as its mean or the local maximum of an estimated density function. The underlying idea, which leads to some robustness in the technique, is that the highly perturbed estimates will present as outliers that are far enough away from the clusters that they can be disregarded. This is indeed the case, but, unfortunately, at low SNRs, some of these cluster-center estimates can degrade significantly as shown in Figure 2.1.

Barone's technique [8] can be interpreted as a model-free method that uses multiple measurements to implicitly construct an empirical distribution for the complex frequencies and then seeks to identify local maxima in that distribution (see Figure 2.1). That interpretation suggests that if we had access to a model for the distribution of the complex frequencies under given noise and timing error models, we could develop

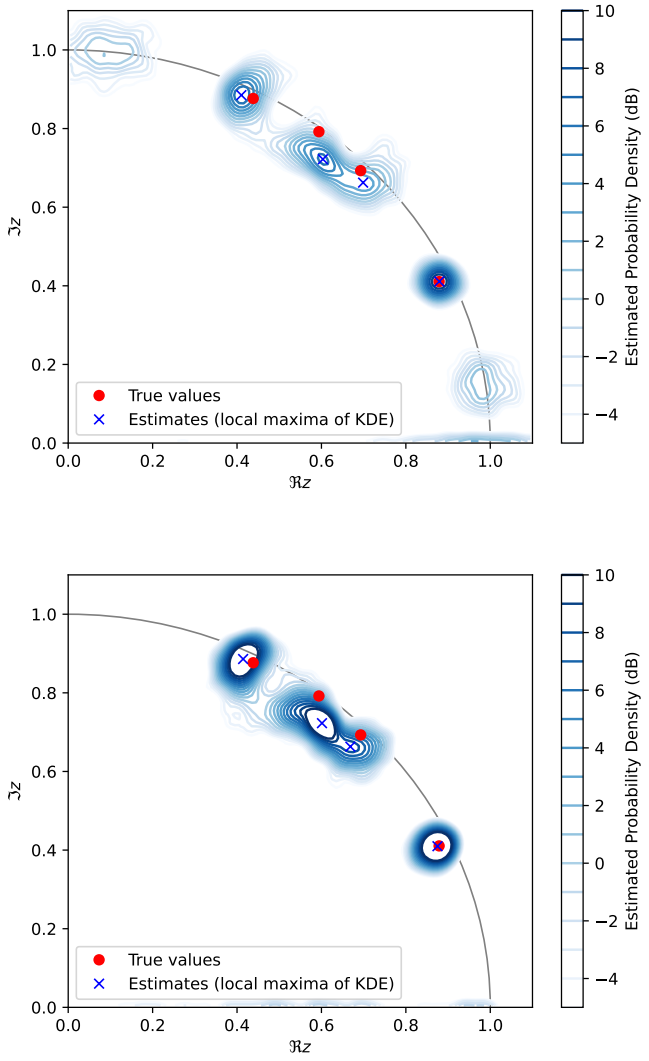

Figure 2.1: Contours (blue lines) of a kernel density estimate (KDE) [83] of the distribution of the poles $\left\{e^{\check{s}_k T}\right\}$ of a 4-resonance signal at an SNR of 10 dB sampled at a rate $1/T$. The KDE was generated from 1600 48-sample measurement shots with the matrix-pencil method (MPM) [56] (top), or the generalized pencil-of-functions method (GPoF) [49] (bottom), being used to extract the complex-frequency estimates of each measurement shot. Using the principles of Barone's multi-shot technique in [8], we have plotted (with blue crosses) the local maxima of the KDE. In the example, note that while the GPoF's filtering leads to a cleaner empirical distribution, it does not improve the estimates. In particular, the distances to the true values in the MPM case are 0.001, 0.031, 0.071, 0.029, and in the GPoF case, they are 0.004, 0.039, 0.070, 0.026.

a model-based variant of Barone's method. Such a method would optimize the values of the estimates of the complex frequencies so that the model-based distribution comes as close as possible to the empirical distribution in an appropriate sense. If it were possible to construct such an estimator, it could be viewed as a method-of-moments estimator that simultaneously considers all the moments. Unfortunately, the complex frequencies are a complicated function of the signal measurements, and hence a model for their distribution has yet to be derived, even in the case of white Gaussian noise. Thus, the envisioned estimator cannot be constructed.

In an earlier work [10], Barone and Ramponi partly overcome this by deriving an approximate model of the complex-frequency distribution using a result from Hammersley regarding the distribution of roots of polynomials with Gaussian coefficients [40]. Since the complex frequencies can be derived from the roots of the denominator of the Padé approximant of the z-transform of the finite-length signal (see Section 2.3.3), its coefficients are approximated with Gaussian ones so as to use Hammersley's result. Unfortunately, even in the case of an input signal with additive Gaussian noise, which is of predominant interest to us, the coefficients of the denominator of the Padé approximant may deviate significantly from a Gaussian distribution, so this method necessarily incurs errors due to model mismatch.

The principle that underlies the approach proposed in this chapter is analogous to that for the model-based variant of Barone's method that was envisioned above. However, instead of attempting to work with an empirical distribution of the complex frequencies, we construct an empirical distribution of the roots of the z-transforms of the finite-length measurements. Since the received signal is of finite length, its z-transform is a polynomial. In the case of additive Gaussian noise, the coefficients

of that polynomial are Gaussian, and we obtain a model for the distribution of the roots that is an explicit function of the complex frequencies and amplitudes of the resonant modes. This is also based on Hammersley's result in [40]. However, in the Gaussian case, there is no model mismatch. Since we work directly with the received signal rather than the Padé approximant, the complex frequencies are estimated by seeking to minimize the distance between the empirical and modeled distributions.

We first develop this technique for the case of no timing errors, and then we show how it can be extended to accommodate timing jitter and synchronization offsets. In our numerical results, we show that the proposed estimator provides good-quality estimates at low SNRs.

The rest of the chapter is arranged as follows: In Section 2.2, we establish models for single- and multi-shot measurements of LTRs. In Section 2.3, we review some established methods of single-shot estimation of the complex frequencies, and in Section 2.4, we review some established methods for the multi-shot setting. The final preparatory step for the development of the proposed method is a review of Hammersley's closed-form expression for the distribution of the roots of a polynomial with Gaussian coefficients [40], which we provide in Section 2.5. The proposed method is introduced in Section 2.6, and the approximations that enable extension to scenarios with significant timing jitter or synchronization offsets are introduced in Section 2.7. The simulation results in Section 2.8 demonstrate the performance advantages of the proposed approach at low SNRs, and conclusions are drawn in Section 2.9.

2.2 System Model

In this section, we describe our system models for the single-shot and multi-shot cases. We consider a signal model that consists of K complex-exponential components in additive noise. The time origin is set at the time that the LTR arrives at the receiver. Thus, the signal at the receiver input can be written as

$$y(t) = \sum_{k=1}^{K} \breve{r}_k e^{\breve{s}_k t} + \breve{v}(t), \tag{2.1}$$

where $\breve{r}_k, \breve{s}_k \in \mathbb{C}$ and $\breve{v}(t)$ represents the additive noise. In the case of real-valued measurements, K is even, $\breve{r}_{k+K/2} = \breve{r}_k^*$ and $\breve{s}_{k+K/2} = \breve{s}_k^*$. To simplify our exposition, in this section we will focus on the complex-valued case.

2.2.1 Single-Shot Model

In this model, the receiver obtains N samples of the incoming continuous time signal $y(t)$, starting at some time t_0. In the absence of sampling jitter, the period T between samples is constant and the samples can be written as

$$y[n] = y(t_0 + nT),\ n = 0, \ldots, N-1. \tag{2.2}$$

Therefore, we can write

$$y[n] = x[n] + [n] = \sum_{k=1}^{K} r_k e^{s_k n} + v[n], \tag{2.3}$$

where $x[n]$ denotes the signal component of the sampled received signal, the normalized complex frequencies are $s_k = \breve{s}_k T$, the residues are $r_k = \breve{r}_k e^{\breve{s}_k t_0}$, and $v[n]$

represents the filtered and sampled additive noise. At times, it will be convenient to refer to e^{s_k} as a *pole*; see Section 2.3.3. Using ideas from [84], we can rewrite the model (2.3) as

$$\mathbf{y} = \mathbf{F}(\mathbf{s})\,\mathbf{r} + \mathbf{v}, \tag{2.4}$$

where $\mathbf{y} \in \mathbb{C}^N$ with $[\mathbf{y}]_n = y[n]$, $\mathbf{s} \in \mathbb{C}^K$ with $[\mathbf{s}]_k = s_k$, $\mathbf{r} \in \mathbb{C}^K$ with $[\mathbf{r}]_k = r_k$, $\mathbf{v} \in \mathbb{C}^N$ with $[\mathbf{v}]_n = v[n]$, and

$$\mathbf{F}(\mathbf{s}) = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ e^{s_1} & e^{s_2} & \cdots & e^{s_K} \\ e^{s_1 2} & e^{s_2 2} & \cdots & e^{s_K 2} \\ \vdots & \vdots & \ddots & \vdots \\ e^{s_1(N-1)} & e^{s_2(N-1)} & \cdots & e^{s_K(N-1)} \end{bmatrix} \in \mathbb{C}^{N\times K}. \tag{2.5}$$

For scenarios in which the additive noise has zero mean, we define the average SNR over the observation window to be

$$\frac{\sum_{n=0}^{N-1} |x[n]|^2}{\sum_{n=0}^{N-1} \mathrm{E}\left(|v[n]|^2\right)} = \frac{\mathbf{r}^{\mathrm{H}}\mathbf{F}(\mathbf{s})^{\mathrm{H}}\mathbf{F}(\mathbf{s})\,\mathbf{r}}{\mathrm{tr}\left(\mathrm{E}\left(\mathbf{v}\mathbf{v}^{\mathrm{H}}\right)\right)}$$

where $\mathrm{E}(\cdot)$ denotes statistical expectation. In [84], this quantity is called the essential SNR.

2.2.2 Multi-Shot Model

If we have M measurements of the form in (2.3), over which the (normalized) complex frequencies s_k do not change, then the m-th of these measurements can be written as

$$y_m[n] = \sum_{k=1}^{K} r_{k,m} e^{s_k n} + v_m[n], \tag{2.6}$$

and also as

$$\mathbf{y}_m = \mathbf{F}(\mathbf{s})\,\mathbf{r}_m + \mathbf{v}_m. \tag{2.7}$$

The concatenation of these measurements can be written as

$$\tilde{\mathbf{y}} = \big(\mathbf{I}_M \otimes \mathbf{F}(\mathbf{s})\big)\tilde{\mathbf{r}} + \tilde{\mathbf{v}}, \tag{2.8}$$

where $\mathbf{I}_M$ is the identity matrix of size M, $\otimes$ denotes the Kronecker product and the tildes denote concatenation, e.g., $\tilde{\mathbf{y}} = \begin{bmatrix}\mathbf{y}_1^{\mathrm{T}} & \mathbf{y}_2^{\mathrm{T}} & \cdots & \mathbf{y}_M^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}}$. In a scenario in which $r_{k,m} = r_k$ for all m, we can write

$$\tilde{\mathbf{y}} = \big(\mathbf{1}_M \otimes \mathbf{F}(\mathbf{s})\big)\mathbf{r} + \tilde{\mathbf{v}}, \tag{2.9}$$

where $\mathbf{1}_M$ is the M-dimensional column vector whose elements are all one.

2.3 Single-Shot Estimation: Some Established Methods

In this section, we examine estimation methods that operate on one measurement of the form in (2.4) at a time. Our goal is to estimate the (normalized) complex frequencies $\mathbf{s}$. In this sense, the residues $\mathbf{r}$ are nuisance parameters in which we are not interested, but which must nonetheless be incorporated in the estimator. To expedite the exposition, we will consider the case in which K, the number of exponential components in the received signal, is known. In practice, there are many techniques that can be used to estimate K; see, e.g., [42], [54].

2.3.1 Maximum Likelihood for Additive Gaussian Noise

Let us consider a scenario in which the (filtered and sampled) noise vector $\mathbf{v}$ has a circular complex Gaussian distribution with zero mean and covariance matrix $\mathbf{C}$; i.e., $\mathbf{v} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C})$. If we treat $\{r_k\}$ and $\{s_k\}$ as non-random parameters, the conditional distribution of $\mathbf{y}$ in (2.4) is

$$f(\mathbf{y}|\mathbf{s}, \mathbf{r}) \sim \mathcal{CN}(\mathbf{F}(\mathbf{s})\mathbf{r}, \mathbf{C}).$$

Hence, by taking the log of $f(\cdot)$, we can show that the ML problem of finding $\arg\max_{\mathbf{s},\mathbf{r}} f(\mathbf{y}|\mathbf{s}, \mathbf{r})$ is equivalent to (e.g., [20], [56])

$$\arg\min_{\mathbf{s},\mathbf{r}} \|\mathbf{y} - \mathbf{F}(\mathbf{s})\mathbf{r}\|^2_{\mathbf{C}^{-1}}, \tag{2.10}$$

where $\|\mathbf{x}\|^2_{\mathbf{C}^{-1}} = \mathbf{x}^{\mathrm{H}}\mathbf{C}^{-1}\mathbf{x}$ is the Mahalanobis norm [59].

As outlined in [84] and elsewhere, given an estimate for $\mathbf{s}$, denoted $\hat{\mathbf{s}}$, finding the corresponding estimate of $\mathbf{r}$ in (2.10) becomes a standard linear weighted least-squares problem, which has the closed-form solution,

$$\hat{\mathbf{r}}_{\mathrm{ML}}\left(\hat{\mathbf{s}}\right) = \mathbf{A}\left(\hat{\mathbf{s}}\right)^{-1}\mathbf{B}\left(\hat{\mathbf{s}}\right)\mathbf{y}, \tag{2.11}$$

where $\mathbf{A}\left(\mathbf{s}\right) = \mathbf{F}\left(\mathbf{s}\right)^{\mathrm{H}}\mathbf{C}^{-1}\mathbf{F}\left(\mathbf{s}\right)$ and $\mathbf{B}\left(\mathbf{s}\right) = \mathbf{F}\left(\mathbf{s}\right)^{\mathrm{H}}\mathbf{C}^{-1}$. By substituting (2.11) into (2.10), we can find the ML estimates of $\mathbf{s}$ directly, without the need to estimate $\mathbf{r}$, by solving a nonlinear least-squares problem. That problem is equivalent to the following "concentrated" ML problem

$$\hat{\mathbf{s}}_{\mathrm{ML}} = \arg\max_{\mathbf{s}} \mathbf{y}^{\mathrm{H}}\mathbf{B}\left(\mathbf{s}\right)^{\mathrm{H}}\mathbf{A}\left(\mathbf{s}\right)^{-1}\mathbf{B}\left(\mathbf{s}\right)\mathbf{y}. \tag{2.12}$$

Due to its complicated optimization landscape, this can be quite a difficult problem to solve [84].

2.3.2 Pencil-of-Functions Methods

The matrix-pencil method (MPM) [50] and generalized pencil-of-functions (GPoF) method [49] are based on insights into the structure of the noise-free sampled signal in (2.3). In particular, for a window of length W such that $K \leq W \leq N-K$, we can construct the Hankel matrices, $\mathbf{X}_0, \mathbf{X}_1 \in \mathbb{C}^{(N-W)\times W}$ such that

$$\left[\mathbf{X}_p\right]_{qr} = x\left[p+q+r\right]. \tag{2.13}$$

The key insight that underlies this method is that the set of poles $\{e^{s_k}\}$ is equal to the set of K *non-zero* generalized eigenvalues of the *matrix pencil* $(\mathbf{X}_0, \mathbf{X}_1)$; i.e., the solutions $\{\lambda_k\}$ to

$$\det\left(\lambda \mathbf{X}_0 - \mathbf{X}_1\right) = 0. \tag{2.14}$$

In the absence of noise, the remaining $W - K$ generalized eigenvalues are zero. One way to compute the generalized eigenvalues of $(\mathbf{X}_0, \mathbf{X}_1)$ is to compute the conventional eigenvalues of $\mathbf{X}_0^{\dagger}\mathbf{X}_1$, where $(\cdot)^{\dagger}$ denotes the Moore-Penrose pseudo-inverse.

In practical settings, the MPM and GPoF techniques are applied to the noisy measurements $y[n] = x[n] + v[n]$ with the corresponding pencil $(\mathbf{Y}_0, \mathbf{Y}_1) = (\mathbf{X}_0 + \mathbf{V}_0, \mathbf{X}_1 + \mathbf{V}_1)$, under the assumption that the noise is sufficiently small. The GPoF obtains some robustness to noise by modifying the pseudo-inverse of $\mathbf{Y}_0$ so that it only includes the K largest singular values of $\mathbf{Y}_0$, with the remainder assumed to be the result of additive noise alone [49].

The method in [37] offers an alternative approach to compensate for the noise that addresses the noise in $\mathbf{Y}_1$ as well as in $\mathbf{Y}_0$. The method is based on the observation that in the limit of long signal measurements, $\mathbf{Y}_p^{\mathrm{H}}\mathbf{Y}_q \to \mathbf{X}_p^{\mathrm{H}}\mathbf{X}_q + \mathbf{V}_p^{\mathrm{H}}\mathbf{V}_q$. Furthermore, using ergodicity arguments, $\mathbf{V}_p^{\mathrm{H}}\mathbf{V}_q \to \mathbf{C}_{pq}$, where the elements of $\mathbf{C}_{pq}$ can be determined from the covariances $\mathrm{E}\left(v[m]^* v[n]\right)$. This observation suggests that, when the noise covariance is known and the signal observation window is long, improved estimates in the presence of noise can be obtained from the eigenvalues of $\left(\mathbf{Y}_0^{\mathrm{H}}\mathbf{Y}_0 - \mathbf{C}_{00}\right)^{-1}\left(\mathbf{Y}_0^{\mathrm{H}}\mathbf{Y}_1 - \mathbf{C}_{01}\right)$.

While noise compensation techniques do offer performance improvements, the analysis in [37] emphasizes that when the MPM and its generalizations are applied to a noisy signal, they are finding the generalized eigenvalues of the perturbed matrix

pencil $(\mathbf{X}_0 + \mathbf{V}_0, \mathbf{X}_1 + \mathbf{V}_1)$. Although the perturbations on the eigenvalues do scale with the "size" of the noise, it is well known that, in general, the eigenvalues of a perturbed pencil can be very sensitive to the perturbation [73]. That means that in some scenarios, even a small perturbation $(\mathbf{V}_0, \mathbf{V}_1)$ of the noise-free pencil $(\mathbf{X}_0, \mathbf{X}_1)$ can have a significant impact on the estimates of the eigenvalues $\{e^{s_k}\}$. As a result, matrix-pencil methods can be rather sensitive to noise, particularly for short block lengths and at low SNRs.

2.3.3 Padé Methods

The complex frequencies can also be estimated using the principles of Padé approximation [7]. This involves approximating the z-transform of the time-reversed sampled (noisy) received signal

$$Y_N(z) = \sum_{n=0}^{N-1} y[n] z^n, \tag{2.15}$$

which is a polynomial of order $N-1$, by the *Padé approximant*, which is a rational function; i.e.,

$$Y_N(z) \approx \frac{P_{N-K}(z)}{Q_{K+1}(z)},$$

where $P_{N-K}(z)$ and $Q_{K+1}(z)$ are polynomials of order $N-K-1$ and K, respectively. The estimates of $\{e^{s_k}\}$ are the K roots of $Q_{K+1}(z)$, i.e., the poles of the Padé approximant. To find these poles, we first need the coefficients of $Q_{K+1}(z)$.

In the absence of noise, i.e., when $y[n] = x[n]$, see (2.3), we can write

$$P_{N-K}(z) = X_N(z) Q_{K+1}(z),$$

where $X_N(z)$ is the z-transform of the time-reversed version of $x[n]$ in (2.3). This

enables us to show that the vector of coefficients $\mathbf{q}$ of $Q_{K+1}(z)$ lies in the nullspace of the system matrix $\mathbf{X} \in \mathbb{C}^{(K-1)\times K}$ with

$$[\mathbf{X}]_{ij} = x[N-K-1+i-j], \tag{2.16}$$

i.e., we are looking for a nontrivial solution to $\mathbf{Xq} = \mathbf{0}$. Such a solution can be thought of [39] as a generalized eigenvector of the matrix pencil $(\mathbf{W}, \mathbf{X})$ that corresponds to a generalized eigenvalue of 0 for a class of matrices $\mathbf{W} \in \mathbb{C}^{(K-1)\times K}$. The choice $\mathbf{W}^{\dagger} = \mathbf{X}^{\mathrm{H}}$ leads to a numerically robust algorithm [39], which corresponds to finding the eigenvector of $\mathbf{X}^{\mathrm{H}}\mathbf{X}$ corresponding to an eigenvalue of 0.

Unfortunately, like eigenvalues, eigenvectors are sensitive to small perturbations [73]. Thus, in the presence of noise, when $x[n]$ is replaced by $y[n] = x[n] + v[n]$, as in (2.3), the estimate of $\mathbf{q}$ can be quite sensitive of the realization of $v[n]$. There exist methods that attempt to filter the noise by truncating the singular-value decomposition of $\mathbf{Y}^{\mathrm{H}}\mathbf{Y}$, where $\mathbf{Y}$ is defined by analogy to (2.16) but with the possibility of having more than $(K-1)$ rows. More generally, we can solve variations of

$$\min_{\mathbf{q}} \|\mathbf{Yq}\|_{\mathbf{M}}^2 = \min_{\mathbf{q}} \mathbf{q}^{\mathrm{H}}\mathbf{Y}^{\mathrm{H}}\mathbf{MYq}$$

for different "heights" of $\mathbf{Y}$ and different choices of the Hermitian matrix $\mathbf{M}$ [19]. Ultimately, these techniques retain the noise sensitivity of matrix-pencil techniques, with the additional computational cost of finding the roots of $Q_{K+1}(z)$. Thus, the GPoF method is typically preferred for the envisioned applications.

2.4 Multi-Shot Estimation: Some Established Methods

In this section, we consider methods that operate on a set of multiple measurements of the same target, all which have the same complex frequencies $\{s_k\}$. We show that while it is difficult to develop a computationally viable maximum likelihood estimator, there is a method based on the empirical distribution of the complex frequencies $\{s_k\}$ that is viable. In some applications, high-speed acquisition methods are available, naturally providing multiple physical measurements. For other applications, several strategies have been suggested for generating a large set virtual shots from a small set of physical measurements. We summarize two such alternative strategies and their associated trade-offs in Appendix 2.C.

2.4.1 Maximum Likelihood

In the multi-shot model in (2.6)-(2.8), if we model the additive noise as $\mathbf{v}_m \sim \mathcal{CN}(\mathbf{0}, \mathbf{C})$ and as being independent between shots, there are several approaches that we could take to develop an ML estimator for the complex frequencies $\mathbf{s}$. These approaches are based on different underlying models.

First, we could model the residues at each shot, $\{r_{k,m}\}$ in (2.6), as being non-random parameters. In that case, the ML estimator can be "concentrated" in a similar way to the single-shot case, resulting in

$$\hat{\mathbf{s}}_{\mathrm{ML}} = \arg\max_{\mathbf{s}} \sum_{m=1}^{M} \mathbf{y}_m^{\mathrm{H}} \mathbf{B}(\mathbf{s})^{\mathrm{H}} \mathbf{A}(\mathbf{s})^{-1} \mathbf{B}(\mathbf{s}) \mathbf{y}_m, \tag{2.17}$$

where $\mathbf{A}(\mathbf{s})$ and $\mathbf{B}(\mathbf{s})$ are defined as in (2.11). However, this problem has an even more complicated optimization landscape than the problem for in the single-shot case in (2.12). Furthermore, modeling the residues at each shot as arbitrary parameters means that the number of nuisance parameters grows linearly in the number of shots.

A naive way to address the growth in the nuisance parameters would would be to assume that there is sufficient synchronization between the shots and sufficiently low jitter between the samples that the residues are all the same (see (2.9)). In that case, if we model the residues as non-random parameters, the ML estimator concentrates to

$$\hat{\mathbf{s}}_{\mathrm{ML}} = \arg\max_{\mathbf{s}} \mathbf{y}_{\mathrm{sum}}^{\mathrm{H}} \mathbf{B}(\mathbf{s})^{\mathrm{H}} \mathbf{A}(\mathbf{s})^{-1} \mathbf{B}(\mathbf{s}) \mathbf{y}_{\mathrm{sum}}, \tag{2.18}$$

where $\mathbf{y}_{\mathrm{sum}} = \sum_{m=1}^{M} \mathbf{y}_m^{\mathrm{H}}$. The difficulty of solving this problem is essentially the same as that of the problem in the single-shot case in (2.12).

While the model that leads to (2.17) is arguably pessimistic in terms of our knowledge of the system, the model that leads to (2.18) is rather optimistic and suffers from model mismatch in practical scenarios. As we explain in Appendix D (in the supplementary material), it is possible to develop multi-shot ML estimators based on more sophisticated models that incorporate synchronization offsets and sampling jitter. However, in the case where the offsets and jitter are modeled as non-random parameters, the resulting optimization over the concentrated (log) likelihood function has many more parameters than those in (2.17) and (2.18). In the case where the offsets and jitter are modeled as random parameters, the objective to be maximized is the marginalized likelihood, which is difficult to concentrate and each evaluation of the objective requires the evaluation of a multidimensional integral with many dimensions. Since both of these approaches to ML estimation under more sophisticated

models result in optimization problems that are significantly more difficult to solve than the problems in (2.17) and (2.18), which are themselves difficult to solve, there is considerable interest in the development of alternative estimation techniques.

2.4.2 Approximation of the Pole Distribution

The principles that underlie the model-free method of Barone [8] extend naturally to the multi-shot case. In particular, we can form an estimate of the complex frequencies $\{s_k\}$ from each shot using any of the single-shot estimators outlined in Section 2.3. The M different estimates of the set $\{s_k\}$ can then be used to construct an empirical distribution for the complex frequencies, or for the poles $\{e^{s_k}\}$. The local maxima of this empirical distribution are selected as the estimates of the complex frequencies, as illustrated in Figure 2.1. Barone [8] employs a simple, but somewhat ad-hoc, technique to determine the maxima. Instead, we will use the M estimates of $\{s_k\}$ to construct a kernel density estimate (KDE) [83] of the underlying distribution of the poles, and will then use smooth optimization techniques to determine the positions of the peaks of the KDE.

Although this technique is simple to describe and straightforward to implement, it retains a significant fraction of the noise sensitivity of the underlying single-shot estimator. As illustrated in Figure 2.1, this sensitivity can have a marked impact on the estimates of the complex frequencies, especially at lower SNRs. That said, Barone's (model-free) technique plays an important role in the initialization step of the optimization algorithm that is inherent in the model-based technique that we will develop in Section 2.6; see Algorithm 1. (Although we will focus on the use of Barone's method in the initialization of our approach, variations of the histogram

technique described in [41] could also be used.)

2.5 Distribution of Roots of Random Polynomials

As alluded to in the Introduction, and explained in more detail in the next section, the proposed method for estimating the complex frequencies will be based on a model for the distribution of the roots of the polynomial $Y_N(z)$ in (2.15), rather than the empirical distribution of the poles $\{e^{s_k}\}$ that is used in the approach of Barone. In preparation for the exposition of the proposed method, in this section we summarize the results of the Hammersley [40] on the distribution of the roots of a generic polynomial with Gaussian coefficients. Our summary is guided, in part, by that in [77].

For a generic $(N-1)$-th order polynomial $P(z) = \sum_{n=0}^{N-1} p_n z^n$ with $z \in \mathbb{C}$, if we are given a domain $\mathcal{D} \in \mathbb{C}$ with closed boundary $\partial\mathcal{D}$ that does not pass through any roots, the argument principle can be used to show that the number of roots of $P(z)$ that lie in $\mathcal{D}$ is

$$R_{\mathcal{D}} = \frac{1}{2\pi j} \oint_{\partial\mathcal{D}} \frac{P'(z)}{P(z)} \mathrm{d}z, \tag{2.19}$$

where $P'(z) = \sum_{n=1}^{N-1} p_n n z^{n-1}$. If the coefficients p_n of $P(z)$ are random with a known distribution, we would like to find a density function $\mathrm{pdf}(z)$ such that the expected fraction of roots of $P(z)$ in $\mathcal{D}$ can be computed as

$$\mathrm{E}\left(\frac{R_{\mathcal{D}}}{N-1}\right) = \int_{\mathcal{D}} \mathrm{pdf}(z)\, \mathrm{d}z.$$

By applying Green's theorem to (2.19), Hammersley [40] obtained such a density function for the cases of complex-valued and real-valued Gaussian coefficients.

For the complex-valued case, let us define $\mathbf{p} = \begin{bmatrix} p_0 & p_1 & \cdots & p_{N-1} \end{bmatrix}^{\mathrm{T}} \in \mathbb{C}^N$ and $\mathbf{z} = \begin{bmatrix} z^0 & z^1 & \cdots & z^{N-1} \end{bmatrix}^{\mathrm{T}} \in \mathbb{C}^N$ so that we can write $P(z) = \mathbf{p}^{\mathrm{T}}\mathbf{z}$. Let us also define the real-valued isomorphs $\check{\mathbf{p}} := \begin{bmatrix} \mathrm{Re}\left\{\mathbf{p}^{\mathrm{T}}\right\} & \mathrm{Im}\left\{\mathbf{p}^{\mathrm{T}}\right\} \end{bmatrix}^{\mathrm{T}} \in \mathbb{R}^{2N}$ and

$$\check{\mathbf{Z}} = \begin{bmatrix} \mathrm{Re}\ \mathbf{z}^{\mathrm{T}} & -\mathrm{Im}\ \mathbf{z}^{\mathrm{T}} \\ \mathrm{Im}\ \mathbf{z}^{\mathrm{T}} & \mathrm{Re}\ \mathbf{z}^{\mathrm{T}} \end{bmatrix} \in \mathbb{R}^{2\times 2N}.$$

The (Gaussian) distribution of the coefficients is specified through the mean $\check{\boldsymbol{\mu}} = \mathrm{E}(\check{\mathbf{p}})$ and covariance $\check{\boldsymbol{\Psi}} = \mathrm{E}\left(\check{\mathbf{p}}\check{\mathbf{p}}^{\mathrm{T}}\right) - \check{\boldsymbol{\mu}}\check{\boldsymbol{\mu}}^{\mathrm{T}}$. For notational convenience, we let $\check{\boldsymbol{\Phi}} = \mathrm{E}\left(\check{\mathbf{p}}\check{\mathbf{p}}^{\mathrm{T}}\right)$. If $\check{\mathbf{Z}}\check{\boldsymbol{\Psi}}\check{\mathbf{Z}}^{\mathrm{T}}$ is non-singular for all relevant z then the density of the roots can be written as [40]

$$\mathrm{pdf}(z) = \frac{\exp\left(-\frac{1}{2}\left(\check{\mathbf{Z}}\check{\boldsymbol{\mu}}\right)^{\mathrm{T}}\left(\check{\mathbf{Z}}\check{\boldsymbol{\Psi}}\check{\mathbf{Z}}^{\mathrm{T}}\right)^{-1}\left(\check{\mathbf{Z}}\check{\boldsymbol{\mu}}\right)\right)}{2\pi\,(N-1)\sqrt{\det\left(\check{\mathbf{Z}}\check{\boldsymbol{\Psi}}\check{\mathbf{Z}}^{\mathrm{T}}\right)}}\mathrm{tr}\left(\check{\boldsymbol{\Xi}}\check{\boldsymbol{\Phi}}\check{\boldsymbol{\Xi}}^{\mathrm{T}}\right), \tag{2.20}$$

where

$$\check{\boldsymbol{\Xi}} = \check{\mathbf{Z}}' - \left(\check{\mathbf{Z}}'\check{\boldsymbol{\Psi}}\check{\mathbf{Z}}^{\mathrm{T}}\right)\left(\check{\mathbf{Z}}\check{\boldsymbol{\Psi}}\check{\mathbf{Z}}^{\mathrm{T}}\right)^{-1}\check{\mathbf{Z}} \in \mathbb{R}^{2\times 2N},$$

and

$$\check{\mathbf{Z}}' = \begin{bmatrix} \mathrm{Re}\ \mathbf{z}'^{\mathrm{T}} & -\mathrm{Im}\ \mathbf{z}'^{\mathrm{T}} \\ \mathrm{Im}\ \mathbf{z}'^{\mathrm{T}} & \mathrm{Re}\ \mathbf{z}'^{\mathrm{T}} \end{bmatrix} \in \mathbb{R}^{2\times 2N},$$

where $\mathbf{z}' := \begin{bmatrix} 0 & z & 2z^1 & \cdots & (N-1)\,z^{N-2} \end{bmatrix}^{\mathrm{T}} \in \mathbb{C}^N$.

The case of real-valued Gaussian coefficients, which is of particular interest in our application, requires more care because $\check{\boldsymbol{\Psi}}$ is generally singular and, more importantly, $\check{\mathbf{Z}}\check{\boldsymbol{\Psi}}\check{\mathbf{Z}}^{\mathrm{T}}$ becomes singular on the real axis. Indeed, there is a discontinuity in the distribution as the imaginary part of z approaches zero. For the real-valued case, we

have

$$\mathrm{pdf}_{\text{real coeffs}}(z) = \begin{cases} \mathrm{pdf}(z), & z \in \mathbb{C}\backslash\mathbb{R} \\ \widetilde{\mathrm{pdf}}(z), & z \in \mathbb{R}, \end{cases} \tag{2.21}$$

where $\widetilde{\mathrm{pdf}}(z)$ is the real-axis density. If we define

$$\begin{aligned} \mathbf{z} &= \begin{bmatrix} z^0 & z^1 & \cdots & z^{N-1} \end{bmatrix}^{\mathrm{T}} \in \mathbb{R}^N, \\ \mathbf{z}' &= \begin{bmatrix} 0 & z & 2z^1 & \cdots & (N-1)\, z^{N-2} \end{bmatrix}^{\mathrm{T}} \in \mathbb{R}^N, \\ \mathbf{p} &= \begin{bmatrix} p_0 & p_1 & \cdots & p_{N-1} \end{bmatrix}^{\mathrm{T}} \in \mathbb{R}^N, \end{aligned}$$

$\boldsymbol{\mu} = \mathrm{E}(\mathbf{p})$, $\boldsymbol{\Psi} = \mathrm{E}\left(\mathbf{p}\mathbf{p}^{\mathrm{T}}\right) - \boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{T}} \in \mathbb{R}^{N\times N}$, $\boldsymbol{\Phi} = \mathrm{E}\left(\mathbf{p}\mathbf{p}^{\mathrm{T}}\right)$, $A_0 = \mathbf{z}^{\mathrm{T}}\boldsymbol{\mu}$, $A_1 = \mathbf{z}'^{\mathrm{T}}\boldsymbol{\mu}$, $U_{00} = \mathbf{z}^{\mathrm{T}}\boldsymbol{\Psi}\mathbf{z}$, $U_{01} = \mathbf{z}^{\mathrm{T}}\mathbf{C}\mathbf{z}'$ and $U_{11} = \mathbf{z}'^{\mathrm{T}}\mathbf{C}\mathbf{z}'$, the real-axis density can be written as

$$\widetilde{\mathrm{pdf}}(z) = \frac{\sqrt{U_{00}U_{11} - U_{01}^2}\exp\left(-\frac{1}{2}A_0^2/U_{00}\right)}{\sqrt{2\pi}\,(N-1)\,U_{00}}\left(r + \sqrt{2/\pi}H_1(r)\right), \tag{2.22}$$

where $r = |U_{00}A_1 - U_{01}A_0|\,/\sqrt{U_{00}\,(U_{00}U_{11} - U_{01}^2)}$,

$$\begin{aligned} H_1(r) &= \int_r^\infty (t-r)\exp\left(-\frac{1}{2}t^2\right)\mathrm{d}t \\ &= \exp\left(-\frac{1}{2}r^2\right) - \sqrt{\frac{\pi}{2}}r\,\mathrm{erfc}\left(\frac{r}{\sqrt{2}}\right), \end{aligned}$$

and the complementary error function is $\mathrm{erfc}(r) = (2/\sqrt{\pi})\int_r^\infty \exp(-t^2)\,\mathrm{d}t$.

2.6 Proposed Multi-Shot Estimation Method

As explained in Section 2.4, in a multi-shot estimation method for the complex frequencies, we have M measurement shots of the form in (2.6) and (2.7) over which the complex frequencies remain constant. In the proposed method, we will use those measurements to construct an empirical distribution for the roots of $Y_N(z)$ in (2.15). In scenarios with sufficient synchronization between shots and sufficiently low jitter, the residues in each shot can be assumed to be constant (cf. (2.9)), and hence the coefficients of Y_N can be modeled as Gaussian. Therefore, we can apply the model for the distribution of the roots that was outlined in the previous section. That model is an explicit function of the complex frequencies (and the residues). Therefore, in broad terms, we can develop an estimator by optimizing the estimates of the complex frequencies so as to minimize the difference between the model-based and empirical distributions. In this section, we provide the details of our invocation of that broad idea. We will develop the estimator for the case of real-valued measurements, and then we will show how the estimator can be simplified in the case of complex-valued measurements.

In order to construct the empirical distribution of the roots of $Y_N(z)$, we will use the kernel density estimation (KDE) technique [83]. To do so, let $Y_{N,m}(z)$ denote the z-transform of the time-reversed version of the m-th measurement shot, computed by analogy with (2.15), and let $\mathcal{R}_m$ denote the set of roots of $Y_{N,m}(z)$ with non-zero imaginary components, and let $\tilde{\mathcal{R}}_m$ denote the set of real-valued roots of $Y_{N,m}(z)$.

To construct the KDE of $\text{pdf}(z)$, we take the standard two-dimensional Gaussian kernel

$$g_b(z) = \frac{1}{2\pi b} \exp\left(-\frac{zz^*}{2b^2}\right),$$

where $z \in \mathbb{C}$. The empirical root distribution is then given by

$$\operatorname{kde}(z) = \frac{1}{M} \sum_{m=1}^{M} \frac{1}{|\mathcal{R}_m|} \sum_{r \in \mathcal{R}_m} g_b(z - r), \tag{2.23}$$

for $z \in \mathbb{C}$, where $|\mathcal{S}|$ denotes the cardinality of a set $\mathcal{S}$. The KDE of $\widetilde{\operatorname{pdf}}(z)$ is constructed in an analogous manner using the one-dimensional Gaussian kernel

$$\tilde{g}_b(z) = \frac{1}{b\sqrt{2\pi}} \exp\left(-\frac{z^2}{2b^2}\right),$$

where $z \in \mathbb{R}$. It is given by

$$\widetilde{\operatorname{kde}}(z) = \frac{1}{M} \sum_{m=1}^{M} \frac{1}{|\tilde{\mathcal{R}}_m|} \sum_{r \in \tilde{\mathcal{R}}_m} \tilde{g}_b(z - r), \tag{2.24}$$

for $z \in \mathbb{R}$.

In order to construct the model distribution for the roots, we observe that with sufficient synchronization between shots and sufficiently low jitter, the received signal samples can be modeled as being Gaussian, with the mean being an explicit function of the complex frequencies $\mathbf{s}$ and the residues $\mathbf{r}$, and the covariance being the noise covariance; see Section 2.7.1 for details. We will use the notation $\operatorname{pdf}_{\mathbf{s},\mathbf{r}}(z)$ and $\widetilde{\operatorname{pdf}}_{\mathbf{s},\mathbf{r}}(z)$ to make that dependence on $\mathbf{s}$ and $\mathbf{r}$ explicit. In Sections 2.7.2 and 2.7.3, we show how this model can be adapted to approximate other scenarios.

With these definitions in place, the proposed estimator seeks to estimate the complex frequencies $\mathbf{s}$ by determining the values for both $\mathbf{s}$ and the residues $\mathbf{r}$ (which are nuisance parameters in our application) that position the model-based distribution

close to the empirical distribution in an appropriate sense. To determine an appropriate sense, we first observe that both (2.23) and (2.24) are scaled such that their integrals are individually 1. In contrast, (2.20) and (2.22) are more correctly scaled, in the sense that the overall integral of the density is 1. Thus, if

$$\tilde{W} = \int_{-\infty}^{\infty} \widetilde{\mathrm{pdf}}_{\mathbf{s},\mathbf{r}}(z)\,\mathrm{d}z, \tag{2.25}$$

then $\widetilde{\mathrm{pdf}}_{\mathbf{s},\mathbf{r}}(z)$ should be compared to $\tilde{W}\widetilde{\mathrm{kde}}(z)$, and $\mathrm{pdf}_{\mathbf{s},\mathbf{r}}(z)$ should be compared to $W\mathrm{kde}(z)$, where $W = 1 - \tilde{W}$. In practice, the integral in (2.25) is computed numerically. Since the density is concentrated near zero, that computation is straightforward. In addition to this scaling, the appropriate error metric should also incorporate the nature of the kernel used in the KDE. In particular, we observe that in the limit of a large number of measurements $W\mathrm{kde}(z)$ approaches the convolution of $\widetilde{\mathrm{pdf}}_{\mathbf{s},\mathbf{r}}(z)$ and the kernel $g_b(z)$ used to construct $\mathrm{kde}(z)$, which we will denote by $\left(\mathrm{pdf}_{\mathbf{s},\mathbf{r}} * g_b\right)(z)$. Analogously, $\widetilde{\mathrm{kde}}(z)$ approaches $\left(\widetilde{\mathrm{pdf}}_{\mathbf{s},\mathbf{r}} * \tilde{g}_b\right)(z)$. With those observations in mind, let us define

$$\mathrm{KDE}(z) = \begin{cases} W\mathrm{kde}(z), & z \in \mathbb{C}\backslash\mathbb{R} \\ \tilde{W}\widetilde{\mathrm{kde}}(z), & z \in \mathbb{R}, \end{cases}$$

and

$$\mathrm{PDF}_{\mathbf{s},\mathbf{r}}(z) = \begin{cases} \left(\mathrm{pdf}_{\mathbf{s},\mathbf{r}} * g_b\right)(z), & z \in \mathbb{C}\backslash\mathbb{R} \\ \left(\widetilde{\mathrm{pdf}}_{\mathbf{s},\mathbf{r}} * \tilde{g}_b\right)(z), & z \in \mathbb{R}. \end{cases}$$

Then, given some appropriate notion of the difference between these functions, denoted $\text{error}(\cdot)$, the proposed estimator can be written as

$$\{\hat{\mathbf{s}}, \hat{\mathbf{r}}\} = \arg \min_{\{\mathbf{s},\mathbf{r}\} \in \mathcal{X}} \text{error}\big(\text{KDE}(z), \text{PDF}_{\mathbf{s},\mathbf{r}}(z)\big), \tag{2.26}$$

where we use the set $\mathcal{X}$ to denote that for real-valued measurements, the complex conjugate relationships between the elements of $\mathbf{s}$ and $\mathbf{r}$ discussed after (2.1) mean that the problem in (2.26) can be reduced an optimization problem over the s_k with positive imaginary parts, and the corresponding residues.

There are many appropriate metrics that could be considered for the error. In this chapter, we select the symmetrized Kullback-Leibler divergence [26] (also known as the Jeffreys divergence) over a domain $\mathcal{D} \subset \mathbb{C}$; i.e.,

$$\text{error}\big(p(z), q(z)\big) = \int_{\mathcal{D}} p(z) \log \frac{p(z)}{q(z)} + q(z) \log \frac{q(z)}{p(z)} \mathrm{d}z. \tag{2.27}$$

The choice of domain the $\mathcal{D}$ in (2.27) involves a judicious trade-off between accuracy and computational cost. Since we consider the case of real-valued measurements, we can restrict $\mathcal{D}$ to the closed upper half of the complex plane, i.e., $\text{Im}(z) \geq 0$. Evaluation of the error along the real axis, if desired, is comparatively efficient and hence we will focus on domains the open upper half-plane.

In the case of zero-mean uncorrelated Gaussian coefficients (i.e., a "white noise only" signal), the roots are concentrated around the unit circle and uniformly distributed in phase except in the neighborhoods of $z = \pm 1$ [40]. In the case of exponentially decaying signals in Gaussian noise, the roots remain concentrated around the unit circle, but their distribution is significantly different from the noise-only case

near the poles $\{e^{s_k}\}$. It is less different elsewhere. This suggests that $\mathcal{D}$ should be chosen to encompass the envisioned poles. To obtain some guidelines for that region, we observe that in practice the frequencies can only be measured within the bandwidth of the radar. This provides natural bounds on the phase of points $z \in \mathcal{D}$. To obtain guidelines on radial bounds, consider bounds of the form $1 - A \leq |z| \leq 1 + B$, where $A, B \geq 0$. If the resonant response decays to below the noise level faster than one sample period, then it is unlikely to be detectable, and that provides a guideline for choosing A. To choose B, we note that the resonances are decaying and hence will correspond to poles inside the unit circle. However, we are measuring the difference between the empirical and model-based distribution of the roots of $Y_N(z)$. Hence, a reasonable guideline is to choose B to be a reasonably small positive number. This analysis suggests that a good choice for $\mathcal{D}$ is the intersection of the cone specified by the bandwidth of the radar (and the sampling period) with the annulus that is specified by the parameters A and B. That annulus would include the unit circle. A computationally efficient scheme for carrying out integration over such a domain is outlined in Appendix 2.A. The domain $\mathcal{D}$ that we have used in our numerical experiments is shown in Figure 2.2.

In the case of complex-valued measurements, the structure of the model-based density is simplified in the sense that it is simply $\mathrm{pdf}_{\mathbf{s},\mathbf{r}}(z)$ for all $z \in \mathbb{C}$. Therefore, the estimator is simply

$$\{\hat{\mathbf{s}}, \hat{\mathbf{r}}\} = \arg\min_{\{\mathbf{s},\mathbf{r}\}} \mathrm{error}\left(\mathrm{kde}(z), \left(\mathrm{pdf}_{\mathbf{s},\mathbf{r}} * g_b\right)(z)\right).$$

In this setting, however, we no longer have conjugate symmetry in the root distribution and the region $\mathcal{D}$ over which the error is evaluated may need to be significantly

larger.

2.6.1 Implementation

In order to implement the proposed approach, we need to specify the system model, and in particular the functional dependence of the mean $\check{\boldsymbol{\mu}}$ and covariance $\check{\boldsymbol{\Psi}}$ in Section 2.5 on the complex frequencies $\mathbf{s}$ and residues $\mathbf{r}$. As we show in Section 2.7.1, this is straightforward in the case of sufficient synchronization between shots, sufficiently low jitter and additive Gaussian noise. However, we also show how to build appropriate models for scenarios in which the jitter and synchronization errors cannot be neglected; see Sections 2.7.2 and 2.7.3, respectively.

Once the model is in place, we need to develop an effective algorithm to solve the relevant optimization problem. We will focus on the case of real-valued measurements, and hence the relevant problem is that in (2.26). While that problem is smooth, it has a complicated optimization landscape. We will employ a local optimization technique, and hence an effective initialization is important. We will employ a two-stage initialization for the estimates of the complex frequencies, and will subsequently initialize the residues using (2.11).

In the first stage of the initialization of the complex frequencies, we employ a variant of Barone's technique [8]. The first step of that technique involves constructing estimates of the complex frequencies from each of the M measurement shots using a single-shot technique. (We use the GPoF technique described in Section 2.3.2.) Those estimates are used to create a kernel density estimate of the distribution of the poles $\{e^{s_k}\}$ using an expression analogous to (2.23). The first-stage estimates of the complex frequencies are obtained from the peaks of that KDE.

In the second stage of the initialization, we use the estimates of $\mathbf{s}$ from the first stage to initialize a search for good solutions to the problem in (2.18). (Akin to the discussion after (2.26), that problem can be reduced to an optimization problem over the elements of $\mathbf{s}$ with positive imaginary parts.) Given the nature of the optimization landscape of that problem, conventional local optimization from a single initial point might not provide sufficient refinement over the initialization from the first stage. Instead, we spawn multiple initial points in the neighborhood of the result from the first stage. These points constitute an initial population from which the differential evolution technique [79] is used to generate multiple candidates for (2.18). The candidate with the largest objective value is chosen as the initialization for the complex frequencies $\mathbf{s}$ in the proposed method. In this step, we constrain the estimates of $\mathbf{s}$ so that the (normalized) frequencies lie within the bandwidth of the radar and the attenuation rates lie within a physically reasonable range. We will denote the combination of this constraint region and the complex conjugate relationships by $\mathbf{s} \in \mathcal{S}$, and in Figure 2.2 we illustrate the constraint region that we use in our numerical experiments in terms of the constraints that it imposes on the pole estimates $\{e^{\hat{s}_k}\}$. (Any spawned initial point outside $\mathcal{S}$ is projected to its boundary before the differential evolution step begins.) We will refer to the estimates from this second stage as the constrained non-linear least-squares (CNLS) estimates of $\mathbf{s}$.

With this two-stage initialization in place, and the subsequent initialization of $\mathbf{r}$ using (2.11), the final stage of the algorithm is to perform a local optimization of the expression in (2.26) using a bound-constrained BFGS quasi-Newton method (e.g., [29]), that constraints the estimates of $\mathbf{s}$ to $\mathcal{S}$. This approach is summarized in Algorithm 1. Our numerical results in Section 2.8 show that, at low SNRs, each

Algorithm 1 Estimation of complex frequencies using Hammersley's model for the the distribution of of z-domain roots

Input: Measurements $\{y_m[n]\}_{m=1,n=0}^{M,N-1}$, number of resonances $K/2$, noise covariance $\mathbf{C}$, model that describes $\check{\boldsymbol{\mu}}$ and $\check{\boldsymbol{\Psi}}$ as a function of $\mathbf{s}$, $\mathbf{r}$ and $\mathbf{C}$, domain $\mathcal{D}$, and region of viable complex frequency estimates $\mathcal{S}$.

Step 1: Obtain an initial estimate of $\mathbf{s} \in \mathcal{X}$, denoted $\hat{\mathbf{s}}_{\text{Bar.}}$, by applying our variant of Barone's technique.

Step 2: Refine $\hat{\mathbf{s}}_{\text{Bar.}}$ to $\hat{\mathbf{s}}_{\text{CNLS}} \in \mathcal{S}$ using differential evolution to find a good solution to the nonlinear least-squares problem in (2.18) over $\mathbf{s} \in \mathcal{S}$.

Step 4: Construct the empirical distribution $\text{kde}(z)$ over $z \in \mathcal{D}$ from the measurements (and $\widetilde{\text{kde}}(z)$ if $\mathcal{D}$ intersects the the real axis).

Step 5: Refine $\hat{\mathbf{s}}_{\text{CNLS}}$ to $\hat{\mathbf{s}}_{\text{Hamm.}} \in \mathcal{S}$ using bound-constrained BFGS to find a local solution of the optimization problem in (2.26).

Output: The estimated complex frequencies $\hat{\mathbf{s}}_{\text{Hamm.}}$.

stage of the algorithm yields significant improvements in the estimates of the complex frequencies.

2.7 Models of Practical Signals

In order to employ the method described in the previous section, we need to construct a model for the coefficients of the z-transform of (the time-reversed version of) the received signal that is Gaussian and in which the mean $\check{\boldsymbol{\mu}}$ and covariance matrix $\check{\boldsymbol{\Psi}}$ of the real-valued isomorphs of the coefficients (see Section 2.5) are explicit functions of the complex frequencies $\mathbf{s}$ and residues $\mathbf{r}$. As we show in Section 2.7.1, this is straightforward when the measurement jitter and synchronization errors are

negligible. However, we also show how approximate models can also be constructed in the cases of non-negligible jitter and synchronization error. While the Gaussian approximation yields models that are inherently mismatched, in Section 2.8, we show that they lead to very effective estimates in practical scenarios.

2.7.1 Additive Gaussian Noise

In the case in which the sampling jitter and synchronization errors are sufficiently small, the received signal in the m-th measurement shot can be modeled as $y_m[n] = x[n] + v_m[n]$, where, as in (2.3), $x[n] = \sum_{k=1}^{K} r_k e^{s_k n}$, and $v_m[n]$ is the additive Gaussian noise, which has zero mean and covariance matrix $\mathbf{C}$. In this case, $[\boldsymbol{\mu}]_n = \mathrm{E}(y_m[n]) = \sum_{k=1}^{K} r_k e^{s_k n}$, and the coefficients are Gaussian with

$$[\boldsymbol{\Psi}]_{nn'} = \mathrm{E}\left(v_m[n]\, v_m[n']^*\right) = [\boldsymbol{C}]_{nn'}.$$

The real-valued isomorphs $\check{\boldsymbol{\mu}}$ and $\check{\boldsymbol{\Psi}}$ can be computed in a similar way. In this setting, it is only the mean of the coefficients $\boldsymbol{\mu}$ that depends on $\mathbf{s}$ and $\mathbf{r}$.

2.7.2 Measurement Jitter

As described in the Introduction, significant sampling jitter can arise in LTR measurements, especially when equivalent sampling is used. To model the effects of this, we assume we have multiple shots which are synchronized, but are subject to jitter, so the m-th measurement shot can be written as

$$y_m[n] = y\left(t_0 + (n + \varsigma\gamma_m[n])\, T\right),$$

where $y(t)$ is given in (2.1) and $\varsigma\gamma_m[n]$ is the jitter, where each $\gamma_m[n]$ is independent and is identically distributed according to $\mathcal{N}(0,1)$, and $\varsigma \geq 0$. In this case, we can write

$$y_m[n] = \sum_{k=1}^{K} r_k e^{s_k(n+\varsigma\gamma_m[n])} + v_m[n], \tag{2.28}$$

where we implicitly assume that the timing jitter does not affect the distribution of the noise samples. Although the distribution of $y_m[n]$ in (2.28) is not Gaussian, we can compute its mean and covariance and apply the technique developed in Section 2.6 to the computed mean and covariance. While that involves a fundamental mismatch in the model, if that mismatch is small, we may still obtain good results. Indeed, the first-order Taylor series approximation of (2.28) is Gaussian, and hence for small values of ς the model mismatch is likely to be small.

As shown in Appendix 2.B, the mean of $y_m[n]$ is

$$[\boldsymbol{\mu}]_n = \mathrm{E}(y_m[n]) = \sum_{k=1}^{K} r_k e^{s_k n + s_k^2\varsigma^2/2} \tag{2.29}$$

and the (n,n')-th element of the cross-correlation takes the form

$$[\boldsymbol{\Phi}]_{nn'} = \mathrm{E}\left(y_m[n]\, y_m[n']^*\right) = \sum_{k=1}^{K}\sum_{k'=1}^{K} r_k r_{k'}^* e^{s_k n + s_{k'} n' + \rho^2_{nn',kk'}/2} + [\mathbf{C}]_{nn'}, \tag{2.30}$$

where $\mathbf{C}$ is the covariance of $v_m[n]$ and

$$\rho^2_{nn',kk'} = \begin{cases} \varsigma^2 (s_k + s_{k'}^*)^2, & n = n' \\ \varsigma^2 \left(s_k^2 + (s_{k'}^*)^2\right), & n \neq n'. \end{cases} \tag{2.31}$$

The definitions in Section 2.6 can then be applied to the real-valued isomorphs of

these expressions.

2.7.3 Synchronization Offsets

Even if the sampling jitter is negligible, we may still need to consider synchronization offsets between the measurement shots. As an example, consider a situation where a target is placed roughly in front of the radar but shifts back and forth. This might occur if we are attempting to identify clutter on the body of a person who fails to stand perfectly still. It is reasonable to assume the shift is constant for any given measurement, but is not the same from one measurement to the next. In our lab, we have noted that a person attempting to stand still will typically wobble up to a few centimeters. Light travels 3 cm in 100 ps; thus a wobble with a span of 3 cm will affect the round-trip time by up to 200 ps, or 4 samples at $T = 50\,\text{ps}$. If the wobble is modeled as being Gaussian, then its standard deviation would be roughly one quarter of that, i.e., up to 0.75 cm or 50 ps or 1 sample. Another case which is modeled equivalently is jitter in the transmitter's trigger, which causes the pulse to be emitted at slightly different times thus also causing a wholesale delay in the received signal.

Referring to our model in (2.6), synchronization errors present as a particular sort of residue uncertainty. The residue can be thought to have a (real) *amplitude* component a_k and *phase* ϕ_k such that $r_k = a_k e^{j\phi_k}$.

For the analysis, we opt to ignore the effect on the amplitudes a_k. In the case of a plane-wave stimulus in the far field, the entire signal scales quadratically with distance; i.e., every a_k scales quadratically with distance. Scaling an entire polynomial by the same constant has no impact on its roots, and thus this scaling would have

no impact at all in the absence of additive noise. However, the signal-to-noise ratio would change in the noisy case. We choose to ignore this also since the effect is often negligible in practice, as we now explain. Assume some uncertainty δ relative to a distance d. The received amplitudes would be

$$\left(\frac{d+\delta}{d}\right)^2 a_k.$$

Typically, d is in excess of 1 m and δ has a standard deviation of roughly 0.75 cm per the analysis above, thus the fraction is approximately 1.

We cannot similarly ignore the effect on the phase ϕ_k. As the target distance to the radar changes, so does the round-trip time. This is much like the jitter case, except the offset is the same across all samples n. The signal can therefore be modeled similarly:

$$y_m[n] = \sum_{k=1}^{K} r_k e^{s_k(n+\nu\varphi_m)} + v_m[n],$$

where φ_m is i.i.d. and distributed as $\mathcal{N}(0,1)$. With the exception of it being constant over all samples, the randomness induced by the synchronization offset has a similar structure to that of jitter, and thus the analysis is similar. The expected value is

$$\mathrm{E}(y_m[n]) = \sum_{k=1}^{K} r_k e^{s_k n + s_k^2 \nu^2/2},$$

and the cross-correlations are

$$\mathrm{E}(\bar{y}_n \bar{y}_{n'}^*) = \sum_{k=1}^{K}\sum_{k'=1}^{K} r_k r_{k'}^* e^{s_k n + s_{k'} n' + \nu^2 \left(s_k + s_{k'}^*\right)^2/2} + [\mathbf{C}]_{nn'}.$$

2.8 Numerical Experiments

In this section, we examine the performance of the proposed estimation approach in a variety of simulated scenarios. We consider real-valued measurements, and hence, as discussed after (2.1), the (normalized) complex frequencies that we seek to identify take the form $s_k = -\alpha_k + j2\pi f_k$ with $f_k > 0$, and $s_{k+K/2} = s_k^*$, for $k = 1, 2, \ldots, K/2$. We consider a radar system with a bandwidth that corresponds to normalized frequencies $f \in [0.05, 0.25]$, and our estimates of $\{s_k\}$ will be constrained so that each f_k lies in that range. (Under this model, the sampling rate has been chosen to be four times the upper edge of the radar's bandwidth.) We also consider an application in which the attenuation rates of interest lie in the range $\alpha_k \in [0, 0.1]$. Therefore, our estimates of the poles $\{e^{s_k}\}_{k=1}^{K/2}$ will be constrained to lie in the region bounded by the black curves in Figure 2.2. With the poles of interest, and their estimates, being constrained to that region, it is sufficient to choose the region $\mathcal{D}$ in (2.27) to be the shaded region in Figure 2.2.

In our experiments we compare the performance of the proposed estimator in Algorithm 1, $\hat{\mathbf{s}}_{\text{Hamm.}}$, against Barone's method [8], $\hat{\mathbf{s}}_{\text{Bar.}}$, which we use as the first stage of the initialization of our method, and the constrained nonlinear least squares (CNLS) estimate, $\hat{\mathbf{s}}_{\text{CNLS}}$, in (2.18), which we use as the second stage of the initialization of our method. In each Monte Carlo trial, a set of 5 conjugate pairs of (normalized) complex frequencies $s_k = -\alpha_k \pm 2\pi j f_k$ was randomly generated, with the attenuation $\alpha_k \in [0.0, 0.1]$ and frequency $f_k \in [0.05, 0.25]$ being generated according to the uniform distribution on the stated intervals. For any two poles e^{s_k} and e^{s_ℓ}, where $\ell \neq k$, a minimal distance was set as $|e^{s_k} - e^{s_\ell}| > 0.015$. Any trial violating this criterion was removed and replaced. The magnitudes of the residues were generated

from the uniform distribution on $[0.8, 1]$, and the corresponding phases were uniformly distributed on $[0, 2\pi]$. For each trial, a set of $M = 40$ measurement shots of length $N = 48$ was then generated using the measurement model that is being investigated.

To provide some context for the errors that we obtain, let us consider some physical quantities. A highly conductive ($\alpha_k \approx 0$) half-wavelength dipole of length 15 cm resonates at approximately 1 GHz. With a sampling frequency of 20 GHz, this is a normalized frequency of $f_k = 0.05$, and the corresponding pole is approximately $e^{j2\pi f_k}$. An error of 0.01 in the estimate of $e^{j2\pi f_k}$ corresponds to an error of up to $0.01/2\pi$ in f_k, which corresponds to around 32 MHz. Assuming a phase velocity of roughly 30 cm/ns, the estimated wavelength would lie between roughly 29 cm and 31 cm, and hence the estimate of the corresponding dipole length would be between 14.5 cm and 15.5 cm. This is likely sufficient to identify the 15-cm dipole as such. In contrast, an error of 0.1 in $e^{j2\pi f_k}$ would result in an estimated dipole length between 11 cm and 22 cm, which is likely insufficient to identify the 15-cm dipole as such. For this reason, especially at lower frequencies, we consider errors on the order of 0.01 to be acceptable and those on the order of 0.1 to be unacceptable.

Before we discuss the results of our Monte Carlo experiments, let us consider an illustrative example of a single typical trial in the case of perfect synchronization and no sampling jitter, with additive white Gaussian noise at an SNR of 0 dB. As shown in Figure 2.2, in this trial the initial estimates from Barone's method (in green) have appreciable errors, but these estimates are refined using the second initialization step (the CNLS estimates in blue). The proposed "Hammersley" estimator is initialized by the CNLS estimates. The improved estimates that it provides are visible (in orange) in the lower plot in Figure 2.2. For context, Figure 2.2 also illustrates the region

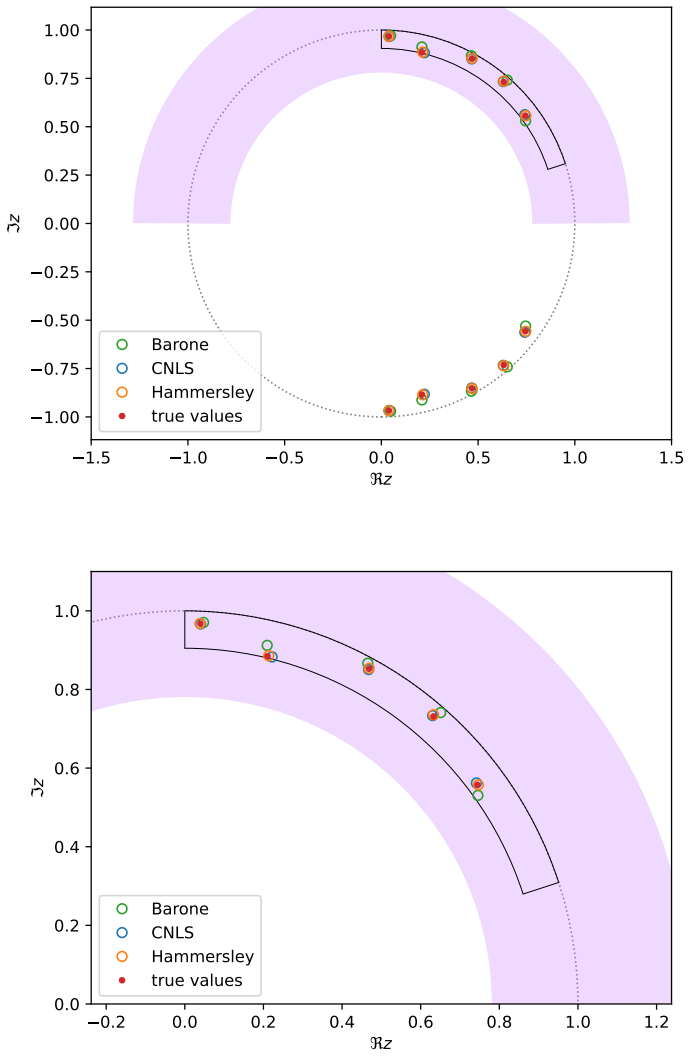

Figure 2.2: Illustrative results for the three considered methods for one trial of the white noise case at 0 dB SNR. Views of all four quadrants (top) and only Quadrant I (bottom) are given. The constraint region for the poles is outlined in black, and the domain $\mathcal{D}$ in (2.27) is shaded.

within which the pole estimates are constrained to lie (by the constraint that $\hat{\mathbf{s}} \in \mathcal{S}$; see Section 2.6.1). This region is outlined by the solid curves, and corresponds to attenuations $0 < \alpha < 0.1$ and frequencies $0.05 < f < 0.25$. The domain $\mathcal{D}$ over which the error in (2.27) is computed is shaded. It corresponds to $-0.25 < \alpha < 0.25$ and $0 < f < 0.5$.

For our Monte Carlo experiments, we will consider both the average and worst-case errors. We observe that having a small worst-case error is likely to lead to accurate classification, and hence the worst-case error is important. However, for targets with multiple resonances, it is possible that misidentifying one or two resonances might still lead to an accurate classification in the subsequent stages. Thus, the average-case performance in the geometric sense is also relevant. For this reason, both are reported.

Figure 2.3 shows the results for the case of additive white Gaussian noise with perfect synchronization and no jitter. We observe that, for SNRs in the range $[0, 10]$ dB, the proposed estimator has an average error that is more than an order of magnitude smaller than that of Barone's method [8] and around two times smaller than that of the CNLS method. Perhaps more interestingly, in the range $[5, 20]$ dB, the worst-case performance of the proposed method is comparable to or better than the average performance of the comparator methods. The worst-case performance of proposed method is below 0.02 in the range $[10, 25]$ dB, which is likely to be acceptably low in a classification problem in our intended application, whereas the comparator methods only achieve this at 20 dB. The grid size for the numerical approximation of the integral in (2.27) (see Appendix 2.A) was chosen to balance performance at low-to-moderate SNRs against computational cost, and leads to a visible error floor

for the proposed method. If increasingly accurate approximations were chosen as the SNR was increased, the proposed approach would extend its advantage over the CNLS method, but with a diminishing performance gap.

Figure 2.4 shows the results for the case of increasing sampling jitter in the presence of additive noise such that the SNR (excluding jitter) is 10 dB. We observe that, for jitter with standard deviation in the range $[0.0625, 1.0]$ samples, the proposed method has an average error that is more than an order of magnitude lower than Barone's method and around two times lower than the CNLS approach. The worst-case error is below 0.02 in the range $[0.0625, 0.5]$ samples, which is likely to be acceptably low in our envisioned classification problem.

Figure 2.5 shows the results for the case of increasing synchronization offsets in the presence of additive noise such that the SNR is 10 dB. We observe that, for offsets with standard deviation in the range $[0.0625, 1.0]$ samples, the proposed method has an average error that is around an order of magnitude lower than Barone's method and around two times lower than the CNLS approach. The worst-case error is at or below 0.02 in the range $[0.0625, 0.5]$ samples, which is likely to be acceptably low in our envisioned classification problem.

In closing, we point out that although the proposed method provides significantly better performance than Barone's method [8] in the settings that we have considered, Barone's method plays a critical role in the initialization of our method. Indeed, although it is possible, in principle, to apply a global optimization technique to the distribution-matching problem in (2.26), it is computationally expensive to do so. Barone's method enables us to efficiently obtain good starting points for the nonlinear least-squares estimator, from which we obtain a good starting point for solving (2.26).

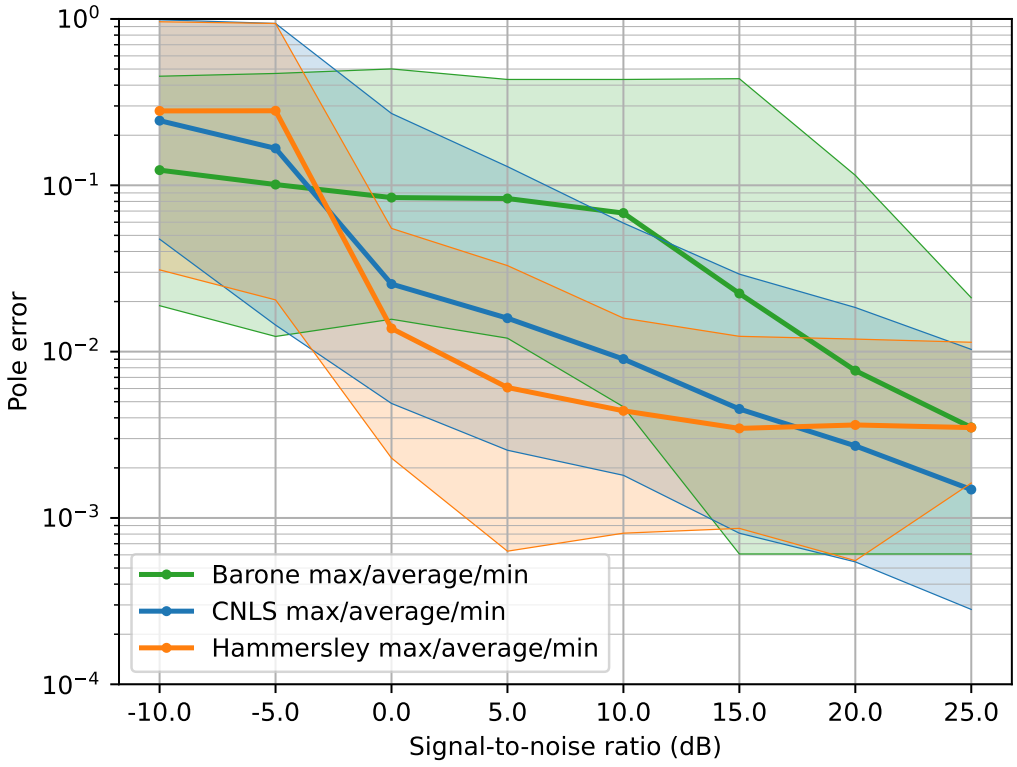

Figure 2.3: Errors in the estimates of the poles $\{e^{s_k}\}$ for the different levels of additive white Gaussian noise.

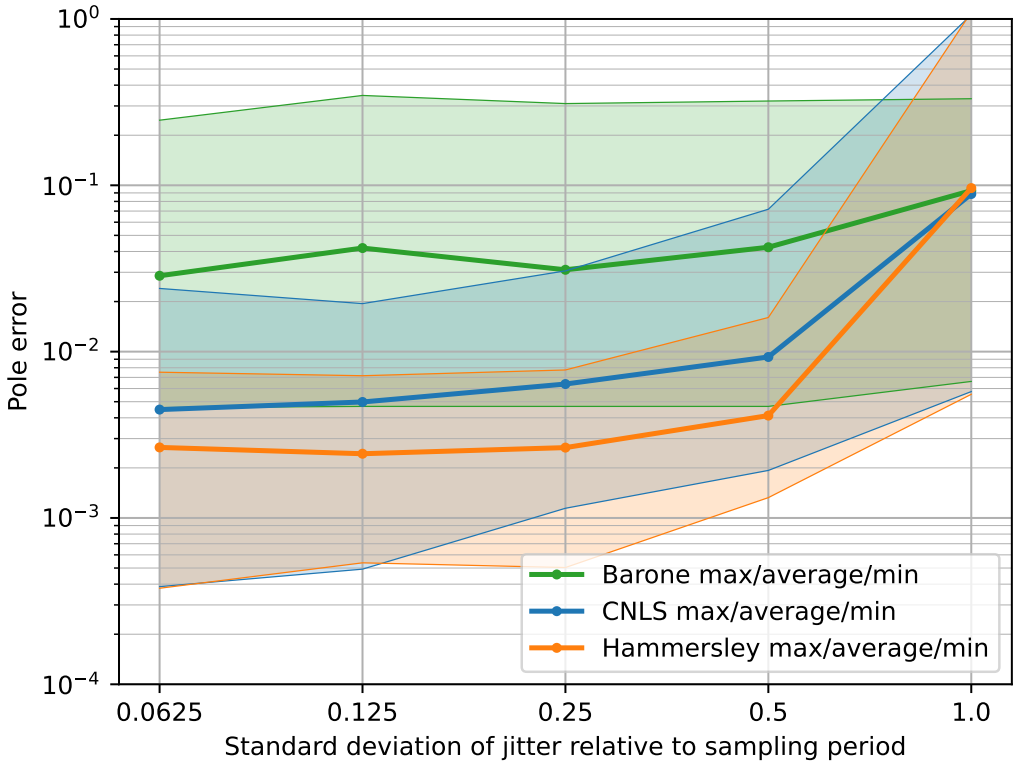

Figure 2.4: Errors in the estimates of the poles $\{e^{s_k}\}$ for the different levels of sampling jitter with additive white Gaussian noise at a fixed SNR of 10 dB.

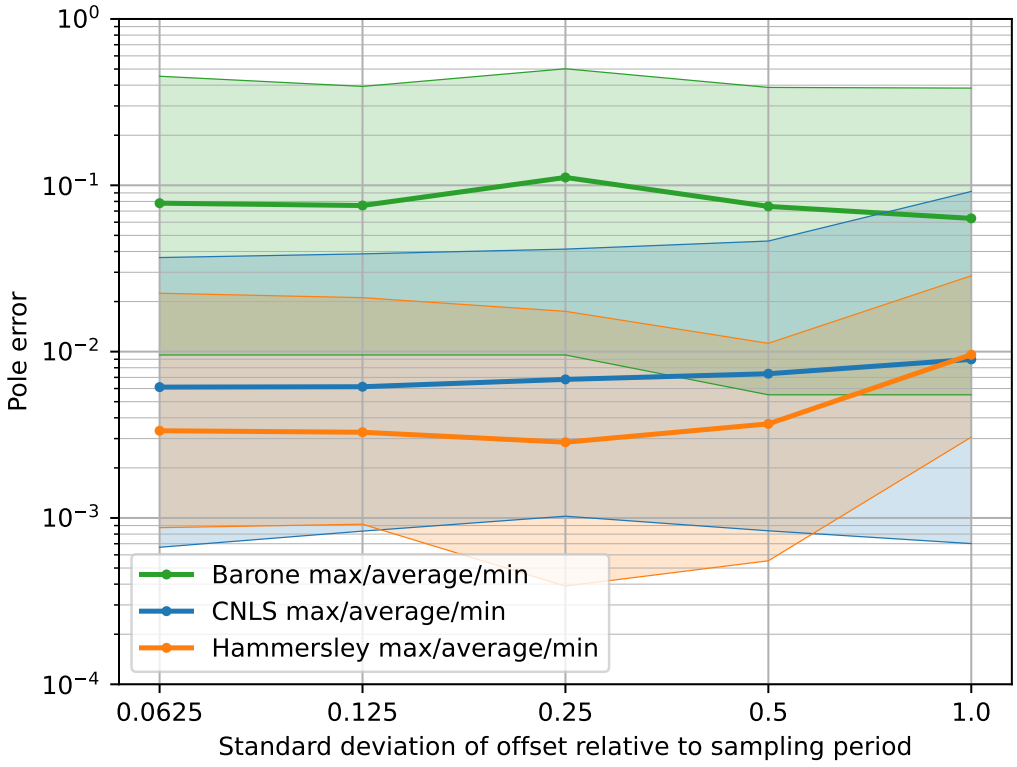

Figure 2.5: Errors in the estimates of the poles $\{e^{s_k}\}$ for the different levels of synchronization offset with additive white Gaussian noise at a fixed SNR of 10 dB.

2.9 Conclusion

We have demonstrated a new method for estimating the damped-sinusoid resonances of signals in noise. The method relies on minimizing the error between the observed distribution of z-domain roots of the signal and its theoretical counterpart. We have shown that this method is effective not only for additive Gaussian noise, for which we have an accurate theoretical counterpart, but also for other sorts of uncertainly for which we have some model mismatch, of which jitter is particularly interesting. We have demonstrated a significant improvement over prior art at low SNR (typically $[0, 10]$ dB). As such, the proposed method is well suited to the extraction of resonance parameters from late-time radar returns.

Appendices

2.A Efficient Computation of the Objective Function

The computation of the objective in (2.26) involves a 2-D convolution on a domain $\mathcal{D} \subset \mathbb{C}$, which is the intersection of a cone and annulus, i.e.,

$$\mathcal{D} = \left\{ e^{-\alpha + j\omega} | \alpha_{\min} \leq \alpha \leq \alpha_{\max}, \omega_{\min} \leq \omega \leq \omega_{\max} \right\}.$$

The domain $\mathcal{D}_s$ of $s = -\alpha + j\omega$ lends itself to a regular rectangular grid, but the original domain $\mathcal{D}$ of $z = e^s$ does not. Our approach is thus to construct a regular rectangular grid $\{s_{p,q}\}$ in $\mathcal{D}_s$, where $s_{p,q} = -\alpha_p + j\omega_q$, $\alpha_p = \alpha_{\min} + (p + 1/2)\left(\alpha_{\max} - \alpha_{\min}\right)/P$

and $\omega_q = \omega_{\min} + (q + 1/2)(\omega_{\max} - \omega_{\min})/Q$ for grid sizes $P, Q \in \mathbb{Z}^+$ and indices $p = 0, \ldots, P-1$ and $q = 0, \ldots, Q-1$. The rectangular grid points $\{s_{p,q}\}$ are then transformed to the log-polar grid points $\{z_{p,q}\}$, where $z_{p,q} = e^{s_{p,q}}$, so as to numerically approximate the convolution in $\mathcal{D}$.

To that end, note that the convolution of $f(z)$ and $g(z)$ over $\mathcal{D}$ is

$$(f * g)(z) = \iint_{\mathcal{D}} f(z')\, g(z - z')\, \mathrm{d}z',$$

where $\mathrm{d}z'$ is a surface element in $\mathcal{D}$. The function that is being swept, $g(z - z')$, is called the *kernel* or *stencil*. An approximation of the convolution is

$$(f * g)_{\mathrm{FD}}(z) = \sum_{p',q'} f(z_{p',q'})\, g(z_{p+p',q+q'})\, \mathrm{area}(z_{p+p',q+q'}),$$

where $\mathrm{area}(z_{p,q})$ denotes the area of the grid cell with corners at $z_{p-1/2,q-1/2}$, $z_{p-1/2,q+1/2}$, $z_{p+1/2,q-1/2}$ and $z_{p+1/2,q+1/2}$. With $u_1 = z_{p-1/2,q+1/2} - z_{p-1/2,q-1/2}$, $u_2 = z_{p+1/2,q+1/2} - z_{p-1/2,q-1/2}$ and $u_3 = z_{p+1/2,q-1/2} - z_{p-1/2,q-1/2}$, this area is

$$\mathrm{area}(z_{p,q}) = \frac{1}{2}\left|\mathrm{Im}(u_1^* u_2)\right| + \frac{1}{2}\left|\mathrm{Im}(u_2^* u_3)\right|. \tag{2.32}$$

To justify (2.32), consider the vector isomorphs $\mathbf{u}_n = \begin{bmatrix} \mathrm{Re}(u_n) & \mathrm{Im}(u_n) & 0 \end{bmatrix}^{\mathrm{T}}$ of u_n. The area of a triangle with sides $\|\mathbf{u}_m\|_2$, $\|\mathbf{u}_n\|_2$ and $\|\mathbf{u}_m - \mathbf{u}_n\|_2$ is $\frac{1}{2}\|\mathbf{u}_m \times \mathbf{u}_n\|_2 = \frac{1}{2}\left|\mathrm{Im}(u_m^* u_n)\right|$, where $\times$ denotes the cross product.

Note that, unlike discrete convolution on a rectangular grid, neither the stencil $g(z_{p+p',q+q'})$ nor the area area$(z_{p+p',q+q'})$ are the same for all (p, q). However, they can be computed once and applied to different inputs $f(z_{p',q'})$, so this does not present a

significant computational overhead.

2.B Cross-Correlation of Complex Exponential Signals with Jitter

To derive an expression for the cross-correlation of a signal of the form in (2.28), we need the complex moments of a log-normal random variable. It is known that the c-th real moment of a log-normal random variable (about zero) is

$$\mathrm{E}\left(e^{(\mu+\sigma X)c}\right) = e^{c\mu+c^2\sigma^2/2}, \tag{2.33}$$

where $X \sim \mathcal{N}(0,1)$ [53], but it is not immediately clear how this generalizes for $\mu, \sigma, c \in \mathbb{C}$. To determine that generalization, we observe that the log-normal PDF is [53]

$$f\left(e^{\mu+\sigma x}\right) = \frac{1}{e^x \sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

Hence,

$$\begin{aligned}\mathrm{E}\left(e^{(\mu+\sigma X)c}\right) &= \int_{-\infty}^{\infty} e^{(\mu+\sigma x)c} f\left(e^{\mu+\sigma x}\right) \mathrm{d}x \\ &= \frac{1}{2} e^{c\mu+c^2\sigma^2/2} \int_{x=-\infty}^{\infty} \mathrm{d}\left(\mathrm{erfc}\left(\frac{\mu+c\sigma^2-x}{\sqrt{2}\sigma}\right)\right)\end{aligned} \tag{2.34}$$

If we can show that the integral converges to 2 then we will have shown that (2.33) holds in the complex-valued case.

To pursue that line of thought, we observe that the complex generalization of the

complementary error function $\mathrm{erfc}(\cdot)$ is commonly given via the Faddeeva function [85], which is defined on $z \in \mathbb{C}$ with $\mathrm{Im}(z) > 0$ as

$$w(z) = \frac{j}{\pi} \int_{-\infty}^{\infty} \frac{e^{-t^2}}{z - t} \mathrm{d}t.$$

It relates to $\mathrm{erfc}(\cdot)$ as [85]

$$w(z) = e^{-z^2} \mathrm{erfc}(-jz),$$

and thus, for $z \in \mathbb{C}$ with $\mathrm{Re}(z) > 0$, we can define

$$\mathrm{erfc}(z) = e^{-z^2} w(jz).$$

In (2.34), we are interested in finding $\lim_{\mathrm{Re}(z)\to\infty} \mathrm{erfc}(z)$. Using the Faddeeva function, we have that

$$\lim_{\mathrm{Re}(z)\to\infty} \mathrm{erfc}(z) = \lim_{\mathrm{Re}(z)\to\infty} e^{-z^2} w(jz) = 0,$$

which is evident since both e^{-z^2} and $w(jz)$ tend to 0 in the limit. Since $\mathrm{erfc}(-z) = 2 - \mathrm{erfc}(z)$, we obtain

$$\lim_{\mathrm{Re}(z)\to-\infty} \mathrm{erfc}(z) = 2.$$

Thus, the integral in (2.34) indeed converges to 2, and (2.33) is valid for $c, \mu, \sigma \in \mathbb{C}$.

Referring to our noise-free signal with sampling jitter $x_m[n] = \sum_{k=1}^{K} r_k e^{s_k(n+\varsigma\gamma_m[n])}$, using the above result, we obtain

$$\mathrm{E}(x[n]) = \sum_{k=1}^{K} r_k e^{s_k n + s_k^2 \varsigma^2/2},$$

by term-wise application of (2.33) with $c = 1$. The cross-correlations are

$$\mathrm{E}\left(x_m\left[n\right]x_m\left[n'\right]^*\right) = \begin{cases} \mathrm{E}\left(\left|x_m\left[n\right]\right|^2\right), & n = n' \\ \mathrm{E}\left(x_m\left[n\right]\right)\mathrm{E}\left(x_m\left[n'\right]^*\right), & n \neq n'. \end{cases}$$

In the $n = n'$ case,

$$\left|x_m\left[n\right]\right|^2 = \sum_{k,k'=1}^{K} r_k r_{k'}^* e^{\left(s_k + s_{k'}^*\right)\left(n + \varsigma\gamma_m[n]\right)}$$

thus

$$\mathrm{E}\left(x_m\left[n\right]x_m\left[n'\right]^*\right) = \sum_{k,k'=1}^{K} r_k r_{k'}^* e^{\left(s_k + s_{k'}^*\right)n + \varsigma^2\left(s_k + s_{k'}^*\right)^2/2}.$$

In the $n \neq n'$ case,

$$\mathrm{E}\left(x_m\left[n\right]\right)\mathrm{E}\left(x_m\left[n'\right]^*\right) = \sum_{k,k'=1}^{K} \left(r_k e^{s_k n + \sigma^2 s_k^2/2}\right)\left(r_{k'}^* e^{s_k^* n' + \varsigma^2\left(s_{k'}^*\right)^2/2}\right),$$

thus

$$\mathrm{E}\left(x_m\left[n\right]x_m\left[n'\right]^*\right) = \sum_{k,k'=1}^{K} r_k r_{k'}^* e^{s_k n + s_{k'}^* n' + \varsigma^2\left(s_k^2 + \left(s_{k'}^*\right)^2\right)/2}.$$

These results are summarized in (2.30) and (2.31).

A noise-free signal with synchronization error is modeled as $x_m\left[n\right] = \sum_{k=1}^{K} r_k e^{s_k(n + \nu\varphi_m)}$. The analysis for $\mathrm{E}\left(x\left[n\right]\right)$ is thus unchanged from the sampling jitter case, and the cross-correlations are

$$\mathrm{E}\left(x_m\left[n\right]x_m\left[n'\right]^*\right) = \sum_{k,k'=1}^{K} r_k r_{k'}^* e^{\left(s_k + s_{k'}^*\right)n + \nu^2\left(s_k + s_{k'}^*\right)^2/2}.$$

2.C Generating Multiple Shots

For applications in which multiple physical measurement shots are not readily available, Barone [8] has noted that additional white noise can be added synthetically to a single measurement to create multiple virtual measurement shots. This comes at the cost of further lowering the SNR of the LTR, which, as noted in the Introduction, is already quite low. Hargrave *et al.* [41] have noted that it is also possible to use different windows on a single LTR measurement to generate multiple virtual shots. This comes at the cost of reduced signal length. As discussed in the Introduction, ensuring a sufficiently long LTR is already a challenge.

Acknowledgment

The authors would like to acknowledge the significant contributions made to this work by Professor Jian-Kang Zhang of McMaster University prior to his passing.

Supplementary Material

2.D Multi-Shot Maximum Likelihood Estimator

Let us consider a multi-shot scenario in which the m-th shot has a synchronization offset of $\tau_m T$ and the sampling jitter on the n-th sample of the m-th shot is $\eta_m[n]T$. (In the models in Section 2.7, $\tau_m = \nu\varphi_m$ and $\eta_m[n] = \varsigma\gamma_m[n]$.) In that setting, if we assume that the synchronization offset and sampling jitter do not impact the distribution of the noise, we can write

$$y_m[n] = \sum_{k=1}^{K} r_k e^{s_k(n+\tau_m+\eta_m[n])} + v_m[n].$$

Using the vectorized notation outlined in Section 2.2, we can rewrite this expression as

$$\mathbf{y}_m = \Big(\mathbf{F}(\mathbf{s}) \odot \big(\mathbf{H}(\mathbf{s};\boldsymbol{\eta}_m)\mathbf{D}(\mathbf{s};\tau_m)\big)\Big)\mathbf{r} + \mathbf{v}_m, \tag{2.35}$$

where where $\mathbf{F}(\mathbf{s})$ is defined in (2.5), $\odot$ denotes the Hadamard (elementwise) product, $[\boldsymbol{\eta}_m]_n = \eta_m[n]$, $[\mathbf{H}(\mathbf{s};\boldsymbol{\eta}_m)]_{nk} = e^{s_k\eta_m[n]}$, and $\mathbf{D}(\mathbf{s};\tau_m) = \operatorname{diag}\big(e^{s_1\tau_m}, e^{s_2\tau_m}, \ldots, e^{s_K\tau_m}\big)$. Let us define $\boldsymbol{\xi}_m = [\boldsymbol{\eta}_m^{\mathrm{T}}, \tau_m]^{\mathrm{T}} \in \mathbb{R}^Q$, where, in the case of synchronization offsets alone, $Q = 1$, and in the case of sampling jitter alone, $Q = N$. Let us also construct $\mathbf{E}(\mathbf{s};\boldsymbol{\xi}_m)$ such that $\mathbf{E}(\mathbf{s};[\boldsymbol{\eta}_m^{\mathrm{T}}, \tau_m]^{\mathrm{T}}) = \mathbf{H}(\mathbf{s};\boldsymbol{\eta}_m)\mathbf{D}(\mathbf{s};\tau_m)$. We can then write (2.35) in the form

$$\mathbf{y}_m = \breve{\mathbf{F}}(\mathbf{s};\boldsymbol{\xi}_m)\mathbf{r} + \mathbf{v}_m, \tag{2.36}$$

in which the residues $\mathbf{r}$ are the same in each shot and $\breve{\mathbf{F}}(\mathbf{s};\boldsymbol{\xi}_m) = \mathbf{F}(\mathbf{s}) \odot \mathbf{E}(\mathbf{s};\xi_m)$.

Under the model in (2.36), if the parameters $\boldsymbol{\xi}_m$ are modeled as non-random

parameters and are concatenated into the vector $\tilde{\boldsymbol{\xi}} = \begin{bmatrix} \boldsymbol{\xi}_1^{\mathrm{T}} & \boldsymbol{\xi}_2^{\mathrm{T}} & \cdots & \boldsymbol{\xi}_M^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}$, the maximum likelihood estimator can be written as

$$\left\{\hat{\mathbf{s}}_{\mathrm{ML}}, \hat{\mathbf{r}}_{\mathrm{ML}}, \hat{\tilde{\boldsymbol{\xi}}}_{\mathrm{ML}}\right\} = \arg\max_{\mathbf{s},\mathbf{r},\tilde{\boldsymbol{\xi}}} \sum_{m=1}^{M} \left\|\mathbf{y}_m - \breve{\mathbf{F}}(\mathbf{s};\boldsymbol{\xi}_m)\mathbf{r}\right\|_{\mathbf{C}^{-1}}^2.$$

If we define $\breve{\mathbf{A}}(\mathbf{s},\tilde{\boldsymbol{\xi}}) = \sum_{m=1}^{M} \breve{\mathbf{F}}(\mathbf{s};\boldsymbol{\xi}_m)^H \mathbf{C}^{-1} \breve{\mathbf{F}}(\mathbf{s};\boldsymbol{\xi}_m)$ and $\breve{\mathbf{B}}(\mathbf{s},\boldsymbol{\xi}_m) = \breve{\mathbf{F}}(\mathbf{s};\boldsymbol{\xi}_m)^H \mathbf{C}^{-1}$, then given ML estimates for $\mathbf{s}$ and $\tilde{\boldsymbol{\xi}}$, the ML estimate of $\mathbf{r}$ is

$$\hat{\mathbf{r}}_{\mathrm{ML}}(\hat{\mathbf{s}}_{\mathrm{ML}}, \hat{\tilde{\boldsymbol{\xi}}}_{\mathrm{ML}}) = \breve{\mathbf{A}}(\hat{\mathbf{s}}_{\mathrm{ML}}, \hat{\tilde{\boldsymbol{\xi}}}_{\mathrm{ML}})^{-1} \sum_{m=1}^{M} \breve{\mathbf{B}}(\hat{\mathbf{s}}_{\mathrm{ML}}, \hat{\boldsymbol{\xi}}_{m,\mathrm{ML}}) \mathbf{y}_m,$$

and if we define $\breve{\mathbf{P}}\left(\tilde{\mathbf{y}};\mathbf{s},\tilde{\boldsymbol{\xi}}\right) = \sum_{m=1}^{M} \mathbf{y}_m \breve{\mathbf{B}}\left(\mathbf{s},\boldsymbol{\xi}_m\right)$, the ML estimator can be concentrated to

$$\left\{\hat{\mathbf{s}}_{\mathrm{ML}}, \hat{\tilde{\boldsymbol{\xi}}}_{\mathrm{ML}}\right\} = \arg\max_{\mathbf{s},\tilde{\boldsymbol{\xi}}} \breve{\mathbf{P}}\left(\tilde{\mathbf{y}};\mathbf{s},\tilde{\boldsymbol{\xi}}\right)^{\mathrm{H}} \breve{\mathbf{A}}\left(\mathbf{s},\tilde{\boldsymbol{\xi}}\right)^{-1} \breve{\mathbf{P}}\left(\tilde{\mathbf{y}};\mathbf{s},\tilde{\boldsymbol{\xi}}\right). \tag{2.37}$$

When $\tilde{\boldsymbol{\xi}} = \mathbf{0}$, there are no synchronization errors, nor any sampling jitter, and the objective in (2.37) collapses to that in (2.18). Due to the presence of $\tilde{\boldsymbol{\xi}} \in \mathbb{R}^{MQ}$, and the way that it enters the objective, the problem in (2.37) is significantly more difficult to solve than those in (2.17) and (2.18).

An alternative approach to estimating $\mathbf{s}$ using a model of the form in (2.36) is to model the vector $\tilde{\boldsymbol{\xi}}$ as a random parameter with a known distribution. Under that model, the "stochastic" ML estimator can be expressed as the joint estimation problem

$$\left\{\hat{\mathbf{s}}_{\mathrm{MLR}}, \hat{\mathbf{r}}_{\mathrm{MLR}}\right\} = \arg\max_{\mathbf{s},\mathbf{r}} \bar{L}(\tilde{\mathbf{y}}|\mathbf{s},\mathbf{r}), \tag{2.38}$$

where $\bar{L}(\tilde{\mathbf{y}}|\mathbf{s},\mathbf{r})$ denotes the likelihood marginalized over $\tilde{\boldsymbol{\xi}}$; i.e.,

$$\bar{L}(\tilde{\mathbf{y}}|\mathbf{s},\mathbf{r}) = \int f\big(\tilde{\mathbf{y}}|\mathbf{s},\mathbf{r},\tilde{\boldsymbol{\xi}}\big) f\big(\tilde{\boldsymbol{\xi}}|\mathbf{s},\mathbf{r}\big) \mathrm{d}\tilde{\boldsymbol{\xi}},$$

in which $f\big(\tilde{\boldsymbol{\xi}}|\mathbf{s},\mathbf{r}\big)$ is the conditional distribution of the random parameters. In our case, this distribution is not dependent on $\mathbf{s}$ nor $\mathbf{r}$, and hence we will write $f\big(\tilde{\boldsymbol{\xi}}|\mathbf{s},\mathbf{r}\big) = f(\tilde{\boldsymbol{\xi}})$. The distribution $f\big(\tilde{\mathbf{y}}|\mathbf{s},\mathbf{r},\tilde{\boldsymbol{\xi}}\big) = \mathcal{CN}(\mathbf{g}, \mathbf{I}_M \otimes \mathbf{C})$, where

$$\mathbf{g} = \begin{bmatrix} \breve{\mathbf{Z}}\left(\mathbf{s};\boldsymbol{\xi}_1\right) \\ \breve{\mathbf{Z}}\left(\mathbf{s};\boldsymbol{\xi}_2\right) \\ \vdots \\ \breve{\mathbf{Z}}\left(\mathbf{s};\boldsymbol{\xi}_M\right) \end{bmatrix} \mathbf{r}.$$

In fact, since the noise is modeled as being independent between shots, $f\big(\tilde{\mathbf{y}}|\mathbf{s},\mathbf{r},\tilde{\boldsymbol{\xi}}\big)$ factorizes to $\prod_{m=1}^{M} f(\mathbf{y}_m|\mathbf{s},\mathbf{r},\boldsymbol{\xi}_m) = \prod_{m=1}^{M} \mathcal{CN}(\breve{\mathbf{Z}}(\mathbf{s};\boldsymbol{\xi}_m)\mathbf{r}, \mathbf{C})$. If each $\boldsymbol{\xi}_m$ is independent from all the others, then $f(\tilde{\boldsymbol{\xi}}) = \prod_{m=1}^{M} f(\boldsymbol{\xi}_m)$, and we can write

$$\bar{L}(\tilde{\mathbf{y}}|\mathbf{s},\mathbf{r}) = \prod_{m=1}^{M} \int f(\mathbf{y}_m|\mathbf{s},\mathbf{r},\boldsymbol{\xi}_m) f(\boldsymbol{\xi}_m)\, \mathrm{d}\boldsymbol{\xi}_m. \tag{2.39}$$

The expression for the marginalized likelihood in (2.38) involves a multidimensional integral over MQ dimensions, and hence simply evaluating the marginalized likelihood for one choice for the pair $(\mathbf{s},\mathbf{r})$ becomes rather unwieldy, let alone optimizing over those parameters as shown in (2.39) . If we can model the vectors $\boldsymbol{\xi}_m$ as being independent from each other, then the evaluation of $\bar{L}(\tilde{\mathbf{y}}|\mathbf{s},\mathbf{r})$ for a given pair $(\mathbf{s},\mathbf{r})$ can be reduced to the computation of M integrals of dimension Q. However, that evaluation remains unwieldy, and the problem in (2.38) remains difficult to solve.

Chapter 3

Multi-Shot Estimation of Resonance Parameters of Late-Time Radar Returns in Clutter

Abstract

The resonance parameters of late-time returns (LTRs) can be used as features in the identification of radar targets. However, reliable estimation of the complex frequency of each resonance is notoriously difficult. This is due to the short duration of the LTR, its low effective signal-to-noise ratio (SNR) and the inherent sensitivity of the estimation problem. These issues are exacerbated when the radar background includes resonating clutter. We develop an effective technique for estimation the complex frequencies of a target's resonances for scenarios in which the radar can obtain multiple

measurement shots of the background (clutter) alone and multiple measurement shots of the target in the presence of the background. The proposed method exploits the fact that the maximum likelihood estimator for measurements in Gaussian noise can be decomposed to estimate the complex frequencies of the resonances separately from their complex amplitudes. This enables us to decouple the estimation of the complex frequencies of the target from those of the background because the the background's complex frequencies remain largely unchanged when the target is introduced. We investigate the performance of the proposed method using a radar that operates in the band of 0.5 GHz to 5 GHz and employs equivalent sampling at a rate of 20 GSa/s. Proof-of-concept experiments on brass rods of known length validate the overall approach, and experiments on more complex targets in clutter demonstrate its potential for practical applications.

3.1 Introduction

Time-domain impulse radar systems seek to locate and identify targets by measuring and processing their early-time responses (ETRs) and late-time responses (LTRs) to the transmitted pulse [16], [30]. In the idealized case of a plane-wave stimulus, a target reflects the wave directly, which is its ETR, which is followed by a largely resonant response, its LTR [46], [80]. In an acoustic analogy, a guitar string may be plucked, which creates a loud twang, its ETR, which is then followed by ringing, its LTR. It is often possible to identify a guitar as such based on this ringing. Similarly, it is often possible to identify a radar target based on the resonant features of its LTR [12], [42]. Although it is notoriously difficult to estimate the parameters of these resonant features [3], this approach remains a compelling option in applications in

which radar-imaging techniques are not appropriate; e.g., due to privacy concerns.

Once the onset and duration of the LTR have been identified, using techniques such as those in [42], the goal is to estimate the number of resonating components, and the frequency and attenuation rate of each component. In some scenarios, it is also of interest to estimate the amplitude and phase of each component. In others, these are "nuisance" parameters, which are unknown and must be estimated, at least implicitly, but are not of interest. In a conventional setup, the parameters of interest are estimated from a single measurement shot. If the number of components is known, this can be done by using linear prediction techniques [56]; (non-linear) least squares techniques [84] derived from the maximum likelihood estimator in Gaussian noise [20]; or methods that rely on the algebraic structure of the resonating signal in the absence of noise, such as Prony's method [24], Padé approximants [39], and matrix-pencil techniques [49], [50], [74]. Unfortunately, the observation window of the LTR tends to be quite short in practice, and its effective signal-to-noise ratio (SNR) tends to be quite low. Worse still, the estimation problems themselves can be rather sensitive to noise and other perturbations [34], [37], [51].

In many potential applications, the expected motion of the target is slow enough that it can be considered to be stationary over the duration of multiple measurement shots. If those measurement shots are processed jointly, then the limitations imposed by the duration of the observation window and the low SNR can be mitigated to some degree [34]. However, that approach does not address the issue of environmental clutter, which we will call the *background*. We consider tackling this problem in scenarios where it is possible to capture multiple measurement shots of the environment before and after the introduction of the *target (TG)* whose resonance parameters are to be

estimated. We thus have a set of *background-only (BG) measurements* and a set of *target-and-background (TG&BG) measurements* available for analysis. We posit that combined processing of both measurement sets should enable us to improve the estimation of the resonant parameters of the target.

In this chapter, we develop such a method and demonstrate its performance using an ultra-wideband time-domain radar to estimate the resonance parameters of several targets in the presence of "light" and "heavy" background clutter. We show that when we use simple rods as targets, our results agree with the resonance parameters obtained by simulation, and that when we use common handheld weapons as targets, we arrive at consistent estimates, even when the clutter partially blocks the target.

The proposed method is based on the observation that the introduction of the target can, and often will, change the amplitude and phase of a clutter resonance. For instance, the target may partially obscure a given object in the background, reducing the associated amplitude. The amplitude and phase of a clutter object's resonance may thus be significantly different in the BG and TG&BG measurement sets, and this difference may be rather difficult to predict. However, the resonant frequencies and attenuation rates are features of an object; they are unaffected by the environment, provided that the object is "separate enough" from that environment (i.e., at a minimum, not in direct contact with conductive background clutter). Therefore, the background's resonant frequencies and attenuation rates should be consistent between the BG and TG&BG measurement sets. Provided that we can decouple the estimation of the frequencies and attenuation rates from that of the amplitude and phase, this would enable to make effective use of the BG measurements in the estimation of the TG resonances from the TG&BG measurements.

The (nonlinear) least-squares estimator that is derived from maximum likelihood estimator in Gaussian noise [20] indeed decouples the estimation in the desired manner [84]. However, it requires the solution of a non-convex optimization problem that has a smooth, but complicated, optimization landscape. We develop a pragmatic approach to generating good estimates from that problem. Our approach involves local smooth optimization from a carefully selected starting point, the application of physically motivated bounds on the optimization, and insight-driven post-processing of the optimization output. Our method for selecting the starting point is inspired by observations by Barone [8] that estimates of the resonant frequencies and attenuation rates from multiple measurement shots tend to cluster, even though the estimates from any one shot may be significantly perturbed. We show that the number of significant clusters is a good estimate of the number of resonating components, and that a good starting point for the nonlinear least-squares optimization can be obtained by finding the "center" of each significant cluster.

After describing the signal model in Section 3.2 and the estimation problem in Section 3.3, we outline our pragmatic approach to solving this problem in the case of multiple measurement shots in Section 3.4. The main contribution of the chapter is outlined in Section 3.5, wherein we describe the proposed approach for estimating the resonance parameters of the target given a set of BG measurements and a set of TG&BG measurements. In Section 3.6, we apply this technique to measurements from a physical radar system. The proposed approach provides significantly better estimates of the resonance parameters than a baseline technique that was constructed from existing ideas in the literature.

3.2 System Model and Problem Statement

We assume that the LTR is composed of damped sinusoids and is uniformly sampled at N instants $n = 0, \ldots, N-1$, leading to the discrete-time signal model

$$x_\bullet[n] = \sum_{k=1}^{K_\bullet} r_{\bullet k} e^{s_{\bullet k} n}, \tag{3.40}$$

where the $\bullet$ indicates whether the signal is from the target (TG), background clutter (BG), or from the target embedded in background clutter (TG&BG). The terms $r_{\bullet k}$ are the *complex amplitudes* or *residues* of the resonances, and the terms $s_{\bullet k} = -\alpha_{\bullet k} + j\omega_{\bullet k}$ are the *complex frequencies* of the resonances which contain their *attenuation rates* $\alpha_{\bullet k}$ and their *frequencies* $\omega_{\bullet k}$. We will often find it convenient to refer to the z-domain *poles* $z_{\bullet k} = e^{s_{\bullet k}}$, and to write $x_\bullet[n] = \sum_{k=1}^{K_\bullet} r_{\bullet k} z_{\bullet k}^n$.

In the scenario that we will consider, we seek to estimate the complex frequencies of the target(s) $\{s_{\mathrm{TG}k}\}_{k=1}^{K_{\mathrm{TG}}}$ from $M_{\mathrm{TG\&BG}}$ measurement shots of the target embedded in the background, assisted by M_{BG} measurement shots of the background clutter (in the absence of any targets). Each measurement shot consists of N samples taken in the presence of zero-mean additive noise $v_{\bullet m}[n]$ and takes the form

$$y_{\bullet m}[n] = x_\bullet[n] + v_{\bullet m}[n], \tag{3.41}$$

where $m = 1, 2, \ldots, M_\bullet$, and $\bullet$ may be BG or TG&BG. (Of course, we cannot measure the target TG directly.) We will develop our estimator for a time-domain pulsed radar, and hence $x_\bullet[n]$ and $y_{\bullet m}[n]$ are real. Thus, $K_\bullet$ is inherently even. We will index the terms in (3.40) so that $r_{\bullet(k+K/2)} = r_{\bullet k}^*$ and $s_{\bullet(k+K/2)} = s_{\bullet k}^*$. Therefore, we only need

to estimate $s_{\mathrm{TG}k}$ for $k = 1, \ldots, K_{\mathrm{TG}}/2$.

We will often find it convenient to work with the matrix-vector form of our model in which the expression in (3.40) is rewritten as

$$\mathbf{x}_{\bullet} = \mathrm{Vand}_N\left(\mathbf{z}_{\bullet}\right)\mathbf{r}_{\bullet}, \tag{3.42}$$

where $[\mathbf{x}_{\bullet}]_n = x_{\bullet}[n]$, $[\mathbf{z}_{\bullet}]_k = z_{\bullet k} = e^{s_{\bullet k}}$, and $\mathrm{Vand}_N(\mathbf{z})$ is an $N \times K_{\bullet}$ Vandermonde matrix with $[\mathrm{Vand}_N(\mathbf{z}_{\bullet})]_{nk} = z_{\bullet k}^n$, where $n = 0, \ldots, N-1$. Similarly, the measurement shots in (3.41) can be rewritten as

$$\mathbf{y}_{\bullet m} = \mathbf{x}_{\bullet} + \mathbf{v}_{\bullet m}, \tag{3.43}$$

where $[\mathbf{y}_{\bullet m}]_n = y_{\bullet m}[n]$ and $[\mathbf{v}_{\bullet m}]_n = v_{\bullet m}[n]$.

Let us also define the *effective signal-to-noise ratio (SNR)* as the average SNR over the N samples, which is equivalent to the ratio of the signal energy to the noise energy over the observation window, i.e.,

$$\mathrm{SNR}_m = \frac{\mathbf{x}_{\bullet}^{\mathrm{T}}\mathbf{x}_{\bullet}}{\mathbf{v}_{\bullet m}^{\mathrm{T}}\mathbf{v}_{\bullet m}}. \tag{3.44}$$

3.3 Least-Squares Pole Estimation

The target-resonance estimation techniques that we will develop in this chapter are based on a nonlinear least-squares estimation method derived from the maximum likelihood (ML) estimator in Gaussian noise; e.g., [20], [84]. Let us first consider the ML estimation problem for a single measurement shot. We observe from the model in

(3.42) and (3.43) that if the measurement noise vector $\mathbf{v}_{\bullet m}$ can be modeled as being a zero-mean random Gaussian vector with covariance matrix $\mathbf{C}_\bullet$ and the number of resonant terms $K_\bullet$ is known, then maximum likelihood estimation of the poles and residues from the m-th measurement shot corresponds to solving the following single-shot (nonlinear) least squares (SSLS) problem:

$$\hat{\mathbf{z}}_{\bullet m\mathrm{SSLS}}, \hat{\mathbf{r}}_{\bullet m\mathrm{SSLS}} = \arg\min_{\mathbf{z},\mathbf{r}} \|\mathbf{y}_{\bullet m} - \mathrm{Vand}_N(\mathbf{z})\,\mathbf{r}\|^2_{\mathbf{C}_\bullet^{-1}}. \tag{3.45}$$

We will refer to (3.45) as the *(nonlinear) least squares estimator* because it can be applied even if the noise is not Gaussian; it is just not the ML estimator in that case. An advantage of the SSLS estimator is that we can decouple the estimation of the poles from the estimation of the residues. In particular, the SSLS estimator for the poles is [84]

$$\hat{\mathbf{z}}_{\bullet m\mathrm{SSLS}} = \arg\max_{\mathbf{z}} \mathbf{y}^{\mathrm{H}}_{\bullet m}\mathbf{Q}_\bullet(z)\,\mathbf{y}_{\bullet m}, \tag{3.46}$$

where $\mathbf{Q}_\bullet(\mathbf{z}) = \mathbf{B}_\bullet(\mathbf{z})^{\mathrm{H}}\,\mathbf{A}_\bullet(\mathbf{z})^{-1}\,\mathbf{B}_\bullet(\mathbf{z})$ with $\mathbf{A}_\bullet(\mathbf{z}) = \mathrm{Vand}_N(\mathbf{z})^{\mathrm{H}}\,\mathbf{C}_\bullet^{-1}\mathrm{Vand}_N(\mathbf{z})$ and $\mathbf{B}_\bullet(\mathbf{z}) = \mathrm{Vand}_N(\mathbf{z})^{\mathrm{H}}\,\mathbf{C}_\bullet^{-1}$. Once an estimate $\hat{\mathbf{z}}_{\bullet m\mathrm{SSML}}$ has been found, if estimates of the residues are needed, they can be determined by solving the linear least-squares problem

$$\hat{\mathbf{r}}_{\bullet m\mathrm{SSLS}} = \arg\min_{\mathbf{r}} \|\mathbf{y}_{\bullet m} - \mathrm{Vand}_N(\hat{\mathbf{z}}_{\bullet m\mathrm{SSLS}})\,\mathbf{r}\|^2_{\mathbf{C}_\bullet^{-1}}. \tag{3.47}$$

In our envisioned application, multiple measurement shots $\{\mathbf{y}_{\bullet m}\}_{m=1}^{M}$ are available, and the poles and residues can be estimated by processing the measurement shots jointly. In particular, the multi-shot least squares (MSLS) estimates of the poles are

(e.g., [34])

$$\hat{\mathbf{z}}_{\bullet\mathrm{MSLS}} = \arg\max_{\mathbf{z}} \sum_{m=1}^{M} \mathbf{y}_{\bullet m}^{\mathrm{H}} \mathbf{Q}_{\bullet}(\mathbf{z})\, \mathbf{y}_{\bullet m}. \tag{3.48}$$

The estimator in (3.48) does not rely on any assumption on the relationship between the residues in different measurement shots. In that sense, it can be viewed as being a universal multi-shot estimator of the poles.[1] Once the pole estimates in (3.48) have been found, we can find residues that are a "best fit" for those estimates:

$$\hat{\mathbf{r}}_{\bullet\mathrm{MSLS}} = \arg\min_{\mathbf{r}} \sum_{m=1}^{M} \|\mathbf{y}_{\bullet m} - \mathrm{Vand}_N(\hat{\mathbf{z}}_{\bullet\mathrm{MSLS}})\, \mathbf{r}\|^2_{\mathbf{C}_{\bullet}^{-1}}. \tag{3.49}$$

These best-fit residues can be useful in making inferences regarding the quality of the pole estimates; see the discussion in the following section.

3.4 Pragmatic Approach to Solving the Pole-Estimation Problem

A significant advantage of the multi-shot least squares estimation technique in the envisioned application is that it separates the estimation of the poles, which we are interested in, from the estimation of the residues, which are of lesser interest; cf. (3.48). However, this estimation technique requires the solution of a non-convex optimization problem, (3.48), that has an optimization landscape that is smooth, but complicated. As a result, methods that guarantee a globally optimal solution tend to be too computationally expensive to be viable in practical applications. In this

[1] An alternative approach would be to assume that the residues in each measurement are the same. The resulting estimation problem for the poles corresponds to applying the SSLS estimator in (3.46) to the averaged received signal $\bar{\mathbf{y}}_{\bullet} = \frac{1}{M}\sum_{m=1}^{M} \mathbf{y}_{\bullet m}$; see, e.g., [34].

section, we will outline a pragmatic approach that generates good solutions in a wide variety of scenarios. Our approach employs a smooth local optimization technique, namely the BFGS quasi-Newton method [64, Ch. 6], initialized by a carefully chosen starting point. We also use our knowledge of the physical capabilities of the radar and of the physical properties of the reflective targets to constrain the search space for the pole estimates. Our method for selecting the starting point of the optimization also provides an initial estimate of the number of resonant sources. The final step is a method to mitigate the impact of errors in that estimate.

3.4.1 Generating an Initial Solution

The first step in this phase is to apply the generalized pencil-of-functions (GPoF) method [49] to each measurement shot, in order to generate M independent estimates of $N/2$ poles, $\{\hat{z}_{k\bullet m\mathrm{GPoF}}\}_{k=1}^{N/2}$, $m = 1, 2, \ldots, M_\bullet$; see Step 1 in Algorithm 2. As detailed in Appendix 3.A, the GPoF estimates are obtained by solving generalized eigenvalue problems that involve matrices assembled from the received signal. In any given measurement shot, the additive noise can significantly perturb the GPoF estimates of the poles, but the estimates will tend to cluster over the $M_\bullet$ measurement shots [8]. To exploit that clustering, we use the $MN/2$ GPoF pole estimates to construct a (smooth) kernel density estimate (KDE) of the distribution of the pole estimates in the presence of noise. This step is described in more detail in Appendix 3.B; see Step 2 in Algorithm 2. The number of poles to be estimated using the MSLS method in (3.48) is chosen to be of the number of significant maxima of the KDE, $\hat{K}_{\bullet\mathrm{KDE}}$, and the locations of the maxima, $\{\hat{z}_{k\bullet\mathrm{KDE}}\}_{k=1}^{\hat{K}_{\bullet\mathrm{KDE}}}$, are used as the starting point for the BFGS-based local optimization. Thus, the number of pole estimates that will

be generated by the BFGS search is $\hat{K}_{\bullet\text{BFGS}} = \hat{K}_{\bullet\text{KDE}}$. Given the nature of our estimation problem, it is preferable to err on the side of overestimating the number of poles to be estimated, and then to resolve any spurious estimates in the final step; see Section 3.4.3.

3.4.2 Restricting the Search Space

To improve the effectiveness of the BFGS solver for the MSLS estimation problem, we constrain the search space. We choose bounds on the frequencies such that the operating band falls well inside the chosen range, with sufficient tolerances to account for the aforementioned susceptibility to large perturbations in such estimation problems. We also place loose bounds on the attenuation rates. Since the length of the measurement shots, N, is often quite short, the attenuation estimates are especially prone to large perturbations. Hence, it can be prudent to allow for slightly negative estimates of the attenuation rates, so that the optimization algorithm can explore a wide variety of paths to good solutions. All the BFGS estimates of the poles, $\{\hat{z}_{k\bullet\text{BFGS}}\}_{k=1}^{\hat{K}_{\bullet\text{BFGS}}}$, will be constrained by these bounds. In our physical experiments described in Section 3.6, we will bound the frequency estimates to the range from 0.25 GHz to 5 GHz, and the estimates of the attenuation rates to the range from -0.3 ns^{-1} to 0.6 ns^{-1}.

3.4.3 Removal of Spurious Pole Estimates

Since our initial estimate of the number of background poles, $\hat{K}_{\bullet\text{BFGS}} = \hat{K}_{\bullet\text{KDE}}$, is chosen so that it errs on the side of overestimation, we must perform some post-processing on the BFGS estimate to remove any spurious pole estimates in out final

MSLS estimate. These spurious pole estimates can arise because the estimated poles $\{\hat{z}_{k\bullet\text{BFGS}}\}_{k=1}^{\hat{K}_{\bullet\text{BFGS}}}$ and residues $\{\hat{r}_{k\bullet\text{BFGS}}\}_{k=1}^{\hat{K}_{\bullet\text{BFGS}}}$ relate to the actual poles $\{z_{k\bullet}\}_{k=1}^{K_\bullet}$ and residues $\{r_{k\bullet}\}_{k=1}^{K_\bullet}$ as

$$\sum_{\ell=1}^{K_\bullet} r_{\ell\bullet} z_{\ell\bullet}^n \approx \sum_{k=1}^{\hat{K}_{\bullet\text{BFGS}}} \hat{r}_{k\bullet\text{BFGS}} \hat{z}_{k\bullet\text{BFGS}}^n.$$

As such, they represent an accurate but non-unique decomposition of the signal.

One case of spurious poles that may arise is that there may be two nearly identical pole estimates $\hat{z}_{j\bullet\text{BFGS}} \approx \hat{z}_{k\bullet\text{BFGS}}$, where $j \neq k$. Such pole estimates should be combined into a single pole estimate $\hat{z}'_{\ell\bullet\text{MSLS}} = (\hat{z}_{j\bullet\text{BFGS}} + \hat{z}_{k\bullet\text{BFGS}})/2$ with estimated residue $\hat{r}'_{\ell\bullet\text{MSLS}} = \hat{r}_{j\bullet\text{BFGS}} + \hat{r}_{k\bullet\text{BFGS}}$. This procedure should be repeated until all of the pole estimates are distinct. Let $\hat{K}'_{\bullet\text{MSLS}} \leq \hat{K}_{\bullet\text{BFGS}}$ denote the number of such estimates.

Once all of the pole estimates $\{\hat{z}'_{\ell\bullet\text{MSLS}}\}_{\ell=1}^{\hat{K}'_{\bullet\text{MSLS}}}$ are distinct, it is possible there may be a pole estimate $\hat{z}'_{\ell\bullet\text{MSLS}}$ with an associated residue $\hat{r}'_{\ell\bullet\text{MSLS}} \approx 0$. This is most likely to occur when the residue addition from the previous step is destructive, but it is also possible that the pole simply has a vanishing residue. Every such estimate should be removed, resulting in a final MSLS set of pole estimates $\{\hat{z}_{\ell\bullet\text{MSLS}}\}_{\ell=1}^{\hat{K}_{\ell\bullet\text{MSLS}}} \subset \{\hat{z}'_{\ell\bullet\text{MSLS}}\}_{\ell=1}^{\hat{K}'_{\ell\bullet\text{MSLS}}}$, where the final number of MSLS-estimated poles is $\hat{K}_{\ell\bullet\text{MSLS}} \leq \hat{K}'_{\ell\bullet\text{MSLS}}$.

3.4.4 Algorithm

In Algorithm 2, we have summarized our pragmatic approach to generating good solutions to the pole estimation problem in (3.48). To place our approach in context, we observe that, prior to its development, the prominent multiple-measurement-shot approaches that were available to estimate the poles independently from the residues

Algorithm 2 Pragmatic Approach to Pole Estimation

1. Perform GPoF on each measurement shot $\mathbf{y}_{\bullet m}$ to generate the pole estimates $\{\hat{z}_{k\bullet m\mathrm{GPoF}}\}_{k=1,m=1}^{N/2,M}$.
2. Generate a KDE from those estimates
3. Find the significant peaks of the KDE $\{\hat{z}_{k\bullet\mathrm{KDE}}\}_{k=1}^{\hat{K}_{\bullet\mathrm{KDE}}}$.
4. Find a local solution of (3.48), $\{\hat{z}_{k\bullet\mathrm{BFGS}}\}_{k=1}^{\hat{K}_{\bullet\mathrm{BFGS}}}$, by applying bounded BFGS starting from $\{\hat{z}_{k\bullet\mathrm{KDE}}\}_{k=1}^{\hat{K}_{\bullet\mathrm{KDE}}}$.
5. Remove any duplicate and spurious poles as described in Section 3.4.3 to yield the final estimate $\{\hat{z}_{\ell\bullet\mathrm{MSLS}}\}_{\ell=1}^{\hat{K}_{\bullet\mathrm{MSLS}}}$.

were based on the principles that underlie the clustering technique developed by Barone [8]. Our approach to obtaining the starting points for the local BFGS optimization can be viewed as a tailored application of the principles of Barone's clustering technique; see Steps 1–3 in Algorithm 2. However, as we will show in our experimental results in Section 3.6, Steps 4 and 5 in Algorithm 2 provide significant improvements in the quality of the pole estimates.

3.5 Proposed Approach to Target Pole Estimation

In our envisioned application, a target is introduced into an environment containing background clutter within the field of view of a radar. Therefore, we can acquire both a set of multiple measurement shots of the background clutter alone, and a set of multiple measurement shots of the target embedded therein. Acquiring a set of measurement shots is very quick, thus we assume that the target is static during the acquisition of its measurement set, and the background is static throughout the

acquisition of both sets. Our goal is to build upon the method described in the previous section to develop an approach for estimating the poles of the target(s) that effectively leverages the availability of the background measurements.

One intuitive approach would be to perform signal subtraction and then perform conventional estimation on the remaining signal. To do that, we would simply subtract an estimate of the background signal $\hat{\mathbf{x}}_{\mathrm{BG}}$ from each of the measurements of the target embedded in the background $\mathbf{y}_{\mathrm{TG\&BG}m}$, and then apply a pole estimator to the measurements $\{\mathbf{y}_{\mathrm{TG\&BG}m} - \hat{\mathbf{x}}_{\mathrm{BG}}\}_{m=1}^{M}$. While we will make use of a subtraction-based approach as a signal preconditioning step in our proposed approach, the subtraction-based approach has the inherent weakness that the presence of the target may change the values of some of the residues associated with the background resonances. For example, if the target blocks part of the background, then its corresponding residue may be smaller in the TG&BG measurements compared to the BG measurements. If the target provides an additional reflection path to a component, then its corresponding residue may be larger and have a different phase in the TG&BG measurements compared to the BG measurements. The combined impact of these effects is that although subtraction may suppress the influence of the background resonances, there may be significant residual components that should be directly addressed.

Our proposed approach embraces the presence of residual background components. It exploits the observation that while the presence of the target may change the residues associated with background resonances, the poles of these resonances are properties of the resonating object itself, and will not be significantly[2] affected by the presence of the target. This observation is used in conjunction with the fact that the (nonlinear) least-squares estimator (3.48) enables direct estimation of the

[2]Assuming the target is not in extreme proximity to the clutter.

poles without the need to estimate the residues. In particular, when this estimator is applied to the TG&BG measurement shots, we can treat the poles of the BG components as being known because they can be estimated from the background-only measurements. This enables us to distinguish the TG poles from the BG poles, and also reduce the dimension of the optimization problem that needs to be solved.

As outlined in the sections below and in Algorithm 3, our approach consists of three phases: a background estimation phase, a signal preconditioning phase, and a target estimation phase.

3.5.1 Background Estimation

In this phase, we estimate the poles of the background signal from the measurements of the background $\{\mathbf{y}_{\mathrm{BG}m}\}_{m=1}^{M_{\mathrm{BG}}}$. This problem can be viewed as a straightforward multi-shot estimation problem. As such, we will employ the pragmatic approach described in the previous section for generating good solutions to the MSLS problem in (3.48) with $\bullet = \mathrm{BG}$; see Step 1 of Algorithm 3.

3.5.2 Preconditioning of the Target Measurement Signal

As we noted above, subtracting an estimate of the background signal is insufficient to remove the influence of the clutter on a target-and-background signal $\mathbf{y}_{\mathrm{TG\&BG}m}$. However, the background also includes other components for which subtraction can provide substantial suppression. One of these is the radar antennas' self resonances. For example, the narrow transmitted pulse induces self resonances in the transmitting antenna which are picked up by the receiving antenna through direct coupling. The pulse itself is also picked up by direct coupling, which induces resonances in the

receiving antenna. Since they arise from direct coupling, such resonances are by far the strongest resonant components of the received signal, and hence they remain a significant component of the received signal over a long time window; see Figure 3.7 in Section 3.6 for an example from our system. This is problematic because a cluster of pole estimates with large residues can conceal a smaller cluster which is nearby; i.e., strong resonances can mask weaker ones. Since the directly coupled signal is not affected by the presence or absence of the target, we will precondition our target measurement shots $\{\mathbf{y}_{\mathrm{TG\&BG}m}\}_{m=1}^{M_{\mathrm{TG\&BG}}}$ by subtracting the mean of the background measurement shots,

$$\bar{\mathbf{y}}_{\mathrm{BG}} = \frac{1}{M}\sum_{m=1}^{M}\mathbf{y}_{\mathrm{BG}m}.$$

That is, we construct

$$\mathbf{y}'_{\mathrm{TG\&BG}m} = \mathbf{y}_{\mathrm{TG\&BG}m} - \bar{\mathbf{y}}_{\mathrm{BG}}. \tag{3.50}$$

This preconditioning step not only suppresses the direct coupling components, but also the resonant components of the clutter whose residues are not significantly affected by the presence of the target. It suppresses non-resonant components of the background, as well. However, it is unable to suppress the transmitting antenna's self-resonance signals that reflect off the target, the receiving antenna's self-resonances that are induced by the reflection of the transmitted pulse from the target, nor any of the background resonances that undergo a significant phase change due to the presence of the target. An example of the effect of this subtraction-based preconditioning in our experimental set up is illustrated in Figure 3.7.

3.5.3 Target Estimation

The key observation in our approach to estimating the poles of the target is that the MSLS approach of (3.48) enables the estimation of the poles independently of the residues. This is important because the residues are are a function of the target, the radar environment, and the preconditioning. In contrast, the poles are an inherent characteristic of the target itself. Therefore, in the preconditioned received signal in (3.50), we expect to find poles which correspond to the background resonances, $\hat{\mathbf{z}}_{\mathrm{BG}}$, and poles which correspond to the target resonances, $\hat{\mathbf{z}}_{\mathrm{TG}}$. The background poles have already been estimated in phase 1 of our method, which was described in Section 3.5.1. Using those estimates, we propose to estimate the poles of the target by applying a reduced-dimension variant of the MSLS estimator in (3.48) to the preconditioned measurements. In particular, given the estimates of the background poles $\hat{\mathbf{z}}_{\mathrm{BG}}$ and an (over)estimate of the number of target poles $\hat{K}_{\mathrm{TG}}$, the reduced-dimension estimator is

$$\hat{\mathbf{z}}_{\mathrm{TG}} = \arg\max_{\mathbf{z}} \sum_{m=1}^{M_{\mathrm{TG\&BG}}} \mathbf{y}'^{\mathrm{H}}_{\mathrm{TG\&BG}m} \mathbf{Q}\left(\left[\begin{smallmatrix}\mathbf{z}\\ \hat{\mathbf{z}}_{\mathrm{BG}}\end{smallmatrix}\right]\right) \mathbf{y}'_{\mathrm{TG\&BG}m}. \tag{3.51}$$

The "best-fit" residues corresponding to both the BG and TG poles can be estimated by applying (3.49), namely

$$\hat{\mathbf{r}}_{\mathrm{TG\&BG}} = \arg\min_{\mathbf{r}} \sum_{m=1}^{M_{\mathrm{TG\&BG}}} \|\mathbf{y}'_{\mathrm{TG\&BG}m} - \mathrm{Vand}_N\left(\left[\begin{smallmatrix}\hat{\mathbf{z}}_{\mathrm{TG}}\\ \hat{\mathbf{z}}_{\mathrm{BG}}\end{smallmatrix}\right]\right) \mathbf{r}\|^2_{\mathbf{C}_\bullet^{-1}}.$$

Even though it has reduced dimensions compared to the MSLS pole estimation problem in (3.48), the optimization landscape of the problem in (3.51) remains complicated. Therefore, we apply the natural variant of the pragmatic approach to generating a good solution to (3.48) that was described in Section 3.4. That natural

Algorithm 3 MSLS Method for Target Pole Estimation

1. Estimate the background poles $\{\hat{z}_{k\mathrm{BG,MSLS}}\}_{k=1}^{\hat{K}_{\mathrm{BG,MSLS}}}$ from $\{\mathbf{y}_{\mathrm{BG}m}\}_{m=1}^{M_{\mathrm{BG}}}$.
2. Generate $\{\mathbf{y}'_{\mathrm{TG\&BG}m}\}_{m=1}^{M_{\mathrm{TG\&BG}}}$ using (3.50).
3. Estimate the target poles $\{\hat{z}_{k\mathrm{TG},MSLS}\}_{k=1}^{\hat{K}_{\mathrm{TG,MSLS}}}$ from $\{\mathbf{y}'_{\mathrm{TG\&BG}m}\}_{m=1}^{M_{\mathrm{TG\&BG}}}$, using $\{\hat{\mathbf{z}}_{k\mathrm{BG,MSLS}}\}_{k=1}^{\hat{K}_{\mathrm{BG,MSLS}}}$ and the modified form of Algorithm 2 described in Section 3.5.3.

variant involves applying Steps 1–3 of Algorithm 2 to the preconditioned signals $\{\mathbf{y}'_{\mathrm{TG\&BG}m}\}_{m=1}^{M_{\mathrm{TG\&BG}}}$, then applying the bounded BFGS technique to obtain a local solution to (3.51), and finally removing any duplicate and spurious pole estimates.

3.5.4 Algorithm

The steps describing how Algorithm 2 is used to to estimate target poles in the presence of clutter are summarized in Algorithm 3.

3.5.5 A Baseline Method

As we observed in Section 3.4.4, the prominent existing multi-shot methods for estimating poles, independently of residues, are based on clustering techniques. In order to establish a baseline against which we can compare the performance of our approach, we will now describe how those clustering techniques could be adapted to the estimation problem that we consider herein.

The first step would be to perform the same signal preconditioning that we use in the proposed technique; see Section 3.5.2. That is, we perform Step 2 of Algorithm 3.

Then, assuming that the preconditioning step has provided sufficient suppression

of the background resonances, we would apply Steps 1–3 of Algorithm 2, and use the significant peaks of the KDE as the estimates of the target poles.

3.6 Results

We now apply the proposed multi-shot least squares (MSLS) method for target pole estimation in Algorithm 3 to estimate the poles of various targets from measurements taken from our physical test bed. As shown in Figure 3.6, our measurement system consists of quad-ridge horn antennas at the transmitter and receiver. The system operates between 0.5 GHz and 5 GHz, with reliable performance between 0.7 GHz and 3.5 GHz. The receiver employs equivalent-sampling techniques to achieve an effective sampling rate of 20 Gsamples/sec. That receiver is described in more detail in [67], and the transmitter is described in more detail in [66]. The structure behind the target in Figure 3.6 is a "quiet" chamber that ensures that we will have an LTR window that is free of multi-path reflections, aside from the multi-path reflection from the floor; see Figure 3.7. It also allows us to acquire relatively good "clean" measurements, which we will use to establish a "ground truth" for the experiments in which we will add clutter to the environment.

A typical received signal for the setup in Figure 3.6 is shown in Figure 3.7. The latter figure shows that in between the initial reflection or ETR, which is marked in red, and the subsequent multi-path reflection via the floor, which is marked in orange, there is a relatively short window of 2.4 ns, which is marked in blue, during which we can analyze the LTR. At our equivalent sampling rate of 20 Gsamples/sec, this corresponds to an observation window of $N = 48$ samples.

Figure 3.6: Measurement setup with a sample target, a facsimile of a Colt-1911 pistol. The transmitting quad-ridge horn antenna is visible on the left, and the receiving one on the right.

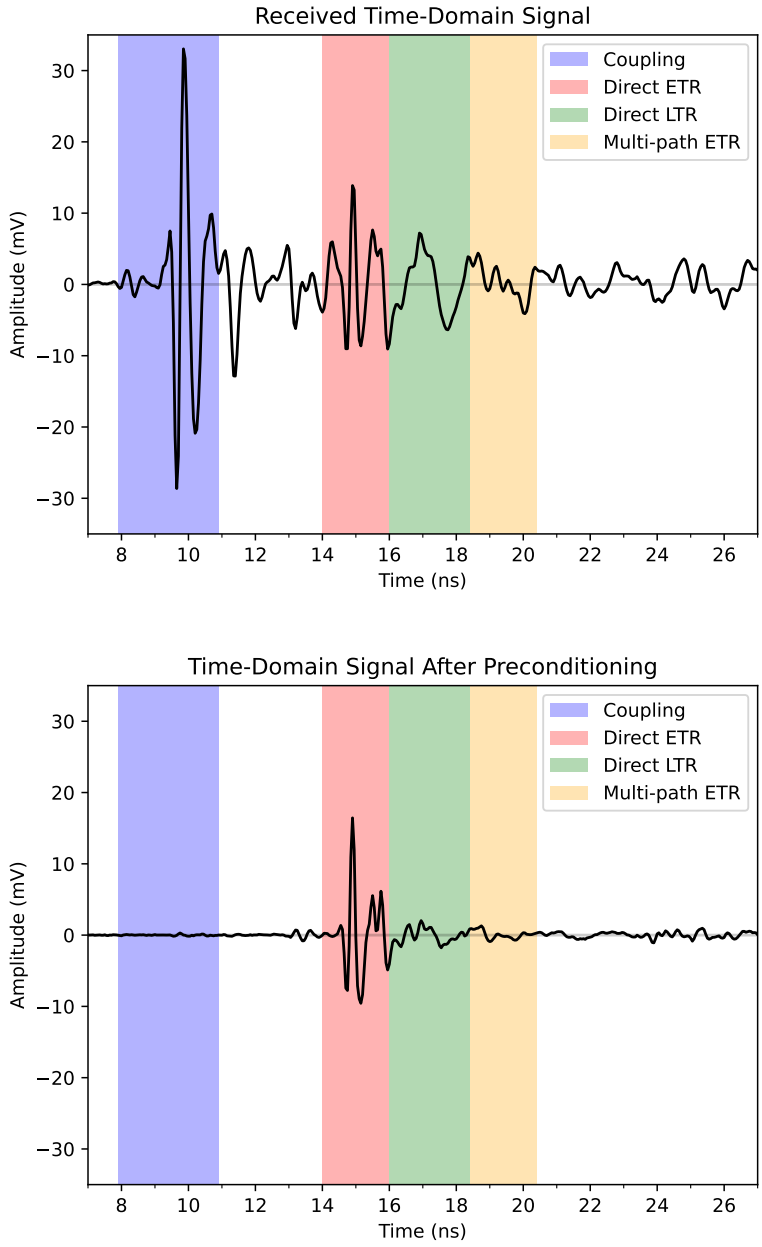

Figure 3.7: A sample measurement of the raw received signal, $\mathbf{y}_{\mathrm{TG\&BG}m}$, (top), and the corresponding preconditioned signal, $\mathbf{y}'_{\mathrm{TG\&BG}m}$ in (3.50), (bottom). The preconditioning removes almost all of the coupling signal (blue), and suppresses other strong signal components throughout. A weak but significant multi-path reflection (orange) curtails the usable LTR (green), which follows the ETR (red).

3.6.1 Proof-of-Concept Experiments

To validate the core claims of our approach, we tested our pole-estimation algorithm on two brass rods, one of length 15 cm and the other of length 10 cm. Both rods were 4.8 mm thick. In these experiments, we only use the horizontal polarization. Each rod was rotated in 22.5° increments, and the resonant pole was estimated using the proposed method. The BG measurement set contained $M_{\mathrm{BG}} = 100$ shots, and the TG&BG measurement sets contained $M_{\mathrm{TG\&BG}} = 100$ shots at each of the sixteen angles. The corresponding resonant frequency, which reflects the length of the rod, is plotted in Figure 3.8 for the 15-cm rod and in Figure 3.9 for the 10-cm rod. As a performance benchmark, both figures include the resonant frequency from a FEKO simulation [27]; see the green dashed line. Both figures also include a performance comparison with the baseline scheme described in Section 3.5.5.

Figure 3.8 shows that in the case of the 15-cm rod, both the proposed method and the baseline method provide consistently good estimates, except at angles where the rod is perpendicular to the polarization of the transmitted signal. The scenario in Figure 3.9 is more difficult for both methods because the shorter rod naturally has a smaller radar cross-section. For example, at the 0° orientation (at which the rod is aligned with the polarization) the effective SNR (cf. (3.44)) for the shorter rod is 12 dB whereas that for longer rod is 20 dB. However, it can be seen that the performance of the proposed approach (orange line) is significantly better than that of the baseline method (blue line). We carried out analogous experiments with rods of length 12.5 cm, 17.5 cm and 20 cm, and the results were consistent with those presented here.

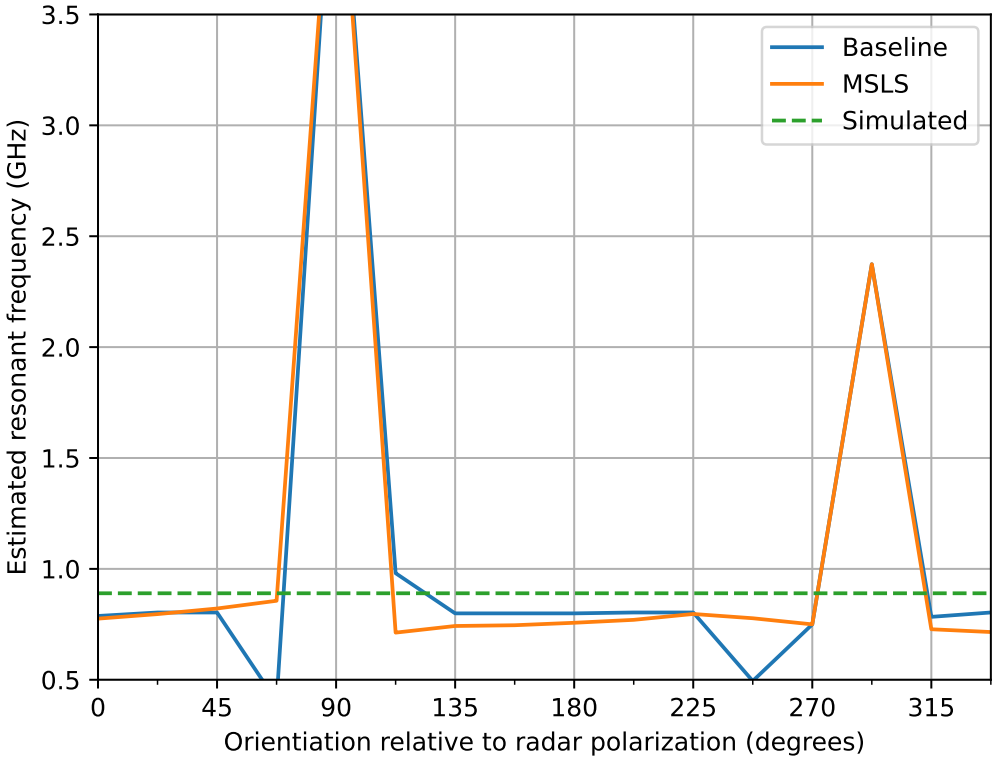

Figure 3.8: The estimated resonance frequency of a 15-cm-long 4.8-mm-thick brass rod at various orientations relative to the radar polarization. Note that when the rod is orthogonal to the radar polarization, the estimate is poor, as can be expected, but the proposed MSLS method yields consistently good estimates in all other cases.

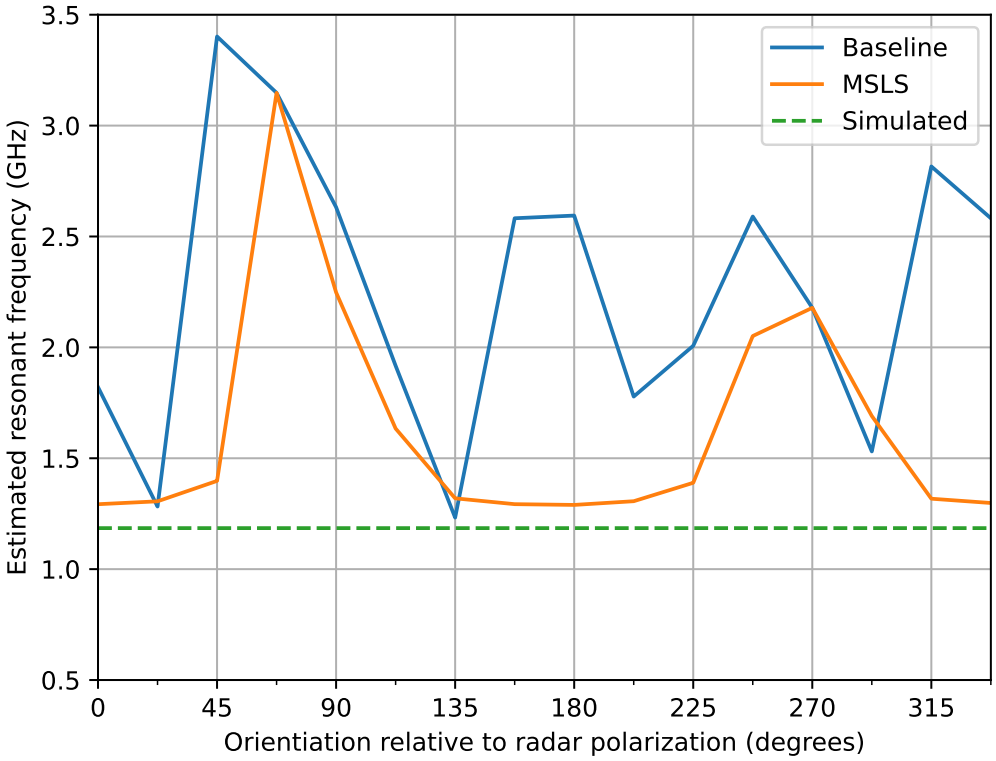

Figure 3.9: The estimated resonance frequency of a 10-cm-long 4.8-mm-thick brass rod at various orientations relative to the radar polarization. Note that proposed MSLS method maintained good performance, despite the smaller target.

3.6.2 Realistic Targets in Clutter

In this section, we show how the proposed approach performs on realistic targets in the presence of background clutter. The targets are a hatchet, a (facsimile of a) Colt 1911 pistol, a kukri, and a Bowie knife; see Figure 3.10. The targets are first measured in the absence of significant clutter (referred to as a "clean" measurement), as shown in Figure 3.10. The resulting estimates of the resonant frequencies and attenuation rates obtained by the proposed MSLS method represent our best estimate of the ground truth for each target, and are marked by green filled circles in Figures 3.12–3.15. We then measure the targets in the presence of "light" clutter, where two orthogonal rods are placed 8 cm in front of the target, and in the presence of "heavy" clutter, where a hammer and unfolded scissors are placed 8 cm in front of the target; see Figure 3.11. Both of these represent a challenging clutter environment because the clutter cannot be obscured by a polarization mismatch or by the target itself. For these experiments, the BG measurement set again contains $M_{\text{BG}} = 100$ shots, but the TG&BG set contains 100 shots at each of two orthogonal polarizations corresponding to orientations of 22.5° and 112.5° for a total of $M_{\text{TG\&BG}} = 200$ shots. Because the estimator in (3.48) does not require consistent residues between measurement shots, we are able to simply combine measurement shots in this fashion.

The estimates of the resonant frequencies and attenuation rates for the hatchet are given in Figure 3.12. The baseline method is unable to provide estimates that resemble the ground truth in either the light clutter or heavy clutter scenarios. However, in both of these scenarios, the proposed method provides good estimates of both resonant frequencies and good estimates of the attenuation rate of the resonance at around 1 GHz. The estimates of the attenuation rate of the resonance around 2.75 GHz

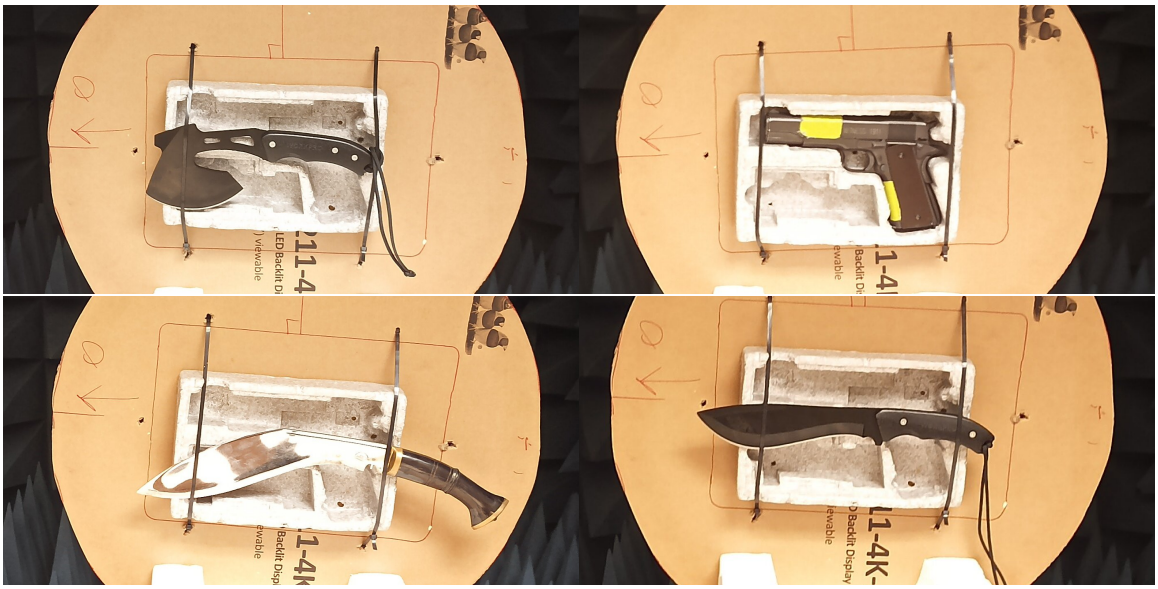

Figure 3.10: Measurement set-ups for the "clean" measurements of a small hatchet, a Colt-1911 facsimile, a kukri and a Bowie knife.

exhibit greater variation than the estimates of the frequency.

The parameter estimates for the facsimile of the Colt 1911 are given in Figure 3.13. The proposed method is quite accurate throughout, particularly in terms of frequency. The baseline method is also quite effective in the case of light clutter, but in the case of heavy clutter it does not detect the "ground truth" pole around 2.6 GHz.

The parameter estimates for the kukri are given in Figure 3.14. The "ground truth" resonant frequencies, which correspond to the ordinate of the green filled circles, are around 1 GHz and 2.75 GHz. As such, they are quite similar to those of the hatchet. Since the metallic portions of the kukri and the hatchet are of similar size, that is expected. However, the estimators perform somewhat differently in this case. In the case of light clutter, the proposed method provides good estimates of the poles,

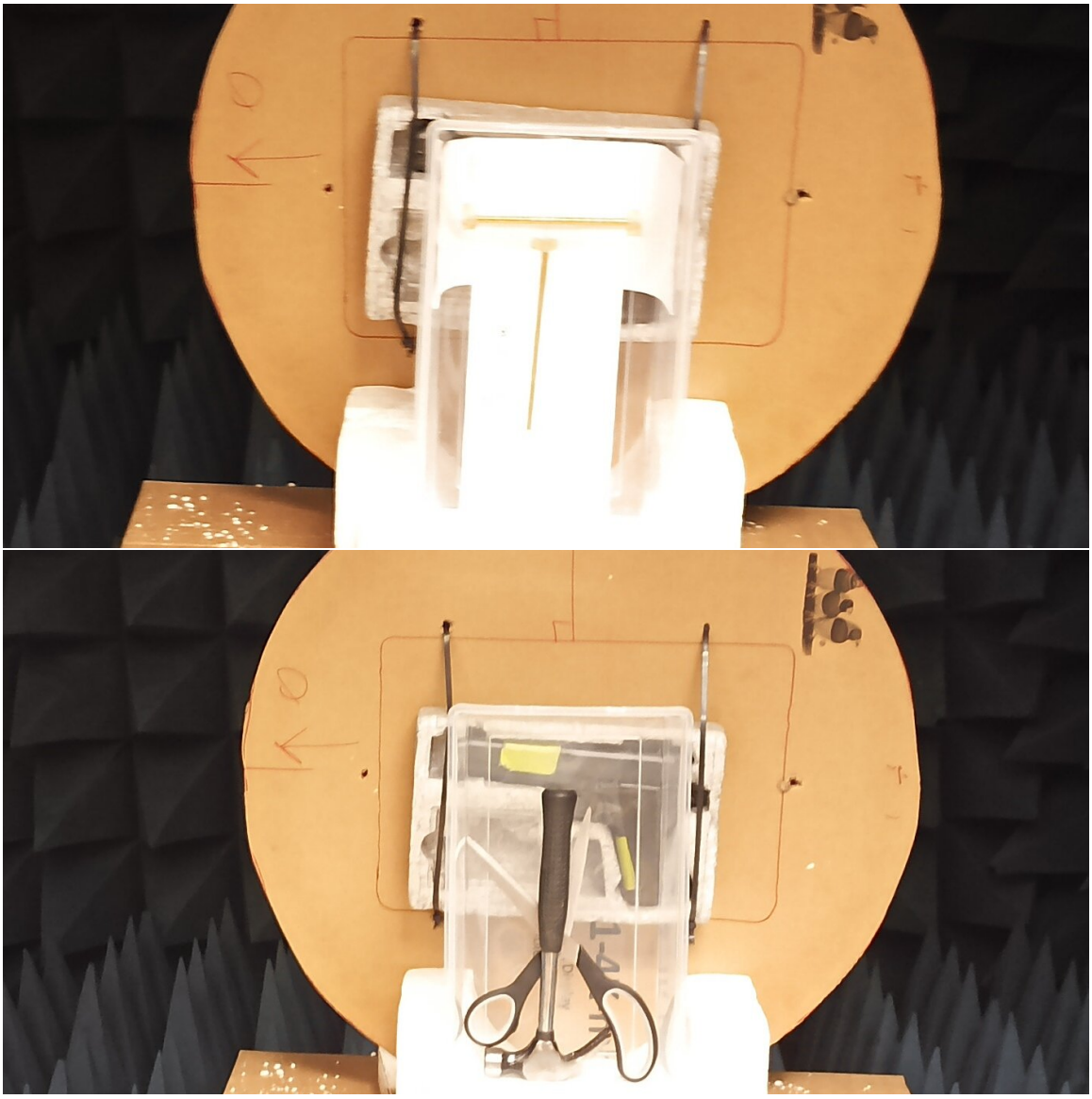

Figure 3.11: Measurement setups for the facsimile of the Colt-1911 in the light clutter environment (top) consisting of two orthogonal rods 8 cm in front of the target, and in the heavy clutter environment consisting of a hammer and a pair of scissors 8 cm in front of the target.

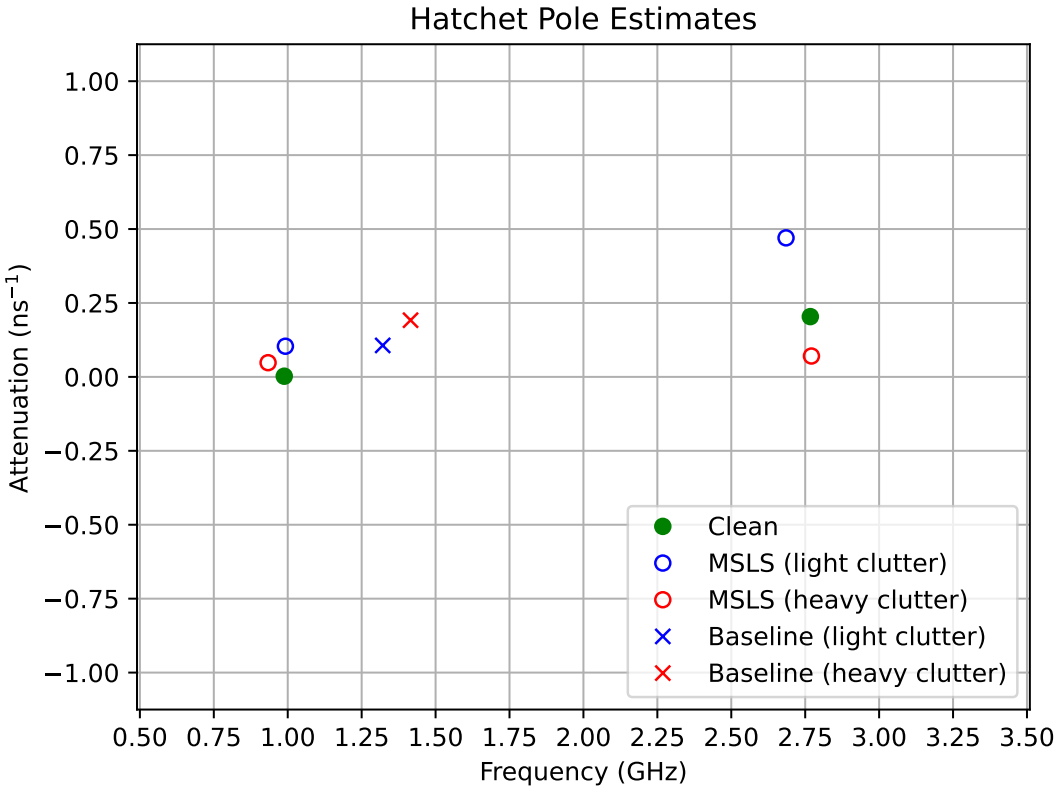

Figure 3.12: Estimates of the resonant frequency and attenuation rate for the hatchet.

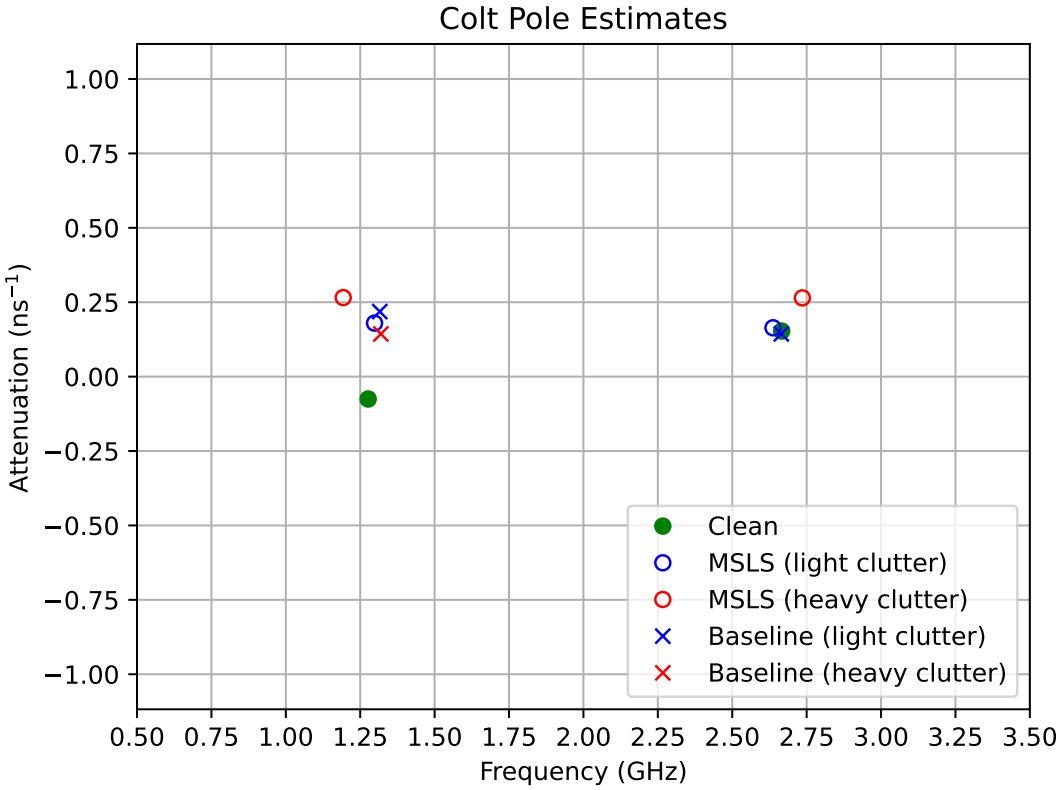

Figure 3.13: Estimates of the resonant frequency and attenuation rate for the Colt 1911.

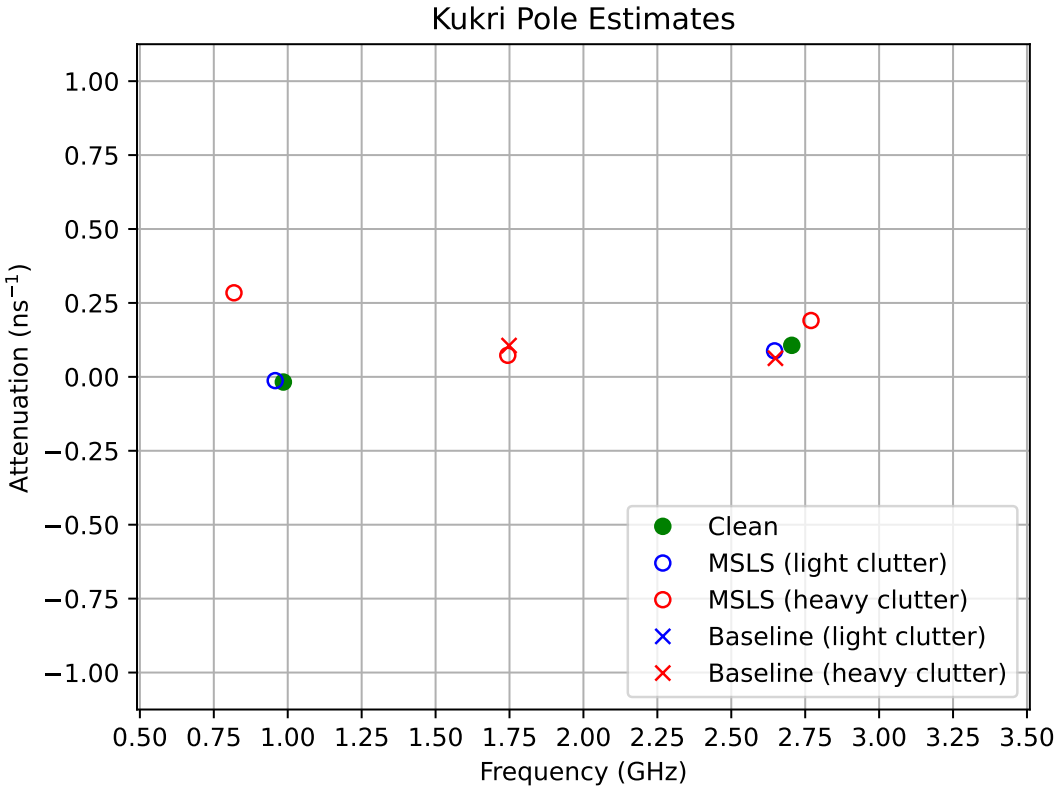

Figure 3.14: Estimates of the resonant frequency and attenuation rate for the kukri.

but the baseline method did not generate any pole estimates within the range of values provided on the figure, which is the region where the radar performs reliably. In the case of heavy clutter, the proposed method provides a reasonable estimate of the higher frequency pole, but that of the lower frequency pole is somewhat degraded. Furthermore, the proposed method does not sufficiently suppress a clutter pole at around 1.75GHz. The baseline method also picks up this clutter pole, along with the higher-frequency "ground truth" pole, but it does not detect the lower-frequency "ground-truth" pole.

The parameter estimates for the Bowie knife are given in Figure 3.15. In this

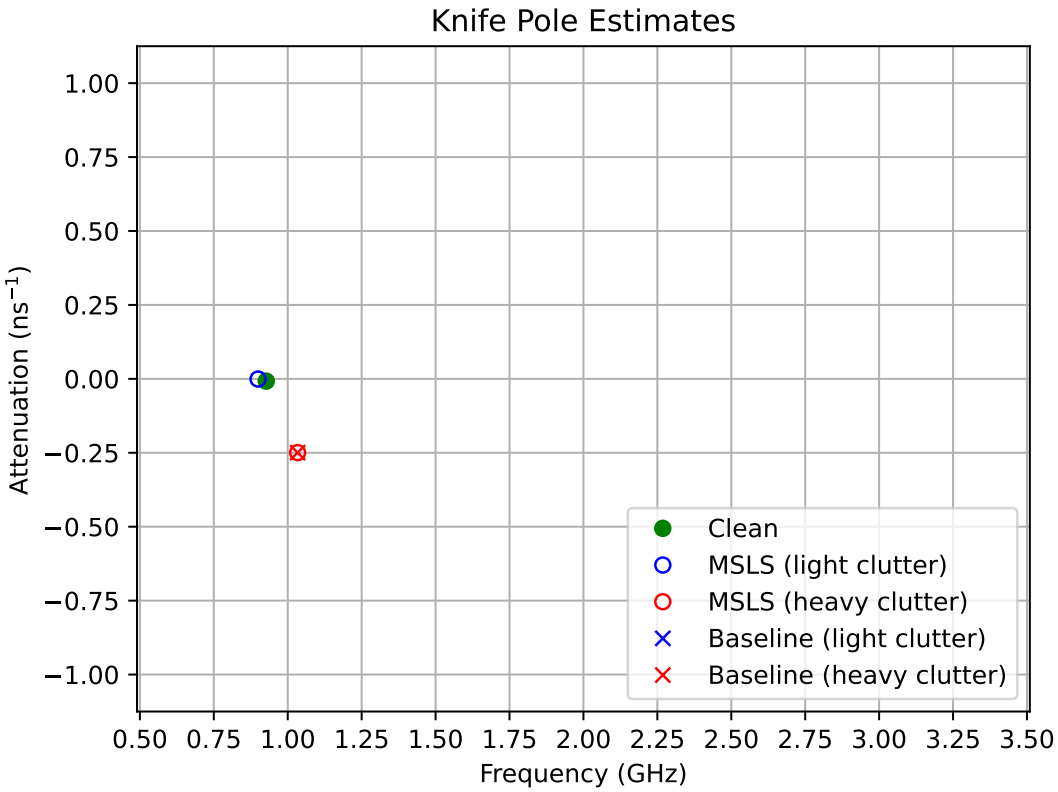

Figure 3.15: Estimates of the resonant frequency and attenuation rate for the Bowie knife.

setting we only have one "ground truth" pole, at around 0.9 GHz. In light clutter, the proposed method provides a good estimate of this pole, but the baseline method does not produce any estimate within the scope of the figure, which is the region where the radar performs reliably. In heavy clutter, both the proposed method and the baseline method identify a clutter pole at just over 1 GHz.

Looking at these results collectively, it can be seen that the proposed approach provides very good results in the light clutter environment, whereas the baseline method fails in the cases of the hatchet, kukri and Bowie knife. As can be seen

from Figure 3.11, the heavy clutter environment poses a serious challenge to any estimator. However, for the hatchet and the (facsimile of the) pistol, the proposed method continues to provide very good estimates, while the baseline method fails. For the kukri and the Bowie knife, the proposed method performs at least as well as the baseline method.

3.7 Conclusion

We have developed a novel algorithm for estimating the complex frequencies of the resonances of targets in the presence of significant environmental clutter. The algorithm is inspired by the observation that the interplay between the resonances of a target and the clutter is made complicated predominantly due to their complex amplitudes. That suggested the development of an estimation algorithm that is minimally dependent on those amplitudes. We developed such an algorithm by using the structure of the maximum likelihood estimator for scenarios in which the noise is Gaussian. We have experimentally demonstrated that the proposed algorithm gives consistent estimates, even in the presence of significant clutter, particularly with regard to the estimation of the resonant frequency.

Appendices

3.A Generalized Pencil-of-Functions Method

The generalized pencil-of-functions (GPoF) method [49], a generalization of the matrix-pencil method (MPM) [50], is a technique for extracting poles $\{z_k\}_{k=1}^{K}$ from a noise-free signal $\mathbf{x} = \mathrm{Vand}_N(\mathbf{z})\,\mathbf{r}$ of length N, where $[\mathbf{z}]_k = z_k$ and $K \leq N/2$. Both methods rely on the structure of $\mathbf{x}$. In particular, they are based on the observation that if one chooses a window-length parameter W such that $K \leq W \leq N - K$, and constructs Hankel matrices $\mathbf{X}_0, \mathbf{X}_1 \in \mathbb{C}^{(N-W)\times W}$ such that

$$[\mathbf{X}_p]_{qr} = x\,[p + q + r]\,, \tag{3.52}$$

then the set of poles $\{z_k\}$ is equal to the set of K *nonzero* generalized eigenvalues of the *matrix pencil* $(\mathbf{X}_0, \mathbf{X}_1)$; i.e., the non-zero solutions $\{\lambda_k\}$ to

$$\det(\lambda\mathbf{X}_0 - \mathbf{X}_1) = 0. \tag{3.53}$$

Given the absence of noise, the remaining $W-K$ generalized eigenvalues are zero. One way to compute the generalized eigenvalues of $(\mathbf{X}_0, \mathbf{X}_1)$ is to compute the conventional eigenvalues of $\mathbf{X}_0^{\dagger}\mathbf{X}_1$, where $(\cdot)^{\dagger}$ denotes the Moore-Penrose pseudo-inverse. In the GPoF method, a particular truncated singular-value decomposition (SVD) of $\mathbf{X}_0$ is employed, which allows (3.53) to be solved in a more robust way than the MPM.

In practice, the MPM and GPoF methods are applied to noisy measurements of $\mathbf{x}$, i.e., $\mathbf{y} = \mathbf{x} + \mathbf{v}$. In that case, the related pencil $(\mathbf{Y}_0, \mathbf{Y}_1)$ and system matrix $\mathbf{Y}_0^{\dagger}\mathbf{Y}_1$ are randomly perturbed from their noise-free counterparts. It is known that even a

small perturbation can lead to a large perturbation in the eigenvalues [73], and hence the MPM and GPoF methods can be quite sensitive to noise. The SVD truncation employed in the GPoF method mitigates this somewhat, but does not eliminate it. Alternatively, if the properties of the noise are known and the signal is long enough, the pencil can be augmented to compensate for the cumulative effects of the noise on the system matrix [37].

3.B Pole Clustering

In [8], Barone acknowledges the noise sensitivity of the GPoF method, but also makes the useful observation that repeated application thereof leads to clustering of the pole estimates; i.e., the probability distribution PDF (z) of the pole estimates from the noisy measurement shots is concentrated around the true poles. To use Barone's observation in the multi-shot measurement scenario, we note that we have M measurement shots $\mathbf{y}_m = \mathbf{x} + \mathbf{v}_m$ of length N, where $\mathbf{x} = \mathrm{Vand}_N(\mathbf{z})\,\mathbf{r}$ with poles $\{z_k\}_{k=1}^K$ arranged in a vector $\mathbf{z}$ where $[\mathbf{z}]_k = z_k$. We then apply the GPoF to each $\mathbf{y}_m$ to generate $K \leq W \leq N/2$ estimates of the poles $\{\hat{z}_{km}\}_{k=1}^W$. For a sufficiently high effective SNR, the pole estimates $\bigcup_{m=1}^{M} \{\hat{z}_{km}\}_{k=1}^W$ cluster in the vicinity of each z_k, and the K most prominent "centers" of these clusters can be used to form multi-shot estimates of the poles $\{\hat{z}_k\}_{k=1}^K$.

To avoid the various difficulties associated with a true clustering technique (cf. [8]), in our implementation of this approach, we opt to use kernel density estimation (KDE) [83] to identify the "centers" of the clusters. The KDE approximates the distribution of the poles as a linear combination of basis functions or *kernels* $g_b(z)$;

i.e.,

$$\mathrm{KDE}\left(z\right) = \frac{1}{M}\sum_{m=1}^{M}\frac{1}{W}\sum_{k=1}^{W} g_b\left(\hat{z}_{km} - z\right).$$

Here, we use a Gaussian kernel with standard deviation or *bandwidth* b; i.e.,

$$g_b\left(z\right) = \frac{1}{2\pi b}\exp\left(-\frac{zz^*}{2b^2}\right).$$

We simply take the K largest local maxima of the KDE as the cluster "centers." If the number of poles K is not known, then the number of "significant" maxima can be used as an estimate thereof. Notice that as $M \to \infty$, $\mathrm{KDE}\left(z\right) \to \left(\mathrm{PDF} * g_b\right)\left(z\right)$, where $*$ denotes convolution. Thus, it is important to choose the bandwidth b to be wide enough that each cluster of poles only has one peak, but not so wide that the clusters blur together.

We further make the observation that when $W > K$, the excess pole estimates often either have very small residues or high rates of attenuation, both of which imply low energy. Therefore, a straightforward way to suppress the effect of spurious poles in the KDE is to weight the samples by their energies. That is, each resonant component $\mathbf{u}_{km}$ with elements $\left[\mathbf{u}_{km}\right]_n = r_k z_k^n$ for $n = 0, \ldots, N-1$ has an associated energy $E_{km} = \mathbf{u}_{km}^{\mathrm{H}}\mathbf{u}_{km}$, where $\left(\cdot\right)^{\mathrm{H}}$ denotes the Hermitian transpose. Hence, we use the weighted KDE

$$\mathrm{KDE}_{\mathrm{weighted}}\left(z\right) = \frac{1}{E}\sum_{m=1}^{M}\sum_{k=1}^{W} E_{km} g_b\left(\hat{z}_{km} - z\right),$$

where $E = \sum_{m=1}^{M}\sum_{k=1}^{W} E_{km}$ is the the total energy of the W resonant components.

This weighted KDE is biased in favor of energetic poles, but when one is only interested in the cluster centers, this can be beneficial when a limited number of measurement shots is available.

Chapter 4

Classification of Radar Targets via Distribution Matching of Late-Time Resonance Parameters

This work is published in the *IEEE Transactions on Radar Systems* as [38]. Its reproduction here is in accordance with the IEEE's Copyright Policy.

Abstract

A promising non-imagining approach to the classification of radar targets is to use the frequencies and attenuation rates of the resonant modes that present during a target's late-time response (LTR) as features. Unfortunately, the estimation of these resonance parameters is rather sensitive to noise. However, we observe that when a large number of measurements of the LTR can be taken in a short time, the probability distribution of the estimates of the parameters can be estimated, and then

matched against a database of such distributions. That has the potential to reduce the sensitivity of the classification problem to noise. In this chapter, we develop a pragmatic approach to target classification using this distribution-matching approach, and demonstrate its effectiveness through physical experiments. The proposed approach is shown to be highly robust to environmental clutter and somewhat robust to target orientation.

4.1 Introduction

Beyond the eponymous tasks of detection and ranging, radar systems also offer a number of other opportunities, including the opportunity to identify targets from their signal returns. The techniques for target identification can be grouped into imaging techniques (e.g., [4]), and non imaging techniques (e.g., [12]). Imaging techniques, like airport mm-wave scanners, have the advantage of producing an image that is intuitive for a human to interpret, but they also have the drawback that they require an array of antennas, or a moving antenna, to generate this image. Non-imaging techniques have the advantage of only needing a few antennas, sometimes just one, but they do not produce anything that is quite as intuitive as an image. That can actually be an advantage in certain applications where privacy may be a concern. In this chapter, we will focus on non-imaging systems that employ time-domain impulse radar, e.g., [31]. Time-domain impulse radars transmit a narrow pulse and hence they probe the environment across a wide band of frequencies. The transmitted pulse stimulates a response from a target or targets in the environment, and this response is measured at the receiver. The initial part of the response is the reflection of the pulse by the target, and called the *early-time response (ETR)*. The pulse also causes the

target to start resonating for some time, which is called its *late-time response (LTR)*. By analogy, the initial twang when a guitar string is plucked is akin to its ETR, and the subsequent sustain (ringing) is akin to its LTR. The frequencies and attenuation rates of the LTR resonances, which we refer to collectively as the *resonance parameters* hereafter, are indicative of the target [12], and have been used as features in the classification of aircraft [75], and buried ordnance [21], [47]. A variety of techniques have been developed for estimating the resonance parameters from a single measurement shot of an impulse radar (e.g., [17], [20], [39], [49], [56], [84]), but in typical applications the LTR has a rather low effective[1] signal-to-noise ratio (SNR), and the estimation problem itself is quite sensitive to perturbations caused by noise and other impairments [73], [48], [37]. This has prompted the development of techniques that can mitigate the sensitivity to some degree. One class to such techniques is based on performing single-shot estimation on different measurements of a target and looking for agreement between the estimates. These different measurements may include different observation windows of the same measurement shot [41], or repeated measurement shots that differ only in their noise content [8]. One can then combine these individual estimates to construct a "best-fit" estimator for the resonance parameters [8]. Another class of techniques that mitigate the sensitivity of resonance parameter estimates involves the construction of an estimator that jointly processes multiple measurements (in contrast to processing the estimates from each measurement); e.g., [34], [36], [55]. However, such techniques can be sensitive to the assumptions that are made regarding the relationships between the measurements. In some applications, the ultimate goal is more than simply estimating the resonance parameters of

[1]The effective SNR is the ratio of the energy of the signal over the observation window of the LTR to the energy of the noise over the observation window of the LTR.

the LTR of an unknown target. Instead, we seek to use the LTR measurements to classify the unknown target against a database of known targets. Early approaches to this problem included methods that synthesize a transmitted waveform that yields a certain "signature" in the returned signal if the target corresponds to a particular entry in the database [22], [52], [75]. However, in an impulse radar system, the transmitted waveform typically has a fixed pulse shape. An approach that is applicable to impulse radar systems (with a single measurement shot) begins with storing the resonance parameters of known targets in a database. Classification of an unknown target then involves determining which entry in the database is the most likely to have generated the measured LTR [62], or which entry has the largest canonical correlation with the measured LTR [23]. However, such an approach retains much of the sensitivity of the underlying estimation problem for the resonance parameters that is used in the construction of the database. It is also sensitive to the presence of clutter in the measurements of the unknown target. An alternative approach to target classification would be to construct a database of the distributions of the resonance parameter estimates for each target, rather than a database of the estimates of the resonance parameters themselves. In that setting, multiple-hypothesis testing can be used to decide to which distribution a resonance estimated from a single measurement of the LTR of an unknown target is most likely to belong. When multiple measurements of the unknown target can be taken in a time frame within which the environment is static, we can avoid explicit estimation of each resonance of the unknown target and instead estimate the distribution of the resonances of the unknown target. That estimated distribution can then be tested against the distributions of known targets; this is called *distribution matching*. An advantage of this approach is that we do not

have to determine accurate estimates of the resonance parameters themselves, neither during the construction of the database, nor during operation, and hence the impact of the sensitivity of that estimation problem is reduced. In this chapter, we consider an impulse radar system that has a high pulse repetition rate, and we describe a technique for classifying unknown targets based on multiple measurements of the LTR of the unknown target and distribution matching. In the proposed technique, the resonance parameters from the LTR of a single measurement shot are estimated with the relatively simple generalized pencil-of-functions (GPoF) method [49]. Repeating this for all available shots of a given target yields a large number of estimates. Their distribution is empirically approximated via kernel density estimation (KDE) [83]. In the preparatory phase of the technique, empirical distributions of known targets are constructed in this way and stored in a database. In the operational phase, the constructed empirical distributions of unknown targets are matched against those in the database. Inspired by the principles of hypothesis testing, the empirical distributions are compared using a carefully chosen approximation of the Kullback-Leibler divergence. The approximation is a variant of a "plug-in" type estimate that we have refined for our application by using insight into the radar system and its environment to mitigate the impact of clutter. The effectiveness of the proposed technique is demonstrated using physical experiments on an ultra-wideband impulse radar operating in a scenario that involves classification of handheld weapons. The chapter is organized as follows: After establishing the system model in Section 4.2, we describe how the resonance parameters can be estimated from an individual measurement shot using the relatively simple generalized pencil-of-functions (GPoF) method [49]; see Section 4.3. In Section 4.4, we elaborate on the principles of the proposed method. As

described above, this involves constructing empirical distributions of the parameter estimates using kernel density estimation (see Section 4.5.1), and careful construction of a metric for comparing distributions; see Sections 4.5.2, 4.6.1 and 4.6.2. With that mathematical framework in place, the method itself is described in Section 4.6.3. Finally, the results from the physical experiments on an ultra-wideband impulse radar system are reported in Section 4.7.

4.2 Signal Model

The model of the late-time response (LTR) to an impulsive stimulus, in the absence of noise, is

$$x(t) = \sum_{k=1}^{K} r_k z_k^t, \tag{4.54}$$

which, assuming that the timebase is normalized to the sampling period, is uniformly sampled at instants $t = n$ to yield a discrete-time model

$$x[n] = \sum_{k=1}^{K} r_k z_k^n, \tag{4.55}$$

where $r_k \in \mathbb{C}$ are the *residues* and $z_k \in \mathbb{C}$ are the *poles*. Each residue $r_k = a_k e^{j\phi_k}$ has an *amplitude* $a_k \in \mathbb{R}$ and a *phase* $\phi_k \in \mathbb{R}$. Each pole $z_k = e^{-\alpha_k + j\omega_k}$ has an *attenuation rate* $\alpha_k \in \mathbb{R}$ and a *frequency* $\omega_k = 2\pi f_k \in \mathbb{R}$. Collectively, ω_k and α_k will be referred to as the resonance parameters of the k-th pole, as discussed in the Introduction. The resonance parameters can also be combined into a *complex frequency* $\tilde{\omega}_k = \omega_k + j\alpha_k$ or an *s-domain pole* $s_k = j\tilde{\omega}_k = -\alpha_k + j\omega_k = \log z_k$. It follows that the estimation of the resonance parameters $\{(\alpha_k, \omega_k)\}_{k=1}^{K}$ is equivalent to

the estimation of the poles $\{z_k\}_{k=1}^K$. Our proposed classification method is based on the distribution of the estimates of the poles $\{z_k\}_{k=1}^K$ from multiple measurements of the LTR. In the idealized case, the radar stimulus can be modeled as being sufficiently impulsive, and the noise can be modeled as being additive. In that case, the m-th measurement can be modeled as

$$x(t) + v_m(t),$$

for $m = 1, 2, \ldots, M$, where $v_m(t)$ denotes the additive noise, which is often modeled as being Gaussian and white. A more accurate model for the measurements would take into account the linear effects of the antennas, filters and amplifiers at the transmitter and the receiver, the structure of the transmitted pulse, and a variety of noise sources. If the component of the received signal that is due to the noise in the transmitter being reflected off the target lies below the receiver noise level, then a more accurate model for the the m-th measurement would be

$$y_m(t) = (x * h)(t) + w_m(t), \tag{4.56}$$

where $h(t)$ is the impulse response of the concatenation of the transmitter and receiver components, $*$ is the convolution operator, and $w_m(t)$ is the sum of the environmental noise, filtered by the front end components of the receiver, and the noise generated by the receiver itself. Using standard Laplace-transform analysis, the first term in (4.56) can be rewritten as

$$(x * h)(t) = \sum_{k=1}^{K} r'_k z_k^t + \sum_{k=K+1}^{K+L} r'_k z_k^t, \tag{4.57}$$

for some residues $\{r'_k\}$ that may be different from those in (4.54), where L is the number of poles of the equivalent filter $h(t)$. The expression in (4.57) indicates that the residues are altered by the radar, but the poles are not. Therefore, we can estimate the poles from the measurements of the form in (4.56), without the need to deconvolve the measured signal (and the extensive calibration needed to perform deconvolution; e.g., [43]). Our approach to classifying the targets will be based on M measurement shots, each of which consists of N samples of the signal in (4.56); i.e.,

$$y_m[n] = \sum_{k=1}^{K+L} r'_k z_k^t + w_m[n], \quad n = 0, 1, \ldots, N-1, \tag{4.58}$$

where $m = 1, 2, \ldots, M$. As suggested by (4.56), assuming that the noise at the receiver is uncorrelated may result in significant model mismatch. In order to avoid the calibration required to estimate the noise correlation, we will develop a technique that makes minimal assumptions on the noise.

4.3 Pole Estimation

Before we discuss the proposed method, which is based on the distributions of the pole estimates, let us review a popular method for the pole estimation itself. The poles $\{z_k\}_{k=1}^{(K+L)}$ can be estimated using the generalized pencil-of-functions (GPoF) method [49]. That method is based on the observation that the poles are the $(K+L)$ non-zero solutions to the following generalized eigenvalue problem: *Given* $x[n]$ *in* (4.55) *for* $n = 0, 1, \ldots, N$, *and* $W \in [K+L, N-1]$, *find* z_k *and* $\mathbf{q}_k$ *such that*

$$\mathbf{X}_0 \mathbf{q}_k = z_k \mathbf{X}_1 \mathbf{q}_k, \tag{4.59}$$

where $\mathbf{X}_\ell \in \mathbb{R}^{(N-W)\times W}$, $[\mathbf{X}_\ell]_{ij} = x[\ell + i + j]$. The remaining $(W - K - L)$ generalized eigenvalues are all zero. The name of the method is derived from that fact that the pair $(\mathbf{X}_0, \mathbf{X}_1)$ is called a matrix pencil. The principle of the GPoF method is to apply the observation in (4.59) to a noisy measurement $y_m[n]$ (cf. (4.58)), to determine W pole estimates $\{\hat{z}_{km}\}_{k=1}^{W}$. Since the excess $(W - K - L)$ generalized eigenvales will be non-zero in this setting, we must then determine which of the W estimates correspond to the $(K+L)$ poles of the LTR. Unfortunately, the generalized eigenvalues of a matrix pencil are notoriously sensitive to perturbations in the matrices [37], [48], [73], [34]. While the GPoF method incorporates a truncated singular value decomposition that mitigates that sensitivity to some degree, it remains difficult to obtain reliable estimates of the poles from a single measurement unless that measurement has a large effective signal-to-noise ratio (SNR) and a large number of samples, N. In scenarios in which the radar can take M measurement shots in a time interval over which the resonance parameters remain constant, it has been observed [8], [10] that the W generalized eigenvalues from each measurement shot, $\{\hat{z}_{km}\}_{k=1}^{W}$, $m = 1, 2, \ldots, M$, tend to form clusters near the "true" poles $\{z_k\}$ of the (sampled) LTR, $x[n]$, with the excess eigenvalues scattered across the complex plane, typically near the unit circle. That suggests that if we have a sufficient number of measurement shots, M, at a sufficiently large signal-to-noise ratio, we should be able to obtain good estimates of the poles by jointly processing the pole estimates, $\left\{\{\hat{z}_{km}\}_{k=1}^{W}\right\}_{m=1}^{M}$, from all the measurement shots $\{y_m[n]\}_{m=1}^{M}$. One approach is to cluster the pole estimates [8]. That approach is appropriate if parameter estimation is the ultimate goal. For applications in which target classification is the ultimate goal, we seek a method in which the pole estimates $\left\{\{\hat{z}_{km}\}_{k=1}^{W}\right\}_{m=1}^{M}$ are jointly processed with the classification

goal in mind, instead of being jointly processed to achieve an interim goal of parameter estimation. In the next section, we will outline the principles of our proposed approach to doing so.

4.4 Principles of proposed method

In the proposed method, we combine the pole estimates from all the measurement shots into a flat set $\{\hat{z}_j\}_{j=1}^{(WM)}$, where $\hat{z}_{(k-1)W+m+1} = \hat{z}_{km}$. Due to the additive noise in the measurement shots, each $\hat{z}_j$ can be viewed as being a random variable drawn from an (unknown) distribution that describes the pole estimates. Each target will have a different distribution, and hence one can envision a classification scheme in which the pole estimates from the unknown target, $\{\hat{z}_j\}_{j=1}^{(WM)}$, are used to construct an empirical distribution, and classification is performed by comparing that distribution to distributions generated by known targets. In the following section, we will describe the basic mathematical tools that underlie our approach, and in Section 4.6, we will put the principles of our approach into practice by modifying these basic mathematical tools so that they account for practical considerations, including the bandwidth of the radar and the presence of environmental clutter.

4.5 Mathematical Preliminaries

4.5.1 Constructing an Empirical Distribution

As described in the previous section, the pole estimates generated from the measurements can be viewed as being drawn from an unknown distribution, which we will

denote by $\mathcal{P}$; i.e., $\hat{z}_p \sim \mathcal{P}$. In this section we describe a well-established method for constructing an approximation of the corresponding probability density function (PDF), $\text{PDF}_{\mathcal{P}}(z)$. To simplify the exposition of the material in the following section, we will adopt a mild abuse of notation and use the index of the pole estimates to indicate the distribution with which they are associated. In the case of a distribution $\mathcal{P}$, that will mean that we index the pole estimates by p. In particular, given M measurement shots, each of length N, $\{y_m[n]\}_{m=1}^{M}$ in (4.58), we let $\{\hat{z}_p\}_{p=1}^{(WM)}$ denote the pole estimates generated by the GPoF method described in Section 4.3. We will let $P = WM$ denote the number of pole estimates. To construct the desired empirical distribution, one could simply make a histogram with each bin spanning a range of the frequency and attenuation rate parameters, as in [41]. However, such an approach is very sensitive to the choice of bins, in terms of both their size and where their edges lie. An alternative that is also reasonably simple is the kernel density estimate (KDE), which is a superposition of Gaussian kernels centered at each sample [83]. The KDE does not suffer from the bin-width and bin-edge sensitivities of the histogram approach, and we will opt to use it here. A key tuning parameter of the KDE is its bandwidth, the standard deviation of its Gaussian kernels. These kernels are typically chosen to be radially symmetric. In [34], it is observed that the distribution of pole estimates, $\mathcal{P}$, is also roughly radially symmetric, particularly in the neighborhood of where the pole estimates cluster. (In contrast, the distribution of the estimates of the corresponding s-domain poles or complex frequencies can be significantly skewed.) Therefore, a KDE obtained from $\{\hat{z}_p\}$ with a bandwidth that is smaller than the standard deviation of a cluster should yield a reasonable estimate of the pole probability density. With this in mind, we construct a KDE from a set of

P pole estimates $\{\hat{z}_p\}_{p=1}^{P}$ as

$$\mathrm{KDE}_{\{\hat{z}_p\}}(z) = \frac{1}{P}\sum_{p=1}^{P} g_b(z;\hat{z}_p), \tag{4.60}$$

where $g_b(z;u) = \exp\left(-(z-u)(z-u)^*/2b^2\right)/2\pi b$ is a radially symmetric Gaussian kernel centered at $z = u$ with standard deviation or *bandwidth* b. Due to the construction in (4.60), the KDE is related to the PDF of $\mathcal{P}$ as

$$\lim_{P\to\infty} \mathrm{KDE}_{\{\hat{z}_p\}}(z) = (\mathrm{PDF}_{\mathcal{P}} * g_b)(z).$$

That is, the KDE is an approximation of a smoothed version of the PDF, where the Gaussian kernel is the smoothing function. Therefore, we must thus take care in choosing a bandwidth b that is large enough for the KDE to be relatively smooth for a given set size P, but not so large that adjacent clusters blur together. In preparation for the discussion of our classification method, we define $\mathrm{kde}_{\{\hat{z}_p\}}(z)$ to be the logarithm of the KDE in (4.60); that is

$$\mathrm{kde}_{\{\hat{z}_p\}}(z) = \log \mathrm{KDE}_{\{\hat{z}_p\}}(z). \tag{4.61}$$

4.5.2 Comparing Distributions

In order to describe an appropriate way to compare distributions, we consider the following abstracted problem. Let $\tilde{z}_{\mathcal{P}}$ denote a random variable with an unknown distribution $\mathcal{P}$. We suspect that $\mathcal{P}$ is one of two known distributions; that is $\mathcal{P}$ is either $\mathcal{G}$ or $\mathcal{Q}$. We wish to determine which distribution $\tilde{z}_{\mathcal{P}}$ is more likely to have, i.e., which of the hypotheses $\mathcal{P} = \mathcal{G}$ and $\mathcal{P} = \mathcal{Q}$ is more likely. We will do this based

on a set of observations $\{\hat{z}_p\}$ of $\tilde{z}_{\mathcal{P}}$. Given a single observation $\hat{z}_p$ of $\tilde{z}_{\mathcal{P}}$, $\hat{z}_p$ is more likely to be drawn from $\mathcal{G}$ than $\mathcal{Q}$ if the likelihood ratio, $\Lambda_{\mathcal{Q}}^{\mathcal{G}}(\hat{z}_p)$, where

$$\Lambda_{\mathcal{Q}}^{\mathcal{G}}(z) = \frac{\mathrm{PDF}_{\mathcal{G}}(z)}{\mathrm{PDF}_{\mathcal{Q}}(z)}, \tag{4.62}$$

is greater than 1, [58]. Given multiple independent observations $\{\hat{z}_p\}$, the cumulative likelihood ratio is $\prod_{p=1}^{P} \Lambda_{\mathcal{Q}}^{\mathcal{G}}(\hat{z}_p)$. Alternatively, using the log likelihood ratio $\lambda_{\mathcal{Q}}^{\mathcal{G}}(\hat{z}_p) = \log \Lambda_{\mathcal{Q}}^{\mathcal{G}}(\hat{z}_p)$, the cumulative log likelihood ratio is $\sum_{p=1}^{P} \lambda_{\mathcal{Q}}^{\mathcal{G}}(\hat{z}_p)$, which is positive if the distribution of $\mathcal{P}$ is more likely to be $\mathcal{G}$ than $\mathcal{Q}$. To gain some insight into the classification behavior as the number of observations grows, we note that if the distribution of $\tilde{z}_{\mathcal{P}}$ is indeed $\mathcal{G}$ (that is, if $\mathcal{P} = \mathcal{G}$), then the average of the cumulative log likelihood ratio over the observations $\{\hat{z}_p\}$ of $\tilde{z}_{\mathcal{P}}$ approaches its expected value; i.e.,

$$\lim_{P\to\infty} \frac{1}{P} \sum_{p=1}^{P} \lambda_{\mathcal{Q}}^{\mathcal{G}}(\hat{z}_p) = \mathrm{E}\left(\lambda_{\mathcal{Q}}^{\mathcal{G}}(\tilde{z}_{\mathcal{P}})\right). \tag{4.63}$$

Since $\mathcal{P} = \mathcal{G}$, we can write that expectation as

$$\mathrm{E}\left(\lambda_{\mathcal{Q}}^{\mathcal{G}}(\tilde{z}_{\mathcal{P}})\right) = \mathrm{E}\left(\lambda_{\mathcal{Q}}^{\mathcal{P}}(\tilde{z}_{\mathcal{P}})\right) = \int \mathrm{PDF}_{\mathcal{P}}(z)\lambda_{\mathcal{Q}}^{\mathcal{P}}(z)\mathrm{d}z. \tag{4.64}$$

In the proposed application, we encounter a related classification problem to this abstracted problem. However, we wish to avoid binary hypothesis testing because we wish to classify the measurements from the unknown distribution as belonging to one of more than two existing distributions. It would be easier to simply have a measure of how similar the unknown target's distribution is to an existing one. To that end, we observe that the integral on the right hand side of (4.64) is, in fact, the

Kullback-Leibler (KL) divergence of $\mathcal{P}$ from $\mathcal{Q}$, [58],

$$D_{\mathrm{KL}}(\mathcal{P} \parallel \mathcal{Q}) = \int \mathrm{PDF}_{\mathcal{P}}(z) \lambda_{\mathcal{Q}}^{\mathcal{P}}(z) \mathrm{d}z. \tag{4.65}$$

The KL divergence, which is also known as the *relative entropy* of $\mathcal{P}$ with respect to $\mathcal{Q}$, is a measure of how different $\mathcal{P}$ is from $\mathcal{Q}$. In our application, $\mathcal{P}$ will be the distribution of an unknown target being classified, and $\mathcal{Q}$ will be the distribution of a known target. The corresponding KL divergence will be small if the unknown target is similar to the known one and large otherwise.

4.6 Implementing the Principles

In this section, we seek to use the mathematical tools described in Sections 4.5.1 and 4.5.2 to put the principles outlined in Section 4.4 into practice in the proposed application. Our first observation is that we do not have access to the distribution of the pole estimates of the (unknown) target, $\mathcal{P}$. We only have access to the pole estimates, $\{\hat{z}_p\}_{p=1}^{P}$, that are obtained by applying the GPoF technique in Section 4.3 to each of the multiple measurement shots $y_m[n]$ in (4.58). Similarly, the distributions of the known targets in the database, which we have denoted by $\mathcal{Q}$, are also estimated from measurements. There are a number of available strategies for estimating the KL divergence from the data; e.g., the multi-dimensional methods described in [65] and references therein. To enable us to accommodate practical considerations and environmental clutter in a straightforward way (see Sections 4.6.1 and 4.6.2, respectively), we will employ a "plug-in" estimate of the "resubstitution" type, in the sense of the taxonomy in [13]. To construct that estimate of the KL divergence, we will

begin by approximating the expectation in (4.65) using the finite-sample version of (4.63). That is, for a finite set of samples $\{\hat{z}_p\}_{p=1}^{P}$, we will use the approximation

$$D_{\mathrm{KL}}(\mathcal{P} \parallel \mathcal{Q}) \approx \frac{1}{P} \sum_{p=1}^{P} \lambda_{\mathcal{Q}}^{\mathcal{P}}(\hat{z}_p), \tag{4.66}$$

which is a truncation of the sum in (4.63). However, the fact that the distribution of the pole estimates from the target is inherently unknown, and that the distribution $\mathcal{Q}$ in the database will itself have been constructed from measurements (taken in a scenario in which the target is known), means that we do not have a direct way to calculate the log likelihood ratios $\lambda_{\mathcal{Q}}^{\mathcal{P}}(\hat{z}_p)$. To construct an approximation for the log likelihood ratios, we observe that the likelihood ratio (cf. (4.62)) can be approximated from measurements using the KDE described in Section 4.5.1, namely $\Lambda_{\mathcal{Q}}^{\mathcal{P}}(z) \approx \Lambda_{\{\hat{z}_q\}}^{\{\hat{z}_p\}}(z)$, where the *approximate likelihood ratio* is

$$\Lambda_{\{\hat{z}_q\}}^{\{\hat{z}_p\}}(z) = \frac{\mathrm{KDE}_{\{\hat{z}_p\}}(z)}{\mathrm{KDE}_{\{\hat{z}_q\}}(z)}.$$

Thus, the log likelihood ratio can be approximated by $\lambda_{\mathcal{Q}}^{\mathcal{P}}(z) \approx \lambda_{\{\hat{z}_q\}}^{\{\hat{z}_p\}}(z)$, where the *approximate log likelihood ratio* is

$$\lambda_{\{\hat{z}_q\}}^{\{\hat{z}_p\}}(z) = \mathrm{kde}_{\{\hat{z}_p\}}(z) - \mathrm{kde}_{\{\hat{z}_q\}}(z). \tag{4.67}$$

Using (4.66) and (4.67), we arrive at the following *approximate KL divergence*[2]

$$D_{\mathrm{KL}}\left(\{\hat{z}_p\} \parallel \{\hat{z}_q\}\right) = \frac{1}{P}\sum_{P=1}^{P} \lambda_{\{\hat{z}_q\}}^{\{\hat{z}_p\}}(z_p), \tag{4.68}$$

where, for convenience, we have adopted a mild abuse of notation to distinguish the approximation in (4.68) from the true definition in (4.65) using only the arguments of the function. We will use this approximate KL divergence as an indicator of how much a set of pole estimates from measurements of an unknown target $\{\hat{z}_p\}$ diverges from a set of pole estimates from measurements of a known target $\{\hat{z}_q\}$.

4.6.1 Practical Considerations

When we apply the GPoF method in Section 4.3 to estimate the poles from a single measurement shot $y_m[n]$ in (4.58), we would typically choose the number of poles to be estimated W to be somewhat larger than the sum of the predicted number of poles from the target and radar $(K + L)$. This is to account for poles due to the environment clutter; see Section 4.6.2. Therefore, we expect that some of the pole estimates $\hat{z}_p$ will be artifacts of the noise, rather than being features of the target or the environment. Ideally, if a pole estimate $\hat{z}_p$ is an artifact of the noise, then its log likelihood ratio $\lambda_{\mathcal{Q}}^{\mathcal{P}}(\hat{z}_p)$ will be small with respect to the log likelihood ratio of an estimate that corresponds to a physical pole, and thus it will have little bearing on the KL divergence. Unfortunately, when the number of estimated poles P is finite, it is not guaranteed that the corresponding approximate log likelihood ratio $\lambda_{\{\hat{z}_q\}}^{\{\hat{z}_p\}}(\hat{z}_p)$ is small, and and the resulting errors can accumulate over multiple noise-pole estimates.

[2]As an aside, we point out that, whereas $D_{\mathrm{KL}}(\mathcal{P} \parallel \mathcal{Q}) \geq 0$, [58], the approximation could be slightly negative for very similar distributions, i.e., $D_{\mathrm{KL}}\left(\{\hat{z}_p\} \parallel \{\hat{z}_q\}\right) \gtrsim 0$.

We propose to use some physical insight into the radar system to reduce the impact of these noise poles. In particular, we observe that any pole estimates that lie outside the bandwidth of the radar or attenuate very rapidly are very likely to be noise artifacts, and it makes little sense to include their contributions in the approximation of the KL divergence. Thus, they are removed prior to computing (4.68). This can be modeled by refining our definition of the approximate KL divergence to

$$D_{\mathrm{KL}}\left(\{\hat{z}_p\} \parallel \{\hat{z}_q\}\right) = \frac{1}{P}\sum_{p=1}^{P} M_{\mathrm{phy}}(\hat{z}_p)\lambda_{\{\hat{z}_q\}}^{\{\hat{z}_p\}}(\hat{z}_p), \tag{4.69}$$

where $M_{\mathrm{phy}}(z)$ is a masking (indicator) function such that

$$M_{\mathrm{phy}}(z) = \begin{cases} 1, & z \text{ is a plausible physical pole,} \\ 0, & \text{otherwise.} \end{cases}$$

In our physical experiments described in Section 4.7, the proposed method will be implemented on an ultra-wideband radar that has a bandwidth of approximately $[0.5,\ 5.0]$ GHz, with an equivalent sampling rate of 20 GSa/s. To allow for perturbations in the estimation process, pole estimates with (normalized) frequencies $0.05\pi \leq \omega_k \leq 0.55\pi$ will be considered to be potentially generated by physical resonances. Furthermore, we make the observation that no detectable physical pole will have an estimate with its (normalized) attenuation rate outside $-0.1 \leq \alpha_k \leq 0.2$.[3]

[3]Note that since our targets are passive, no physical pole can have a negative attenuation rate. However, as discussed in in Section 4.3, the estimate of a physical pole by the GPoF method can be significantly perturbed by the noise, and hence an estimate of a physical pole from a single measurement shot may have a slightly negative attenuation rate.

4.6.2 Environmental Clutter

In our intended applications, a target cannot be easily measured in isolation. Typically, some of the extracted poles will belong to background clutter, which can include resonances in the radar itself (such as those described in (4.57)) or something other than the target in the radar's field of view. Various background de-embedding strategies exist. Some are as simple as subtracting a reference background signal prior to any other processing, while others attempt to classify pole estimates as belonging to the background and remove them accordingly. Here, we are inspired by the latter approach, but consistent with our overall strategy, we seek to work with the distribution of the pole estimates. The idea behind our approach to managing environmental clutter is as follows. Given a distribution of background pole estimates $\mathcal{B}$ with log PDF $\text{pdf}_{\mathcal{B}}(z)$ and a distribution of target pole estimates $\mathcal{P}$ with log PDF $\text{pdf}_{\mathcal{P}}(z)$, a pole estimate $\hat{z}_p$ is more likely to be due to the background than the target if the log likelihood ratio $\lambda_{\mathcal{P}}^{\mathcal{B}}(\hat{z}_p)$ is positive. By definition, that is equivalent to $\text{pdf}_{\mathcal{B}}(\hat{z}_p) > \text{pdf}_{\mathcal{P}}(\hat{z}_p)$. Such pole estimates should not play a role in the classification metric. In our implementation of this idea, we use KDEs of the PDFs, and we further refine our notion of the approximate KL divergence to

$$D_{\text{KL}}\left(\{\hat{z}_p\} \parallel \{\hat{z}_q\}\right) = \frac{1}{P}\sum_{p=1}^{P} M_{\neg\text{BG}}(\hat{z}_p) M_{\text{phy}}(\hat{z}_p) \lambda_{\{\hat{z}_q\}}^{\{\hat{z}_p\}}(\hat{z}_p), \tag{4.70}$$

where $M_{\neg\text{BG}}(z)$ is a masking (indicator) function such that

$$M_{\neg\text{BG}}(z) = \begin{cases} 0, & \text{kde}_{\{\hat{z}_p\}}(z) > \text{kde}_{\{\hat{z}_b\}}(z) \\ 1, & \text{otherwise.} \end{cases} \tag{4.71}$$

Since the background that is observed during the classification of an unknown target may be different from the background that was observed during the construction of the database of known targets, it may be to our advantage to replace the condition in (4.71) for $M_{\neg\mathrm{BG}}(z)$ to be zero with the condition

$$\mathrm{kde}_{\{\hat{z}_p\}}(z) > \max\big(\mathrm{kde}_{\left\{\hat{z}_{b,\mathrm{DB}}\right\}}(z), \mathrm{kde}_{\left\{\hat{z}_{b,\mathrm{Cls}}\right\}}(z)\big),$$

where the subscripts DB and Cls identify the background observed during the construction of the database, and the background observed during classification, respectively. Although it is possible for the difference in backgrounds to introduce or remove significant environmental clutter, the most energetic components of the background are usually from the radar system itself. Thus, this modification to the condition usually has a small impact in practice.

4.6.3 Classification Algorithm

The basic structure of the classification problem can be broken down into two phases: building a database from measurements of known targets, which we describe in Algorithm 5; and using measurements of an unknown target to classify against the database, which we describe in Algorithm 6. As can be seen from the algorithm statements, both of these phases require the construction of a KDE of the distribution of the pole estimates. The steps that are taken to construct such a KDE are outlined in Algorithm 4. Algorithm 4 can be viewed as being a summary of Section 4.5.1, but we note that we make use of our radar's polarization diversity, which is necessary when comparing targets at different orientations.

Algorithm 4 PDF Approximation by KDE

1. Acquire $M^{(\mathrm{HH})}$ measurement shots with horizontal-horizontal polarization.
2. Acquire $M^{(\mathrm{VV})}$ measurement shots with vertical-vertical polarization.
3. Extract an N-sample window of the LTR, which lies between the ETR and first multi-path reflection.
4. Extract W poles from each measurement shot using the GPoF method in Section 4.3 to create a set of pole estimates $\{\hat{z}_g\}_{g=1}^{(WM)}$, where $M = M^{(\mathrm{HH})} + M^{(\mathrm{VV})}$.
5. Construct $\mathrm{KDE}_{\{\hat{z}_g\}}(z)$ using (4.60).

Algorithm 5 Database Building

1. For $j = 1, 2, \ldots, J$, construct the KDE for the j-th known target, $\mathrm{KDE}_{\{\hat{z}_{q_j}\}}(z)$, using Algorithm 4.
2. Construct the KDE for the background that was observed when those known targets were measured, $\mathrm{KDE}_{\{\hat{z}_{b,\mathrm{DB}}\}}(z)$, using Algorithm 4.
3. Store these $(J+1)$ KDEs in a database for later use.

4.7 Results

To demonstrate the practical performance of the proposed approach to target classification, we consider an application involving the classification of hand-held weapons. We use a bistatic ultra-wide-band time-domain radar system that uses two quad-ridge antennas to enable full polarization diversity; see Figure 4.16. In our measurements, we will restrict attention to the (co-polarized) horizontal-horizontal (HH) and vertical-vertical (VV) polarizations. The radar's transmitter operates reliably between approximately 0.5 GHz and 5.0 GHz [66], and the receiver employs an equivalent sampling rate of 20 GSa/s [67]. From an estimation perspective, the receiver's

Algorithm 6 Classification

1. Construct the KDE for the unknown target, $\mathrm{KDE}_{\{\hat{z}_p\}}(z)$, using Algorithm 4.
2. Construct the KDE for the background that was observed when that unknown target was measured, $\mathrm{KDE}_{\{\hat{z}_{b,\mathrm{Cls}}\}}(z)$, using Algorithm 4.
3. Load the KDEs for the known targets, $\mathrm{KDE}_{\{\hat{z}_{q_j}\}}(z)$, $j = 1, 2, \ldots, J$, from the database.
4. Load the KDE for the background that was observed when those known targets were measured, $\mathrm{KDE}_{\{\hat{z}_{b,\mathrm{DB}}\}}(z)$, from the database.
5. For $j = 1, 2, \ldots, J$, compute the approximate KL divergence from the j-th known target in the database to the unknown target, $D_{\mathrm{KL}}\left(\{\hat{z}_p\} \parallel \{\hat{z}_{q_j}\}\right)$, using (4.70).
6. Determine $j^\star = \arg\min_{j\in\{1,2,\ldots,J\}} D_{\mathrm{KL}}\left(\{\hat{z}_p\} \parallel \{\hat{z}_{q_j}\}\right)$, the index of the known target from which the unknown target diverges the least.

reliability begins to taper off above 3.5 GHz [36], but for our distribution-matching approach to classification the receiver remains effective up to 5.0 GHz. The targets are placed approximately 1.2 m in front of the radar on a minimally reflectively stand that allows them to be rolled and yawed relative to the radar. In our experiments, we only alter the roll angle; see Figure 4.16. We refer measurements taken in this fashion as "clean." We then take measurements where clutter is added 8 cm in front of the target; see Figure 4.17. In one case, the clutter consists of a pair of orthogonal brass rods, which we refer to as "light clutter." Our choice of the arrangement of the rods ensures that there is at least one problematic clutter resonance regardless of polarization. In the other case, the clutter consists of a hammer and scissors, which we refer to as "heavy clutter." This introduces multiple difficult-to-predict resonances, and weakens the target response somewhat more than the rods. The targets that

we will consider are a (monolithic) hatchet, a (monolithic) Bowie knife, a facsimile of a Colt 1911 pistol, and a kukri; see Figure 4.18. To provide some context for what we might expect from the LTR of these targets, we observe that we can make a very rough estimate of the resonant frequency of a component of a target by approximating that component by a thin rod of the same length. If we denote that length by L, then its fundamental resonant frequency is $f_{\text{res}} = 2c/L$, where c is the speed of light. With this approximation, the fundamental resonant frequency a target of length 30 cm would be 0.5 GHz, which is at the lower limit of what our radar can detect reliably. Since the largest dimensions of our targets are on the order of 30 cm,[4] our classification scheme will rely heavily on the resonances of smaller features and the harmonics of the larger features. As such, the classification task in this setting is quite challenging. In order to employ Algorithm 4 to construct a reliable KDE of the distribution of the pole estimates, we need to make an appropriate choice for the number of pole estimates P in (4.60), and for the bandwidth b of the Gaussian kernel function $g_b(z; \hat{z}_p)$. We have empirically determined that approximately $P = 1000$ pole estimates are sufficient. As discussed at the end of Section 4.5.1, b must be small enough that adjacent clusters of estimates do not blur together, and must be large enough that the KDE does not have artificial peaks in the subdomains that are sparsely sampled. In our experiments we began to observe noticeable blurring when $b \gtrsim 0.06$, and hence we opted for $b = 0.055$. To determine how many measurement shots are required to generate at least this number of pole estimates, we observe that in our set up we have identified an LTR window which is 2.4 ns (48 samples) long, which is nearly the maximum available between the end of the ETR and the arrival

[4] All of the bladed targets are longer, as is the cumulative length of the Colt-1911 pistol's barrel and grip; see Figure 4.18.

of a multi-path reflection from the floor. From the LTR window in each measurement shot, we estimate $W = 10$ poles. This choice is consistent with the "rule-of-thumb" for the GPoF method that from a window of N samples one should seek to estimate at most $N/4$ poles. In our implementation, we take 100 HH measurements and 100 VV measurements for each target for a total of $M = 200$ measurement shots, and hence we obtain 2000 pole estimates. This takes less than 3 seconds; see [68] for more details. In our analysis below, we will only use half of these measurements at a time, and hence our experiments correspond to an acquisition time of 1.5 seconds.

4.7.1 Performance on a Single Set of Measurement Shots

The database of known targets was constructed by applying Algorithm 5 to 100 "clean" measurement shots (50 HH and 50 VV) of each of the four targets at a roll angle of 0°. Classification was carried out by applying Algorithm 6 to 100 measurements of each of the four targets (50 HH and 50 VV), in no, light, and heavy clutter, at roll angles of 0°, +22.5° and −22.5°, for a total of 36 classification tests. (In the uncluttered 0° case, the classification measurements were distinct from the database measurements.) The values for the approximate KL divergence in (4.70) of each test from each entry in the database are shown in the rows of Tables 4.1, 4.2 and 4.3. (The columns represent the known targets in the database.) The smallest divergence in each row, which corresponds to the decision made by the classifier, is highlighted in bold font. In Table 4.1, we observe the effect the clutter alone has on the classification. In the case of no clutter, the proposed classifier is able to confidently classify each target in the sense that the divergence from the database entry that corresponds to the actual target is much smaller than the divergence from the other database

Figure 4.16: The radar setup for taking the "clean" measurements in our experiments.

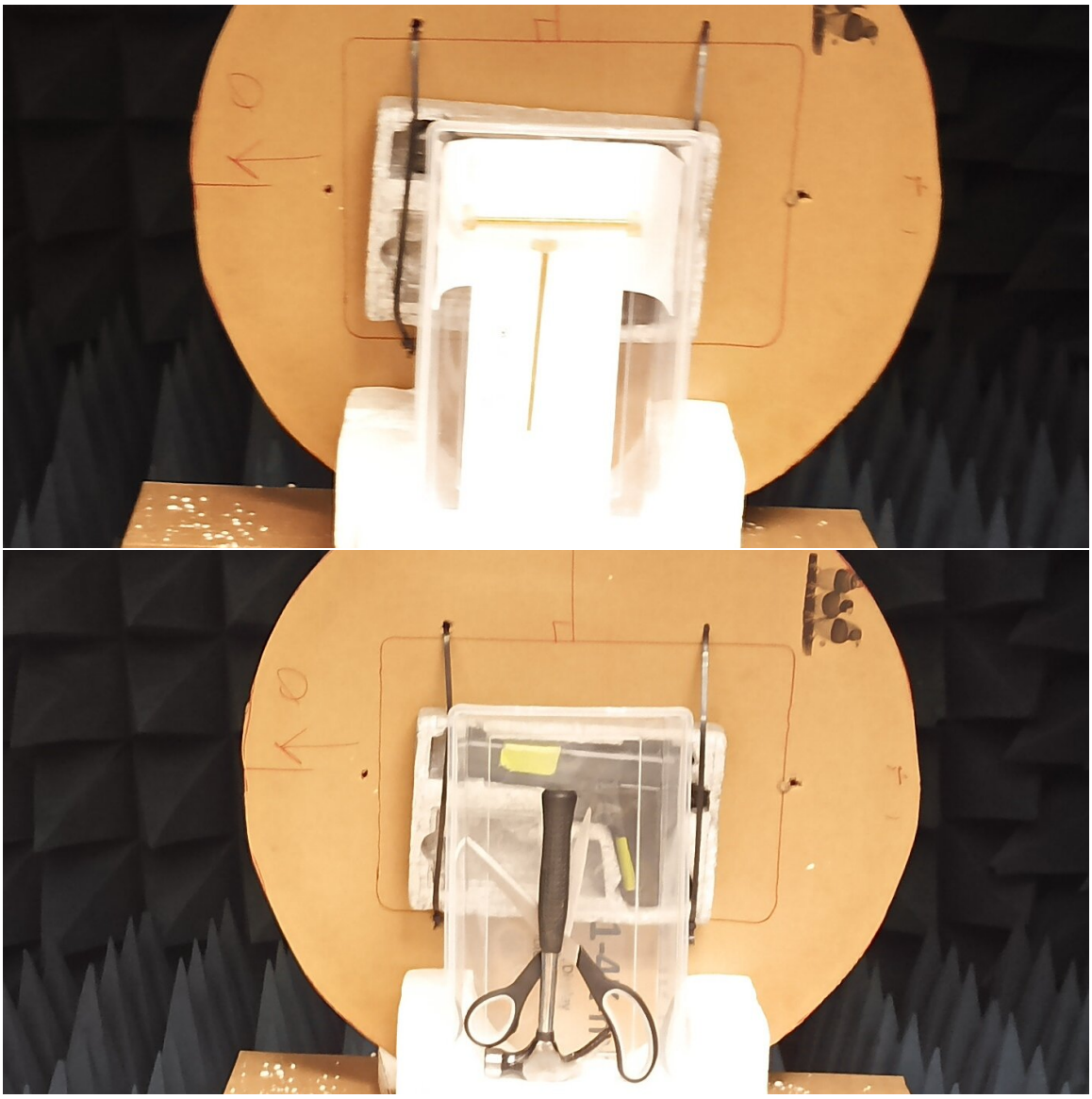

Figure 4.17: The light and heavy clutter scenarios used in the experiments. The facsimile of the Colt 1911 pistol is visible behind the heavy clutter.

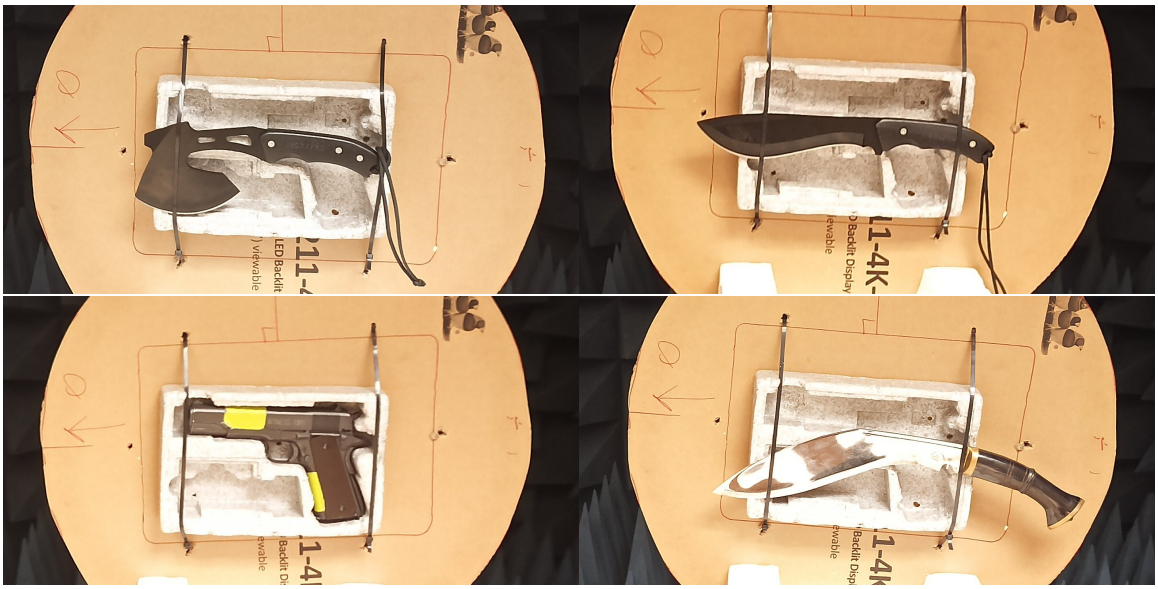

Figure 4.18: The hatchet, knife, facsimile of a Colt 1911 pistol, and kukri used in the experiments.

entries. That is, the diagonal entries of the table are much smaller than the other entries in the corresponding row. In the cases of light and heavy clutter in Table 4.1, the proposed classifier continues to correctly classify the targets. However, the confidence in that classification is lower than the confidence in the no-clutter experiments. When the hatchet is measured in light clutter, the divergences from the database entries of both the hatchet and kukri are close to zero. (Recall from Footnote 2 that the approximate KL divergence may be slightly negative.) For the Colt in light clutter, we can be confident that it is unlikely to be the knife or the kukri, because the divergence from the database entries for those weapons is more than double that from the database entry for the Colt. However, the divergence from the database entry for the hatchet is only 26 % larger than the divergence to the database entry for the Colt. Hence, we are unable to be as confident that the measured weapon is not the

hatchet. When measured in heavy clutter, the Bowie knife can be confidently distinguished from the Colt, but the divergences from the database entries for the hatchet and kukri are only 12 % and 14 % larger than the divergence from the database entry for the knife. In contrast, the kukri can be confidently classified in both light and heavy clutter, as well as in the no-clutter scenario. From Tables 4.2 and 4.3, we observe that the proposed classification method continues to provide good performance even as the orientation of the target is rolled, although some classification errors now arise. In the case of the positive roll with light clutter, and all three cases of the negative roll, one of the four targets was misclassified. (The number of successful classifications is listed in the heading of each sub-table in the tables.) Looking more closely at Tables 4.2 and 4.3, we can see that the kukri is correctly classified in all of the cases. However, in many cases, that classification comes with reduced confidence when compared with the results for the 0° roll. In particular, in the case of the positive roll and heavy clutter, the divergence of the measurement from the database entry for the hatchet is almost the same as the divergence from the database entry for the kukri. The Colt can be confidently identified in all cases bar the case of the positive roll in light clutter, where it is misidentified as a hatchet. However, the correct identification has the second smallest divergence, and that divergence is only 11 % larger than the minimum. Interestingly, if we look at the results for the zero-degree roll in Table 4.1, there is an indication of the potential for this misclassification. In the light clutter results for testing the Colt in that table, the hatchet is the second choice of classification, and has a divergence that is only 26 % larger than that for the correct entry. The hatchet is correctly identified in all but one case in Tables 4.2 and 4.3, namely the case of a negative roll with no clutter, where it is

misidentified as a kukri. However, the divergence for the correct identification is the second smallest, and is also quite small. There is an indication of the potential for this misclassification in the light clutter results for the zero-degree roll case, where the divergence from the data base entry for the kukri is very similar to that for the hatchet. Finally, we observe that while the knife is correctly classified in four of the six cases in Tables 4.2 and 4.3, it is misclassified as a hatchet in the light clutter case of the negative roll, and is misclassified as a kukri in the heavy clutter case of the negative roll. In the light clutter case, the divergence from the database entry for the knife is the second smallest and is only 12 % larger than the smallest divergence. In the heavy clutter case, the divergence from the database entry for the knife is the third smallest, and is 24 % larger than the smallest. However, in this case, the divergences from all of the entries in the database are quite small, which suggests that in this instance, the classification problem may be inherently difficult. It is also worth noting that the knife is the most linear of the tested targets, and hence it is expected to be the most sensitive to rotation. Overall, the results in Tables 4.1–4.3 indicate that the proposed classification approach is quite robust to clutter; it can distinguish between targets with considerable efficacy even when significant clutter is placed in front of them. However, the results also demonstrate that the proposed approach has some sensitivity to the orientation of the target. As such, an avenue for future improvement might be to construct a database that includes KDEs for multiple orientations of each known target, rather than the single orientation case that was considered here.

Table 4.1: KL Divergence with Same Orientation

No Clutter – 4/4				
Target	Hatchet	Knife	Colt	Kukri
Hatchet	**0.512**	87.572	70.370	40.973
Knife	215.167	**0.482**	169.418	79.114
Colt	46.593	158.839	**0.310**	40.385
Kukri	73.784	49.013	89.511	**0.142**
Light Clutter – 4/4				
Target	Hatchet	Knife	Colt	Kukri
Hatchet	**−1.492**	48.672	43.883	−1.151
Knife	176.171	**16.084**	123.037	110.133
Colt	69.369	132.822	**54.961**	113.312
Kukri	101.482	74.651	105.581	**16.256**
Heavy Clutter – 4/4				
Target	Hatchet	Knife	Colt	Kukri
Hatchet	**37.429**	120.853	38.609	83.929
Knife	63.159	**56.395**	89.551	64.272
Colt	70.812	163.967	**34.482**	108.720
Kukri	47.840	133.298	65.818	**26.275**

Table 4.2: KL Divergence with +22.5° Roll Angle

No Clutter – 4/4				
Target	Hatchet	Knife	Colt	Kukri
Hatchet	**44.380**	82.009	120.933	98.800
Knife	141.184	**−7.326**	119.417	87.825
Colt	30.542	79.060	**26.532**	30.609
Kukri	9.512	45.787	27.624	**−2.307**
Light Clutter – 3/4				
Target	Hatchet	Knife	Colt	Kukri
Hatchet	**63.343**	69.292	91.859	88.989
Knife	146.364	**37.885**	118.713	87.427
Colt	**41.946**	82.383	46.027	68.325
Kukri	42.524	38.421	59.407	**20.481**
Heavy Clutter – 4/4				
Target	Hatchet	Knife	Colt	Kukri
Hatchet	**7.242**	130.850	60.660	66.424
Knife	194.733	**57.539**	132.415	139.459
Colt	63.328	147.296	**22.539**	74.927
Kukri	45.719	56.475	101.907	**45.710**

Table 4.3: KL Divergence with $-22.5°$ Roll Angle

No Clutter – 3/4				
Target	Hatchet	Knife	Colt	Kukri
Hatchet	31.005	48.627	84.465	**18.791**
Knife	134.828	**46.033**	160.569	67.198
Colt	45.720	123.301	**8.092**	43.740
Kukri	27.975	94.859	48.877	**18.021**
Light Clutter – 3/4				
Target	Hatchet	Knife	Colt	Kukri
Hatchet	**12.691**	64.664	37.594	28.741
Knife	**30.772**	34.446	69.440	51.409
Colt	65.812	105.278	**20.912**	82.437
Kukri	12.396	23.976	30.016	**10.852**
Heavy Clutter – 3/4				
Target	Hatchet	Knife	Colt	Kukri
Hatchet	**74.482**	183.984	77.598	139.008
Knife	48.921	56.522	71.303	**45.597**
Colt	60.596	120.339	**24.702**	86.565
Kukri	53.957	138.943	105.626	**33.340**

Table 4.4: Confusion Matrix (%) with Same Orientation

No Clutter – 4/4				
Target	Hatchet	Knife	Colt	Kukri
Hatchet	**100.0**	0.0	0.0	0.0
Knife	0.0	**100.0**	0.0	0.0
Colt	0.0	0.0	**100.0**	0.0
Kukri	0.0	0.0	0.0	**100.0**
Light Clutter – 4/4				
Target	Hatchet	Knife	Colt	Kukri
Hatchet	**55.2**	0.0	0.0	44.8
Knife	0.0	**100.0**	0.0	0.0
Colt	0.0	0.0	**100.0**	0.0
Kukri	0.0	0.0	0.0	**100.0**
Heavy Clutter – 4/4				
Target	Hatchet	Knife	Colt	Kukri
Hatchet	**60.2**	0.0	39.8	0.0
Knife	0.9	**98.9**	0.0	0.2
Colt	0.0	0.0	**100.0**	0.0
Kukri	0.0	0.0	0.0	**100.0**

Table 4.5: Confusion Matrix (%) with +22.5° Roll Angle

No Clutter – 4/4				
Target	Hatchet	Knife	Colt	Kukri
Hatchet	**100.0**	0.0	0.0	0.0
Knife	0.0	**100.0**	0.0	0.0
Colt	3.0	0.0	**96.0**	1.0
Kukri	0.3	0.0	0.0	**99.7**
Light Clutter – 3/4				
Target	Hatchet	Knife	Colt	Kukri
Hatchet	**78.6**	21.4	0.0	0.0
Knife	0.0	**100.0**	0.0	0.0
Colt	**93.4**	0.0	6.6	0.0
Kukri	0.0	0.0	0.0	**100.0**
Heavy Clutter – 3/4				
Target	Hatchet	Knife	Colt	Kukri
Hatchet	**100.0**	0.0	0.0	0.0
Knife	0.0	**100.0**	0.0	0.0
Colt	0.0	0.0	**100.0**	0.0
Kukri	**58.2**	0.2	0.0	41.6

Table 4.6: Confusion Matrix (%) with $-22.5°$ Roll Angle

No Clutter – 3/4				
Target	Hatchet	Knife	Colt	Kukri
Hatchet	0.0	0.0	0.0	**100.0**
Knife	0.0	**100.0**	0.0	0.0
Colt	0.0	0.0	**100.0**	0.0
Kukri	0.0	0.0	0.0	**100.0**
Light Clutter – 3/4				
Target	Hatchet	Knife	Colt	Kukri
Hatchet	**100.0**	0.0	0.0	0.0
Knife	**73.7**	26.3	0.0	0.0
Colt	0.0	0.0	**100.0**	0.0
Kukri	31.5	0.0	0.0	**68.5**
Heavy Clutter – 3/4				
Target	Hatchet	Knife	Colt	Kukri
Hatchet	**76.6**	0.0	23.4	0.0
Knife	16.3	0.0	0.0	**83.7**
Colt	0.0	0.0	**100.0**	0.0
Kukri	0.0	0.0	0.0	**100.0**

4.7.2 Performance Over Multiple Sets of Measurement Shots

In order to examine the statistical performance of the proposed method, we repeated the single-measurement-set analysis described in the previous section 1000 times. Each time, a different random subset of 100 measurement shots was taken from the set of the 200 available shots and the database was constructed from the other 100 "clean" measurement shots at the 0° roll angle. The classification result (i.e., the entry of the database with the smallest divergence) was recorded for each realization. These results were collated into the confusion matrices presented in Tables 4.4, 4.5 and 4.6, which are organized in the same way as Tables 4.1, 4.2 and 4.3. Each entry in each confusion matrix indicates the fraction of trials that yielded a particular classification result, presented as a percentage. Each row represents a particular target (which is treated as being unknown), and the four percentages in that row indicate how often it was classified as a particular database target. The most frequent classification result is highlighted in bold. The number of targets for which the most frequent classification is the correct classification is listed in the heading of each subtable. These statistical results align well with the outcomes of the single-measurement-set results in the previous section, which suggests that those single-measurement-set results are typical. In cases for which the statistical classification results are consistent, there is one divergence in the corresponding single-measurement-set results that is significantly smaller than the other divergences in that row. In the cases where the statistical classification is consistently correct, the smallest divergence corresponds to the correct target. However, there are cases where the statistical classification results are consistently incorrect. This is also reflected in the single-measurement-set results, where the smallest divergence corresponds to an incorrect classification; see

the $+22.5°$-roll Colt in light clutter and the $-22.5°$-roll hatchet in the absence of clutter. In cases for which the statistical classification results alternate between two decisions in different realizations, the two smallest divergences for the corresponding single-measurement-set results are of similar size. An example of that is the $0°$-roll hatchet in both light and heavy clutter. The insight that can be gleaned from the statistical results in this section is in agreement with the observations made in the previous section. In particular, the proposed method demonstrates considerable robustness to the rather difficult clutter environments that we have considered, but it can be somewhat sensitive to orientation.

4.8 Conclusion

A promising method for classifying targets based on measurements of their late time responses was proposed. Rather than using direct estimates of the resonance parameters as the classification features, or using estimates of the system poles that capture the information in the resonance parameters, the proposed approach uses the distribution of the pole estimates as the classification feature. An approximation of the Kullback-Leibler divergence that is tailored to the proposed application is employed as the classification metric. That metric incorporates practical considerations regarding the radar, and helps to mitigate the impact of environmental clutter. In physical experiments, the proposed approach was demonstrated to be remarkably robust to environmental clutter, but it did exhibit some sensitivity to target orientation.

Chapter 5

Conclusion and Future Work

The availability of a radar capable of a high rate of repeated measurements has opened up an avenue of exploration involving statistical analysis of relatively large sets of measurements shots of the late-time responses (LTRs) of radar targets. Three approaches for estimating the resonance parameters of a target in such a setting have been presented. These parameters correspond to the poles of a particular Padé approximant of the z-transform of a finite-length window of the target's sampled LTR. These poles can be highly perturbed even if the LTR is only slightly perturbed by noise, making their estimation challenging in low-SNR scenarios. For this reason, probability distributions of the poles (and of closely related features) were studied.

In the first proposed method, the poles are estimated via distribution matching of the roots of the z-transform of the sampled late-time response. The zeros and poles are known to "repel" one another, thus their distributions are closely related. However, a closed-form expression for the distribution of these zeros is known whereas such a distribution for the poles themselves is not. The developed method was demonstrated to be effective on synthetic signals in low-SNR conditions. Moreover, it was

demonstrated to work even in the cases of sampling jitter and synchronization errors.

In the second proposed method, the poles were estimated by using an analysis of an empirical approximation of their distribution to initialize a constrained nonlinear least-squares optimizer. In this way, a good solution to the non-convex optimization problem can usually be found without resorting to global optimization. By applying this resonance estimation method to environment background signals first and then introducing the target, the method is able to accurately estimate the poles of a target separately from its background. This enabled its application in physical experiments with a prototype radar, because measurement shots from such experiments inherently have a background component. The method was shown to be effective at estimating resonance parameters in such cases, even when the targets are quite cluttered.

The third contribution of the thesis directly addresses the problem of target classification, as distinct from the problem of explicitly estimating the resonance parameters. A feature of this contribution is that aggregate estimates of individual resonance parameters are eschewed in favour of studying the empirical distributions of such estimates. Those empirical distributions are obtained by processing the estimates obtained from multiple individual measurements shots. Classification was performed by comparing the distribution of an unknown target against each member of a set of distributions stored in a database of known targets, using a modified version of the Kullback-Leibler (KL) divergence that incorporates knowledge of the background clutter and the bandwidth of radar. This was shown to be a natural generalization of Bayesian binary-hypothesis testing. The proposed classification method was shown to be effective, remarkably robust to clutter, and reasonably robust to changes in

target orientation (aspect). Indeed, to the best of the author's knowledge, the experimental results in Chapter 4 represent the first example of a practical resonance-based classification scheme that is effective in scenarios that involve a somewhat realistic (and perhaps even excessive) degree of clutter.

Nonetheless, there are numerous aspects of the analysis of the resonances in the LTRs acquired from multiple measurement shots that have yet to be explored. The most compelling practical avenue of exploration involves testing the efficacy of the classification method in more realistic kinds of clutter, such as having the target in a bag or on the body of a person.

In the latter case, having the person move introduces a new set of challenges. For example, in the results presented here, each analyzed measurement shot is actually the average of 256 physical measurement shots, which slows acquisition to a point that it takes a few seconds to acquire a set of 100 measurements. However, the radar can operate with averages of just 8 or 16 shots.[1] Thus it should be possible to measure 16 to 32 times faster for the purposes of measuring moving targets without having to account for motion blur. It is not clear exactly what effect the reduced fidelity of the measurement shots would have, be it purely from the decrease in SNR or some other aberrations that the averaging is masking, and it is not clear how the algorithms would have to be adjusted. Such a scenario also has the inherent challenge of having to estimate the LTR window, whereas in the experiments studied here, this was controlled.

It was noted in Chapter 4 that the method is not especially tolerant to changes in orientation, which is a common problem with these techniques. However, it is not

[1] In principle, it can operate without averaging, but offloading and processing such a great volume of data is both very difficult and likely of less value due to the low SNR.

clear how many orientations it is necessary to store in a database, it is not clear what impact different tilt angles might have because only roll angles were studied, and is it not clear whether all of these measurement shots should be clumped together in a single empirical distribution, treated separately, or some combination of the two.

Having polarization diversity proved to be critically important throughout the analysis of the physical experiments. This is most apparent for linear targets, such as the brass rods studied in Chapter 3, but it applies to more complex targets as well. However, only co-pol (HH and VV) responses were analyzed in this thesis, whereas the radar also records cross-pol (HV and VH) responses, which should also contain useful information. In a similar vein, in the bistatic setup that was utilized, the transmitting antennas were collocated as were the receiving antennas, which was done purely for pragmatic reasons. It may be worth exploring alternative configurations.

It has been noted throughout the thesis that having a multi-path return within the LTR window is detrimental. In the experimental setup in this thesis, the first such return comes from a reflection from the floor, forcing the use of an LTR window that is around 2.4 ns in duration. Light travels 72 cm in that time. This suggests that the (partly) anechoic chamber behind the target is not actually needed, but rather that it is only important to have the area up to 36 cm behind the target clear. Demonstrating this would would provide valuable guidance for future implementations in other scenarios, but that is impractical with the current configuration of the laboratory space.

Bibliography

[1] S. Abrahamson, B. Brusmark, G. C. Gaunaurd, and H. C. Strifors, "Target identification by means of impulse radar," in *Automatic Object Recognition*, SPIE, vol. 1471, 1991, pp. 130–141.

[2] J. Accardo and M. A. Chaudhry, "Radiation exposure and privacy concerns surrounding full-body scanners in airports," *Journal of Radiation Research and Applied Sciences*, vol. 7, no. 2, pp. 198–200, 2014.

[3] A. Agurto, Y. Li, G. Y. Tian, N. Bowring, and S. Lockwood, "A review of concealed weapon detection and research in perspective," in *Proc. IEEE Int. Conf. Networking, Sensing, Control*, London, United Kingdom, 2007, pp. 443–448.

[4] R. Appleby and R. N. Anderton, "Millimeter-wave and submillimeter-wave imaging for security and surveillance," *Proc. IEEE*, vol. 95, no. 8, pp. 1683–1690, 2007.

[5] N. Asncombe, "No place to hide [millimeter wave imaging]," *IEE Review*, vol. 51, no. 12, pp. 26–30, 2005.

[6] C. Baard and M. Georgiev, "Methods for detecting target in range and system-level noise estimation," *Electromagn. Vision Res. Lab.*, McMaster University, Hamilton, ON, Tech. Rep. EMVi-R-93, Apr. 2020.

[7] G. A. Baker and P. Graves-Morris, *Padé Approximants*, 2nd ed. Cambridge University Press, 1996.

[8] P. Barone, "A new transform for solving the noisy complex exponentials approximation problem," *J. Approx. Theory*, vol. 155, no. 1, pp. 1–27, 2008.

[9] P. Barone, "On the condensed density of the generalized eigenvalues of pencils of Gaussian random matrices and applications," *J. Multivariate Analysis*, vol. 111, pp. 160–173, 2012.

[10] P. Barone and A. Ramponi, "A new estimation method in modal analysis," *IEEE Trans. Signal Process.*, vol. 48, no. 4, pp. 1002–1014, 2000.

[11] C. E. Baum, "The singularity expansion method," in *Transient Electromagnetic Fields*, L. B. Felsen, Ed., Springer, 1976, pp. 129–179.

[12] C. E. Baum, E. J. Rothwell, K.-M. Chen, and D. P. Nyquist, "The singularity expansion method and its application to target identification," *Proc. IEEE*, vol. 79, no. 10, pp. 1481–1492, 1991.

[13] J. Beirlant, E. J. Dudewicz, L. Györfi, and E. C. Van der Meulen, "Nonparametric entropy estimation: An overview," *Int. J. Math. Statist. Sci.*, vol. 6, no. 1, pp. 17–39, 1997.

[14] D. Belkić, P. Dando, J. Main, and H. Taylor, "Three novel high-resolution nonlinear methods for fast signal processing," *J. Chem. Phys.*, vol. 113, no. 16, pp. 6542–6556, 2000.

[15] M. Bell, "Information theory and radar waveform design," *IEEE Trans. Information Theory*, vol. 39, no. 5, pp. 1578–1597, 1993.

[16] C. Bennett and G. Ross, "Time-domain electromagnetics and its applications," *Proc. IEEE*, vol. 66, no. 3, pp. 299–318, 1978.

[17] A. J. Berni, "Target identification by natural resonance estimation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-11, no. 2, pp. 147–154, 1975.

[18] P. Blacksmith, R. Hiatt, and R. Mack, "Introduction to radar cross-section measurements," *Proceedings of the IEEE*, vol. 53, no. 8, pp. 901–920, 1965.

[19] F. Bonizzoni, F. Nobile, I. Perugia, and D. Pradovera, "Fast least-squares Padé approximation of problems with normal operators and meromorphic structure," *Math. Comput.*, vol. 89, no. 323, pp. 1229–1257, 2020.

[20] Y. Bresler and A. Macovski, "Exact maximum likelihood parameter estimation of superimposed exponential signals in noise," *IEEE Trans. Acoustics, Speech, Signal Process.*, vol. 34, no. 5, pp. 1081–1089, 1986.

[21] C.-C. Chen and L. Peters, "Buried unexploded ordnance identification via complex natural resonances," *IEEE Trans. Antennas and Propag.*, vol. 45, no. 11, pp. 1645–1654, 1997.

[22] K.-M. Chen, D. Nyquist, E. Rothwell, L. Webb, and B. Drachman, "Radar target discrimination by convolution of radar return with extinction-pulses and single-mode extraction signals," *IEEE Trans. Antennas Propag.*, vol. 34, no. 7, pp. 896–904, 1986.

[23] W. C. Chen and N. V. Z. Shuley, "Robust target identification in white Gaussian noise using canonical correlation analysis," *IEEE Trans. Antennas Propag.*, vol. 60, no. 7, pp. 3533–3537, 2012.

[24] C. Chuang and D. Moffatt, "Natural resonances of radar targets via Prony's method and target discrimination," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-12, no. 5, pp. 583–589, 1976.

[25] C. Chui and A. Chan, "A two-sided rational approximation method for recursive digital filtering," *IEEE Trans. Acoustics, Speech, Signal Process.*, vol. 27, no. 2, pp. 141–145, 1979.

[26] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Wiley, 2005.

[27] *FEKO*, `https://www.altair.com/feko/`, version 2023.1, Altair Engineering Inc.

[28] T. Feng, J. Yang, P. Ge, S. Zhang, Y. Du, and H. Wang, "An improved pole extraction algorithm with low signal-to-noise ratio," in *2021 IEEE 4th Advanced Information Management, Communicates, Electronic and Automation Control Conference*, IEEE, vol. 4, 2021, pp. 1146–1150.

[29] R. Fletcher, *Practical Methods of Optimization*, 2nd ed. Wiley, 1987.

[30] R. Fontana, "Recent system applications of short-pulse ultra-wideband (UWB) technology," *IEEE Trans. Microw. Theory Techn.*, vol. 52, no. 9, pp. 2087–2104, 2004.

[31] R. Fontana, "Recent system applications of short-pulse ultra-wideband (UWB) technology," *IEEE Trans. Microw. Theory Tech.*, vol. 52, no. 9, pp. 2087–2104, 2004.

[32] M. Froissart, "Approximation de pade application à la physique des particules élémentaires," *Les rencontres physiciens-mathématiciens de Strasbourg-RCP25*, vol. 16, pp. 1–13, 1973.

[33] M. Georgiev, "Interference filtering and correction of systemic timebase distortion in an equivalent-sampling receiver," *Electromagn. Vision Res. Lab.*, McMaster University, Hamilton, ON, Tech. Rep. EMVi-R-96, Jun. 2020.

[34] M. Georgiev and T. N. Davidson, "Estimating resonances in low-SNR late-time radar returns with sampling jitter," *IEEE Trans. Signal Process.*, vol. 72, pp. 4651–4665, 2024.

[35] M. Georgiev and J. Nguyen, "System noise and signal-to-noise ratio analysis of the UWB equivalent-time-sampling radar," *Electromagn. Vision Res. Lab.*, McMaster University, Hamilton, ON, Tech. Rep. EMVi-R-104, Nov. 2020.

[36] M. Georgiev, A. D. Pitcher, and T. N. Davidson, *Multi-shot estimation of resonance parameters of late-time radar returns in clutter*, Submitted, Dec. 2024.

[37] M. Georgiev, P. Rice, and J.-K. Zhang, "Novel noise-tolerant method for extracting target resonances using pulse radar," in *Proc. 17th European Radar Conf.*, 2021, pp. 258–261.

[38] M. S. Georgiev, A. D. Pitcher, and T. N. Davidson, "Classification of radar targets via distribution matching of late-time resonance parameters," *IEEE Trans. Radar Syst.*, vol. 3, pp. 645–655, 2025.

[39] P. Gonnet, S. Guttel, and L. N. Trefethen, "Robust Padé approximation via SVD," *SIAM Rev.*, vol. 55, no. 1, pp. 101–117, 2013.

[40] J. M. Hammersley, "The zeros of a random polynomial," in *Proc. Third Berkeley Symp. Prob. and Stats*, vol. 2, 1956, pp. 89–111.

[41] C. Hargrave, A. Abbosh, N. Shuley, and V. Clarkson, "Radar target identification by histogram analysis of resonant frequencies," in *Proc. 7th European Conf. Antennas Propag.*, 2013, pp. 3604–3608.

[42] C. O. Hargrave, I. V. L. Clarkson, and H.-S. Lui, "Late-time estimation for resonance-based radar target identification," *IEEE Trans. Antennas Propag.*, vol. 62, no. 11, pp. 5865–5871, 2014.

[43] S. S. Haykin, *Adaptive Filter Theory*, 3rd ed. Prentice Hall, 1996.

[44] W. H. Hayt Jr and J. A. Buck, *Engineering Electromagnetics*, 8th ed. McGraw Hill, 2012.

[45] E. Heyman and L. Felsen, "A wavefront interpretation of the singularity expansion method," *IEEE Trans. Antennas Propag.*, vol. 33, no. 7, pp. 706–718, 1985.

[46] E. Heyman and L. Felsen, "Creeping waves and resonances in transient scattering by smooth convex objects," *IEEE Trans. Antennas Propag.*, vol. 31, no. 3, pp. 426–437, 1983.

[47] K. C. Ho, L. Carin, P. D. Gader, and J. N. Wilson, "An investigation of using the spectral characteristics from ground penetrating radar for landmine/clutter discrimination," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 4, pp. 1177–1191, 2008.

[48] Y. Hua and T. Sarkar, "Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 5, pp. 814–824, 1990.

[49] Y. Hua and T. K. Sarkar, "Generalized pencil-of-function method for extracting poles of an EM system from its transient response," *IEEE Trans. Ant. Propag.*, vol. 37, no. 2, pp. 229–234, 1989.

[50] Y. Hua and T. K. Sarkar, "Matrix pencil method and its performance," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Process.*, 1988, pp. 2476–2477.

[51] Y. Hua and T. K. Sarkar, "Perturbation analysis of TK method for harmonic retrieval problems," *IEEE Trans. Acoustics, Speech, Signal Process.*, vol. 36, no. 2, pp. 228–240, 1988.

[52] P. Ilavarasan, J. Ross, E. Rothwell, K.-M. Chen, and D. Nyquist, "Performance of an automated radar target discrimination scheme using E pulses and S pulses," *IEEE Trans. Antennas Propag.*, vol. 41, no. 5, pp. 582–588, 1993.

[53] S. Kotz, N. Balakrishnan, and N. L. Johnson, *Continuous Multivariate Distributions, Volume 1: Models and Applications*. Wiley, 2004.

[54] H. Krim and M. Viberg, "Two decades of array signal processing research: The parametric approach," *IEEE Signal Process. Mag.*, vol. 13, no. 4, pp. 67–94, 1996.

[55] D. A. Ksienski, "Pole and residue extraction from measured data in the frequency domain using multiple data sets," *Radio Science*, vol. 20, no. 01, pp. 13–19, 1985.

[56] R. Kumaresan and D. Tufts, "Estimating the parameters of exponentially damped sinusoids and pole-zero modeling in noise," *IEEE Trans. Acoustics, Speech, Signal Process.*, vol. 30, no. 6, pp. 833–840, 1982.

[57] H.-S. Lui and N. V. Shuley, "Evolutions of partial and global resonances in transient electromagnetic scattering," *IEEE Antennas Propag. Letters*, vol. 7, pp. 436–439, 2008.

[58] D. J. MacKay, *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, 2003.

[59] P. C. Mahalanobis et al., "On the generalised distance in statistics," in *Proc. Nat. Inst. Sci. India*, vol. 2, 1936, pp. 49–55.

[60] D. McMakin, D. Sheen, T. Hall, J. Tedeschi, and A. M. Jones, "New improvements to millimeter-wave body scanners," in *Proc. 8th Int. Conf. & Exh. 3D Body Scanning & Process. Tech.*, Montreal QC, Canada, 2017, pp. 263–271.

[61] O. Mironenko, "Body scanners versus privacy and data protection," *Computer Law & Security Review*, vol. 27, no. 3, pp. 232–244, 2011.

[62] J. E. Mooney, Z. Ding, and L. S. Riggs, "Robust target identification in white Gaussian noise for ultra wide-band radar systems," *IEEE Trans. Antennas Propag.*, vol. 46, no. 12, pp. 1817–1823, 1998.

[63] J. Nguyen, M. Georgiev, and N. K. Nikolova, "Stage 3 technical progress report to DASA," *Electromagn. Vision Res. Lab.*, McMaster University, Hamilton, ON, Tech. Rep. EMVi-R-108, Apr. 2021.

[64] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed. Springer, 2006.

[65] F. Pérez-Cruz, "Kullback-Leibler divergence estimation of continuous distributions," in *Proc. IEEE Int. Symp. Information Theory*, 2008, pp. 1666–1670.

[66] A. D. Pitcher, C. W. Baard, and N. K. Nikolova, "Design and performance analysis of a picosecond pulse generator," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–14, 2022.

[67] A. D. Pitcher, M. Georgiev, and N. K. Nikolova, "Correcting timebase errors in ultra-wideband equivalent-time sampling receivers," in *Proc. 21st European Radar Conf.*, Paris, 2024, pp. 47–50.

[68] A. D. Pitcher, M. Georgiev, N. K. Nikolova, and N. Nicolici, "Parallelized field-programmable gate array data processing for high-throughput pulsed-radar systems," *Sensors*, vol. 25, no. 1, p. 239, 2025.

[69] A. D. Pitcher and M. Georgiev, "Equivalent-time sampling ultra-wideband pulsed-radar receiver: RF front-end design," *Electromagn. Vision Res. Lab.*, McMaster University, Hamilton, ON, Tech. Rep. EMVi-R-122, Sep. 2024.

[70] A. D. Pitcher and M. Georgiev, "The digital signal processing pipelines (DSPPL) suite," *Electromagn. Vision Res. Lab.*, McMaster University, Hamilton, ON, Tech. Rep. EMVi-R-121, Sep. 2024.

[71] A. D. Pitcher, M. Georgiev, J. Nguyen, and N. K. Nikolova, "Unobtrusive inspection for on-body threats concealed under clothing: Concealed weapon detection using polarization energy ratios," *Electromagn. Vision Res. Lab.*, McMaster University, Hamilton, ON, Tech. Rep. EMVi-R-110, Sep. 2021.

[72] B. Porat and B. Friedlander, "A modification of the Kumaresan-Tufts methods for estimating rational impulse responses," *IEEE Trans. Acoustics, Speech, and Signal Process.*, vol. 34, no. 5, pp. 1336–1338, 1986.

[73] F. Rellich, *Perturbation Theory of Eigenvalue Problems*. CRC Press, 1969.

[74] R. Rezaiesarlak and M. Manteghi, "On the application of short-time matrix pencil method for wideband scattering from resonant structures," *IEEE Trans. Antennas Propag.*, vol. 63, no. 1, pp. 328–335, 2015.

[75] E. Rothwell, D. Nyquist, K.-M. Chen, and B. Drachman, "Radar target discrimination using the extinction-pulse technique," *IEEE Trans. Antennas Propag.*, vol. 33, no. 9, pp. 929–937, 1985.

[76] T. K. Sarkar, S. Park, J. Koh, and S. M. Rao, "Application of the matrix pencil method for estimating the SEM (Singularity Expansion Method) poles of source-free transient responses from multiple look directions," *IEEE Trans. Antennas Propag.*, vol. 48, no. 4, pp. 612–618, 2000.

[77] R. Schober and W. H. Gerstacker, "The zeros of random polynomials: Further results and applications," *IEEE Trans. Commun.*, vol. 50, no. 6, pp. 892–896, 2002.

[78] D. M. Sheen, D. L. McMakin, and T. E. Hall, "Three-dimensional millimeter-wave imaging for concealed weapon detection," *IEEE Trans. Microw. Theory Tech.*, vol. 49, no. 9, pp. 1581–1592, 2001.

[79] R. Storn and K. Price, "Differential evolution–A simple and efficient heuristic for global optimization over continuous spaces," *J. Global Opt.*, vol. 11, pp. 341–359, 1997.

[80] W. Sun, K.-M. Chen, D. P. Nyquist, and E. J. Rothwell, "Determination of the natural modes for a rectangular plate (transient scattering)," *IEEE Trans. Antennas Propag.*, vol. 38, no. 5, pp. 643–652, 1990.

[81] S. S. Swords, *Technical History of the Beginnings of RADAR*. The Institution of Engineering and Technology, 1986.

[82] J. D. Taylor, *Introduction to Ultra-Wideband Radar Systems*. CRC press, 1995.

[83] G. R. Terrell and D. W. Scott, "Variable kernel density estimation," *Ann. Statist.*, pp. 1236–1265, 1992.

[84] S. Umesh and D. W. Tufts, "Estimation of parameters of exponentially damped sinusoids using fast maximum likelihood estimation with application to NMR spectroscopy data," *IEEE Trans. Signal Process.*, vol. 44, no. 9, pp. 2245–2259, 1996.

[85] J. A. C. Weideman, "Computation of the complex error function," *SIAM J. Numer. Anal.*, vol. 31, no. 5, pp. 1497–1518, 1994.

[86] L. Weiss and R. McDonough, "Prony's method, z-transforms, and Padé approximation," *SIAM Review*, vol. 5, no. 2, pp. 145–149, 1963.

[87] K. Xu, G. Marrelec, S. Bernard, and Q. Grimal, "Lorentzian-model-based Bayesian analysis for automated estimation of attenuated resonance spectrum," *IEEE Trans. Signal Process.*, vol. 67, no. 1, pp. 4–16, 2018.