Information-Theoretic Compressive Sensing Kernel Optimization and Bayesian Cramér–Rao Bound for Time Delay Estimation

Yujie Gu, *Senior Member, IEEE*, and Nathan A. Goodman, *Senior Member, IEEE*

***Abstract*—With the adoption of arbitrary and increasingly wideband signals, the design of modern radar systems continues to be limited by analog-to-digital converter technology and data throughput bottlenecks. Meanwhile, compressive sensing (CS) promises to reduce sampling rates below the Nyquist rate for some applications by constraining the set of possible signals. In many practical applications, detailed prior knowledge on the signals of interest can be learned from training data, existing track information, and/or other sources, which can be used to design better compressive measurement kernels. In this paper, we use an information-theoretic approach to optimize CS kernels for time delay estimation. The measurements are modeled via a Gaussian mixture model by discretizing the *a priori* probability distribution of the time delay. The optimal CS kernel that approximately maximizes the Shannon mutual information between the measurements and the time delay is then found by a gradient-based search. Furthermore, we also derive the Bayesian Cramér–Rao bound (CRB) for time delay estimation as a function of the CS kernel. In numerical simulations, we compare the performance of the proposed optimal sensing kernels to random projections and the Bayesian CRB. Simulation results demonstrate that the proposed technique for sensing kernel optimization can significantly improve performance, which is consistent with the Bayesian CRB versus signal-to-noise ratio (SNR). Finally, we use the Bayesian CRB expressions and simulation results to make conclusions about the usefulness of CS in radar applications. Specifically, we discuss CS SNR loss versus resolution improvement in SNR- and resolution-limited scenarios.**

***Index Terms*—Compressive sensing (CS), Cramér–Rao bound (CRB), Gaussian mixture (GM), minimum mean-square error (MMSE), mutual information, sensing kernel optimization, time delay estimation.**

I. Introduction

TIME delay estimation is a fundamental issue in many signal processing applications, such as radar and/or sonar ranging, radiolocation, geolocation, remote sensing, synchronization, and other areas (see, for example, [1]–[6], and the references therein). Moreover, it is a canonical nonlinear estimation problem that can be useful for understanding and demonstrating fundamental concepts. Waveform bandwidth is an essential factor in determining estimation accuracy. Therefore, radar systems typically transmit modulated pulses, such as linear frequency modulation (LFM) waveforms, and then correlate the echo signal with the known transmit waveform via an implementation of the matched filter. Time delay can be estimated from observing the matched filter output over time. Because time delay resolution and estimation accuracy can be improved by increasing the waveform bandwidth, more wideband signals are being adopted to improve the performance of imaging and other tasks.

However, high-rate analog-to-digital (A/D) converters present significant challenges with respect to required sampling rate, system cost, and power consumption. Data throughput from the A/D converter to the processor or storage sub-system can also be a bottleneck. In most cases, the required high-rate A/D converters and data handling hardware are costly and have high power consumption. Because A/D technology can be a significant limiting factor, in this paper we compare classical Nyquist and compressive sampling schemes at a fixed A/D rate. Traditional correlation-based processing mandates that the waveform bandwidth cannot exceed the A/D rate, whereas a compressive approach allows the transmit bandwidth to be increased beyond the A/D rate. Hence, in this paper, we consider fundamental advantages and disadvantages regarding the tradeoff between resolution and signal-to-noise ratio (SNR) loss of Nyquist versus sub-Nyquist radio frequency (RF) systems.

In contrast to the classical Shannon/Nyquist sampling theory, compressive sensing (CS) theory [7] states that some signals can be acquired at sub-Nyquist sampling rates and then reconstructed with accuracy comparable to Nyquist sampling, given some additional side information about the signals being reconstructed. In the case of CS, this side information comprises knowledge of a signal basis in which all signals of interest have a sparse representation, which essentially limits the signal to a low-dimensional manifold in bandlimited signal space. Moreover, in order to solve the resulting ill-conditioned signal recovery problem, an additional constraint requires the measurement kernels to be incoherent with the signal's sparse representation basis [7], [8]. It is now well known that random measurement kernels, such as Gaussian and Bernoulli kernels, have a high

Manuscript received March 15, 2016; revised October 13, 2016 and February 18, 2017; accepted May 2, 2017. Date of publication May 18, 2017; date of current version June 28, 2017. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Mark Plumbley. This work was supported in part by the Defense Advanced Research Projects Agency under Grant #N66001-10-1-4079. (*Corresponding author: Nathan A. Goodman.*)

The authors are with the School of Electrical and Computer Engineering, Advanced Radar Research Center, University of Oklahoma, Norman, OK 73019 USA (e-mail: guyujie@hotmail.com; goodman@ou.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2017.2706187

probability of meeting this incoherence requirement and can provide accurate signal recovery when the representation basis is fixed [9]. Random projections, however, do not exploit any prior knowledge that may be available regarding the signal(s) of interest. On the one hand, this attribute makes random projections robust, applicable to a wide set of applications, and useful for important and rigorous proofs on CS behavior. On the other hand, prior signal knowledge is available in many applications and might improve system performance if properly exploited. For example, target track information can be used to initialize prior distributions on target properties including estimated target range, which then provides the possibility of optimizing CS kernels according to an appropriate sensing performance metric.

In contrast to the random sensing kernels commonly adopted in CS theory and applications [7]–[11], in this paper we develop and apply an information-theoretic metric [12]–[15] to design an optimized sensing kernel for time delay estimation in radar applications. Within the framework of information theory, time delay can be modeled as a random variable with some prior probability density function (pdf). By discretizing the *a priori* distribution of the target time delay, the measurement pdf can be approximated by a Gaussian mixture (GM) given a linear CS measurement model. The GM distributions describe the time delay and, subsequently, the accuracy of the pdf of the radar measurements can be improved by increasing the number of components in the mixture. The optimization formulation then calls for maximizing the Shannon mutual information between the compressive measurements and the target time delay.

In order to optimize the sensing kernel, we derive the approximate gradient of mutual information with respect to the sensing matrix based on a first-order Taylor series expansion. Although this approach is similar to the one we used to optimize CS kernels for radar target profiling and recognition application [16], [17], wall reflections mitigation in urban radar imaging [18], [19] and direction-of-arrival estimation in massive multiple input multiple output (MIMO) systems [20], the time delay estimation problem treated in this paper is more fundamental. The time delay estimation problem clearly demonstrates the fundamental tradeoff in compressive radar systems between resolution and SNR loss, and helps to explain the results obtained in [16]–[20]. Starting from the Cramér–Rao bound (CRB) under Nyquist sampling [21], [22], we then derive the CRB under compressive sampling and extend the applications of CRB analysis on parameter estimation.

Compared with CRB analyses in the framework of random compressive sensing [23]–[26], the CRB analyses presented in this paper clearly state the resolution/SNR loss tradeoff inherent in compressive radar systems. The minimum mean-square error (MMSE) estimator, which has a closed form for the GM approximation to the prior distribution, is used to compare performance of different compression strategies. The relationship between mutual information and MMSE, which has been shown in [27], validates the use of mutual information as an optimization metric. We also derive a Bayesian CRB on the time delay estimate that accounts for both prior information and the compressive measurement structure. We note that the Ziv–Zakai bound (ZZB) is also a Bayesian bound that can be used to evaluate the performance of time delay estimation (see, for example, [4], and references therein). The ZZB is typically applicable to estimation of uniformly-distributed scalar random variables [28]. However, the Bellini–Tartara form [29] of the ZZB can be used to describe estimation performance for an arbitrarily-distributed continuous vector random variable [30]. In this paper, with the connection between mutual information and MMSE, we have used the Bayesian CRB to judge estimator performance.

Generally speaking, radar CS suffers an SNR loss roughly proportional to the compression factor. This SNR loss can manifest as either noise folding or as a loss in signal energy [31]–[33], but the implication is the same - using radar CS for detection applications has some potential difficulties. On the other hand, radar CS can sometimes enable improved resolution without the typically associated increases in system size, weight, and power (SWAP). For example, in the application proposed here, we consider compressive sampling from the perspective of increased bandwidth of the transmit waveform without the associated increase in A/D sample rate. Through compressive sampling, we can capture the additional information inherent in a high-bandwidth waveform without the corresponding increase in system SWAP and cost that would result from a faster A/D converter. In high-SNR parameter estimation scenarios, this additional resolution can be a benefit that outweighs the inherent SNR loss. Our simulation results are consistent with these observations while also showing performance improvement of optimized sensing kernels in comparison to random sensing kernels.

This paper explores a fundamental tradeoff enabled by compressive sensing. In contrast to traditional systems where sampling intervals are linked to signal bandwidth, CS enables the link between these factors to be broken. In other words, native resolution can be increased without a corresponding increase in sampling rate. We demonstrate a potentially transformative consequence of breaking this link by showing, via the CRB, that enhanced resolution outweighs the disadvantage of SNR loss inherent in compressive RF sampling, when the system is operating in the high-SNR asymptotic regime. We demonstrate through CRB analysis and simulation that a system employing high-bandwidth waveforms that necessitate compressive sampling can outperform a traditional system with waveform bandwidth matched to the sampling rate. As additional contributions, we provide Bayesian CRB derivations for CS with optimized kernels and demonstrate how the GM-based measurement kernel optimization can be applied to a continuous parameter estimation problem.

The fundamental tradeoffs and capabilities evaluated in this paper are only part of the analysis required to arrive at a good system design. We hold A/D sampling rate constant in our comparisons because A/D technology can be a significant limiting factor in the bandwidth of RF systems. For example, in a recent survey of high-performance A/D technology, we found that in order to increase sampling rate by a factor of 10, the A/D converter would consume about 7.5 times more power and increase A/D cost by a factor of approximately 35. Meanwhile, the additional high-rate digital-to-analog (D/A) converter required

to implement a compressive approach did not have nearly the same cost or power implications. Clearly these system tradeoffs depend on specific rates and the current state of technology, so they must be evaluated for the desired sampling rates and compression ratios of specific applications. For the same signal bandwidth, full-rate sampling will always outperform compressive sampling, but may be prohibitive in terms of system cost and power consumption. Therefore, we propose that in some situations, the performance and systems analysis will favor the compressive approach.

The rest of the paper is organized as follows. In Section II, we summarize the radar signal model and associated MMSE estimator based on the GM assumption. In Section III, we summarize CS kernel optimization for both fixed and random signal amplitudes, and in Section IV we present the corresponding Bayesian CRBs on time delay estimation. In Section V, we analyze different sensing kernels by comparing pairwise distances, average output SNR, and approximate distributions of squared error. We then compare the mean-square error (MSE) performance of optimized sensing kernels against the corresponding Bayesian CRB and to random sensing kernels. We make our conclusions in Section VI.

II. Signal Model

In radar signal processing, the received echo signal $\boldsymbol{r}(t)$ can be modeled in complex baseband notation as

$$\boldsymbol{r}(t) = \alpha \boldsymbol{\psi}(t-\tau) + \boldsymbol{n}(t), \tag{1}$$

where α is the complex amplitude coefficient of the received signal, $\boldsymbol{\psi}(t-\tau)$ is the radar waveform with unknown time delay τ, and $\boldsymbol{n}(t)$ is complex additive Gaussian noise. The delayed signal can be sparsely represented by defining a waveform basis parameterized by the time delay. According to CS theory, this sparse signal can be recovered at sampling rates much lower than the standard Nyquist rate with accuracy at least as good as that provided by the standard Nyquist rate. (Indeed, our contention is that under certain circumstances, this reconstruction can be significantly *better* than for Nyquist-sampled systems operating at the same sampling rate.) Hence, we need only to measure and encode M linear projections of the received signal $\boldsymbol{r}(t)$ onto a set of measurement kernels $\{\boldsymbol{\phi}_m(t), m = 1, \cdots, M\}$, where M is much less than the number of samples that would be collected at the Nyquist rate. The m-th measurement will be a projection of the noisy received signal $\boldsymbol{r}(t)$ onto the m-th measurement kernel $\boldsymbol{\phi}_m(t)$ as

$$\begin{aligned} y_m &= \langle \boldsymbol{\phi}_m(t), \boldsymbol{r}(t) \rangle \\ &= \int_0^T \boldsymbol{\phi}_m(t) \boldsymbol{r}(t)\, dt \\ &\approx \Delta \sum_{l=1}^{L} \phi_m[l] r[l] \\ &= \Delta \sum_{l=1}^{L} \phi_m[l] \Big[\alpha \psi[l;\tau] + n[l] \Big], \end{aligned} \tag{2}$$

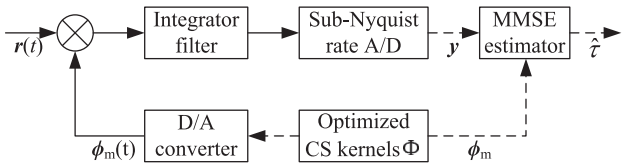


Fig. 1. Compressive sensing radar receiver, where the solid lines denote analog signals and dashed lines denote digital signals.

where $\langle \cdot, \cdot \rangle$ denotes the inner product, T is the maximum duration of the received echo signal, $L = T/\Delta$, and the integrals have been approximated by dividing the maximum duration T into discrete intervals of width Δ. If the interval Δ is much smaller than the reciprocal of the bandwidths of both the measurement kernel and the radar waveform, then the signals do not vary significantly within a discrete interval and the approximation will be accurate.

Defining $\boldsymbol{\phi}_m = [\phi_m[1], \ \phi_m[2], \ \cdots, \ \phi_m[L]]$, $\boldsymbol{\psi}(\tau) = [\psi[1;\tau], \ \psi[2;\tau], \ \cdots, \psi[L;\tau]]^{\mathrm{T}}$, and $\boldsymbol{n} = [n[1], n[2], \cdots, n[L]]^{\mathrm{T}}$, yields

$$\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{r} = \boldsymbol{\Phi}\big(\alpha \boldsymbol{\psi}(\tau) + \boldsymbol{n}\big), \tag{3}$$

where $\boldsymbol{y} = [y_1, y_2, \cdots, y_M]^{\mathrm{T}} \in \mathbb{C}^M$ is the measurement vector, $\boldsymbol{\Phi} = \big[\boldsymbol{\phi}_1^{\mathrm{T}}, \boldsymbol{\phi}_2^{\mathrm{T}}, \cdots, \boldsymbol{\phi}_M^{\mathrm{T}}\big]^{\mathrm{T}} \in \mathbb{C}^{M \times L}$ is the sensing matrix (a discrete representation of the M analog sensing kernels) with m-th row comprising the measurement kernel $\boldsymbol{\phi}_m$, $\boldsymbol{\psi}(\tau) \in \mathbb{C}^L$ is the discrete-time representation of the radar waveform delayed by τ, and $\boldsymbol{n} \sim \mathcal{CN}(\boldsymbol{0}, \mathbf{C}_{\boldsymbol{nn}})$ is zero-mean complex Gaussian noise independent of time delay τ. Here, $[\,\cdot\,]^{\mathrm{T}}$ denotes the transpose. We are careful to note here that although approximated in discrete time, the models of the received signal, noise, and sensing kernels are representations of analog signals as seen in the compressive receiver block diagram of Fig. 1. In Fig. 1, we see the analog multiplication of the noisy received waveform with the sensing kernels, an analog integration component to complete the projection of the noisy waveform onto each measurement kernel, and finally A/D conversion to obtain the measurement vector $\boldsymbol{y}$. Therefore, discrete-time representations must be at a rate that is at least as high as the Nyquist rate for the signal at the output of the analog multiplier. This output has higher bandwidth than the noisy input waveform due to the convolution of the frequency spectra of $\boldsymbol{r}(t)$ and $\boldsymbol{\phi}_m(t)$. Only the elements in the measurement vector $\boldsymbol{y}$ are actually discrete samples.

The probability distribution of the time delay τ is assumed to be known *a priori*. Among other scenarios, this assumption is reasonable for radar tracking where a current estimate of radar parameters must be updated with a new measurement. Our goal is to design the sensing matrix $\boldsymbol{\Phi}$ (namely, the corresponding analog compressive kernels $\boldsymbol{\phi}_m(t)$) to estimate the time delay τ as accurately as possible from the measurement $\boldsymbol{y}$. Defining the transmit waveform bandwidth as B, the underlying dimensionality of the received signal $\boldsymbol{r}(t)$ is $N = BT$, and the compressive sampler should have $M \ll N$.

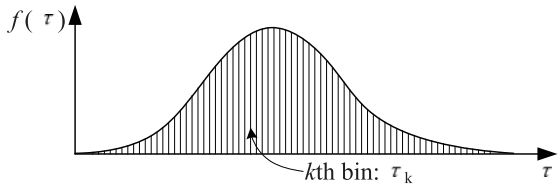


Fig. 2. Discretization of the pdf of time delay.

The pdf of the observation vector $\boldsymbol{y}$ can be expressed as

$$f(\boldsymbol{y}) = \int f(\boldsymbol{y}|\tau) f(\tau)\, d\tau \approx \sum_{k\in\mathcal{K}} f(\boldsymbol{y}|\tau_k) f(\tau_k)\Delta\tau, \quad (4)$$

where the integral is over the support of the pdf of τ and the cardinality of the set $\mathcal{K} = \{1, 2, \cdots, K\}$ is $|\mathcal{K}| = K$. Here, K is the number of bins resulting from dividing the support of $f(\tau)$ into bins of width $\Delta\tau$. This discrete approximation of $f(\tau)$, depicted in Fig. 2, results in

$$f(\boldsymbol{y}) \approx \sum_{k\in\mathcal{K}} p_k f(\boldsymbol{y}|\tau_k), \quad (5)$$

where $p_k = f(\tau_k)\Delta\tau$. When the realization of the time delay is τ_k, the corresponding measurement is

$$\begin{aligned} \boldsymbol{y}_k = \boldsymbol{y}|_{\tau=\tau_k} &= \boldsymbol{\Phi}\big(\alpha\boldsymbol{\psi}(\tau_k) + \boldsymbol{n}\big) \\ &= \alpha\boldsymbol{\Phi}\boldsymbol{\psi}(\tau_k) + \boldsymbol{\Phi}\boldsymbol{n}. \end{aligned} \quad (6)$$

When the signal amplitude α is fixed, then the observation vector in (6) is complex Gaussian with a mean of $\bar{\boldsymbol{y}}^{(k)} = \alpha\boldsymbol{\Phi}\boldsymbol{\psi}(\tau_k)$ and covariance matrix $\mathbf{C}_{\boldsymbol{yy}} = \boldsymbol{\Phi}\mathbf{C}_{\boldsymbol{nn}}\boldsymbol{\Phi}^{\mathrm{H}}$. Therefore, the conditional pdf is

$$f(\boldsymbol{y}|\tau_k) = \frac{1}{\pi^M\,|\mathbf{C}_{\boldsymbol{yy}}|} e^{-\left[\boldsymbol{y}-\bar{\boldsymbol{y}}^{(k)}\right]^{\mathrm{H}} \mathbf{C}_{\boldsymbol{yy}}^{-1}\left[\boldsymbol{y}-\bar{\boldsymbol{y}}^{(k)}\right]}, \quad (7)$$

where $[\,\cdot\,]^{\mathrm{H}}$ denotes the Hermitian transpose. When the signal amplitude α is zero-mean, complex normal according to $\alpha \sim \mathcal{CN}(0, \sigma_\alpha^2)$, then the observation vector is a zero-mean complex Gaussian random vector with covariance matrix $\mathbf{C}_{\boldsymbol{yy}}^{(k)} = \boldsymbol{\Phi}\big(\sigma_\alpha^2\boldsymbol{\psi}(\tau_k)\boldsymbol{\psi}^{\mathrm{H}}(\tau_k) + \mathbf{C}_{\boldsymbol{nn}}\big)\boldsymbol{\Phi}^{\mathrm{H}}$. Therefore, the conditional pdf is

$$f(\boldsymbol{y}|\tau_k) = \frac{1}{\pi^M\left|\mathbf{C}_{\boldsymbol{yy}}^{(k)}\right|} e^{-\boldsymbol{y}^{\mathrm{H}}\left[\mathbf{C}_{\boldsymbol{yy}}^{(k)}\right]^{-1}\boldsymbol{y}}. \quad (8)$$

Hence, under either amplitude model, the pdf of $\boldsymbol{y}$ is a weighted sum of K Gaussian distributions, which forms a Gaussian mixture distribution. In the fixed-amplitude case, the set $S_F = \big\{p_k, \bar{\boldsymbol{y}}^{(k)}, \mathbf{C}_{\boldsymbol{yy}}; k \in \mathcal{K}\big\}$ defines the GM parameters of the measurement signal $\boldsymbol{y}$. In the random-amplitude case, the set $S_R = \big\{p_k, \mathbf{0}, \mathbf{C}_{\boldsymbol{yy}}^{(k)}; k \in \mathcal{K}\big\}$ defines the GM parameters of the measurement signal $\boldsymbol{y}$. If we allow the bin size $\Delta\tau$ of the approximation to go to zero, then the approximated measurement vector density approaches the true density at the expense of an infinite number of components in the mixture.

The MMSE estimator is commonly used for estimating random parameters. Although there is not a closed-form MMSE estimator for the full pdf in (4), the MMSE estimate can be numerically computed for the approximated pdf. The MMSE estimate is the posterior expected value of the parameter. Therefore, the MMSE estimate of time delay is

$$\hat{\tau} = \mathrm{E}_{\tau|\boldsymbol{y}}\{\tau\} = \int \tau f(\tau|\boldsymbol{y})\, d\tau, \quad (9)$$

where $\mathrm{E}\{\,\cdot\,\}$ denotes statistical expectation. Using the same discretized pdf approximation as above, an approximate MMSE estimate of time delay is

$$\begin{aligned} \hat{\tau} &\approx \sum_{k\in\mathcal{K}} \tau_k f(\tau_k|\boldsymbol{y})\Delta\tau \\ &= \Delta\tau \sum_{k\in\mathcal{K}} \tau_k \frac{f(\boldsymbol{y}|\tau_k) f(\tau_k)}{f(\boldsymbol{y})} \\ &= \frac{1}{f(\boldsymbol{y})} \sum_{k\in\mathcal{K}} p_k \tau_k f(\boldsymbol{y}|\tau_k), \end{aligned} \quad (10)$$

where Bayes' theorem was used. Obviously, the MMSE estimate depends on the measured data, which depend on the sensing kernel $\boldsymbol{\Phi}$. In contrast to randomly generated sensing kernels in the CS literature, the sensing kernel $\boldsymbol{\Phi}$ now can be optimized based on the *a priori* knowledge of the time delay.

The discrete-pdf approximation to the MMSE estimate will be valid as long as the bin width is much smaller than the MSE of the non-approximated MMSE estimator. In other words, the quantization error introduced by the approximation should be small compared to the error due to measurement noise. As SNR increases, MSE is reduced, implying that finer bins are required in order to prevent quantization error from affecting the outcome. Another interpretation is that there must be enough bins such that the posterior pdf of τ can be smoothly represented. As SNR increases, the posterior pdf will become increasingly narrow and sharp, and the discrete bins must be much narrower than the variations in this pdf. Consequently, it is necessary to use more bins at high SNR so that the discrete approximation does not adversely affect either the sensing kernel optimization (via an inaccurate GM approximation) or the MMSE estimation (via an insufficiently sampled posterior pdf).

III. Compressive Sensing Kernel Optimization

In this section, we adopt the mutual information maximization criterion to optimize the sensing kernel for time delay estimation in two different amplitude cases. Considering that the optimization variable $\boldsymbol{\Phi}$ is a high-dimensional matrix, we prefer a gradient-based search method as we did in [17], which requires the gradient of the Shannon mutual information $I(\boldsymbol{y};\tau)$ between the time delay τ and the measurement $\boldsymbol{y}$ with respect to the sensing matrix $\boldsymbol{\Phi}$

$$\nabla_{\boldsymbol{\Phi}} I(\boldsymbol{y};\tau) = \nabla_{\boldsymbol{\Phi}} h(\boldsymbol{y}) - \nabla_{\boldsymbol{\Phi}} h(\boldsymbol{y}|\tau), \quad (11)$$

where $h(\boldsymbol{y}) = -\mathrm{E}\big\{\log[f(\boldsymbol{y})]\big\}$ denotes the differential entropy of the measurement $\boldsymbol{y}$, and $h(\boldsymbol{y}|\tau) = -\mathrm{E}\big\{\log[f(\boldsymbol{y}|\tau)]\big\}$ denotes the conditional differential entropy of the measurement $\boldsymbol{y}$ conditioned on the time delay τ, respectively. Here, $\nabla_{\boldsymbol{\Phi}}\{\cdot\}$ denotes the gradient with respective to $\boldsymbol{\Phi}$. Although there is not a closed-form expression for our information metric for most

pdfs of interest in practical radar applications, the previous discretization of the pdf of time delay allows an approximate path forward. By performing a first-order Taylor series expansion of the logarithm of the GM distribution, the differential entropy is approximated as [17]

$$h(\boldsymbol{y}) \approx -\log\left[\sum_{k\in\mathcal{K}} p_k f(\boldsymbol{y}_0|\tau_k)\right], \tag{12}$$

where $\boldsymbol{y}_0 = \mathrm{E}\{\boldsymbol{y}\}$ is the Taylor series expansion point.

A. Fixed Amplitude Case

In the fixed-amplitude case, we have

$$f(\boldsymbol{y}=\boldsymbol{y}_0|\tau_k) = \frac{1}{\pi^M\left|\mathbf{C}_{\boldsymbol{yy}}\right|} e^{-|\alpha|^2\left[\bar{\boldsymbol{\psi}}-\boldsymbol{\psi}(\tau_k)\right]^{\mathrm{H}}\boldsymbol{\Phi}^{\mathrm{H}}\mathbf{C}_{\boldsymbol{yy}}^{-1}\boldsymbol{\Phi}\left[\bar{\boldsymbol{\psi}}-\boldsymbol{\psi}(\tau_k)\right]}, \tag{13}$$

where $\bar{\boldsymbol{\psi}} = \sum_{k\in\mathcal{K}} p_k\boldsymbol{\psi}(\tau_k) \approx \int f(\tau)\boldsymbol{\psi}(\tau)\,d\tau$ is the mean of $\boldsymbol{\psi}(\tau)$ over the prior pdf of τ, and $\boldsymbol{y}_0 = \alpha\boldsymbol{\Phi}\bar{\boldsymbol{\psi}}$ is the mean of the measurement vector $\boldsymbol{y}$. We note that when the time delay τ is given, the only random contribution is due to noise, which leads to $h(\boldsymbol{y}|\tau) = h(\boldsymbol{\Phi n})$. Hence, the gradient of the mutual information with respect to the sensing matrix is approximated as $\nabla_{\boldsymbol{\Phi}} I(\boldsymbol{y};\tau)$ (14), shown at the bottom of this page, where the derivation is

$$\begin{aligned}
&\nabla_{\boldsymbol{\Phi}}\left\{\left[\bar{\boldsymbol{\psi}}-\boldsymbol{\psi}(\tau_k)\right]^{\mathrm{H}}\boldsymbol{\Phi}^{\mathrm{H}}\mathbf{C}_{\boldsymbol{yy}}^{-1}\boldsymbol{\Phi}\left[\bar{\boldsymbol{\psi}}-\boldsymbol{\psi}(\tau_k)\right]\right\}\\
&\quad= \mathbf{C}_{\boldsymbol{yy}}^{-1}\boldsymbol{\Phi}\left[\bar{\boldsymbol{\psi}}-\boldsymbol{\psi}(\tau_k)\right]\left[\bar{\boldsymbol{\psi}}-\boldsymbol{\psi}(\tau_k)\right]^{\mathrm{H}}\\
&\quad\quad-\mathbf{C}_{\boldsymbol{yy}}^{-1}\boldsymbol{\Phi}\left[\bar{\boldsymbol{\psi}}-\boldsymbol{\psi}(\tau_k)\right]\left[\bar{\boldsymbol{\psi}}-\boldsymbol{\psi}(\tau_k)\right]^{\mathrm{H}}\boldsymbol{\Phi}^{\mathrm{H}}\mathbf{C}_{\boldsymbol{yy}}^{-1}\boldsymbol{\Phi}\mathbf{C}_{\boldsymbol{nn}}.
\end{aligned}$$

In a special case where the complex additive Gaussian noise $\boldsymbol{n}$ in (3) is white (namely, $\mathbf{C}_{\boldsymbol{nn}} = \sigma_{\boldsymbol{n}}^2\mathbf{I}$) and the sensing kernel $\boldsymbol{\Phi}$ has orthonormal rows (namely, $\boldsymbol{\Phi}\boldsymbol{\Phi}^{\mathrm{H}} = \mathbf{I}$), the above approximate gradient (14) can be simplified as (15), shown at the bottom of this page, where $\mathbf{I}$ denotes the identity matrix and $\frac{|\alpha|^2}{\sigma_{\boldsymbol{n}}^2}$ denotes the input SNR of the target signal.

B. Random Amplitude Case

In the random-amplitude case, we have the approximate gradient of the differential entropy of the measurement with respect to the sensing matrix [16], [17], [20]

$$\nabla_{\boldsymbol{\Phi}} h(\boldsymbol{y}) \approx \frac{\sum_{k\in\mathcal{K}} p_k\left|\mathbf{C}_{\boldsymbol{yy}}^{(k)}\right|^{-1}\left[\mathbf{C}_{\boldsymbol{yy}}^{(k)}\right]^{-1}\boldsymbol{\Phi}\left(\sigma_\alpha^2\boldsymbol{\psi}(\tau_k)\boldsymbol{\psi}^{\mathrm{H}}(\tau_k)+\mathbf{C}_{\boldsymbol{nn}}\right)}{\sum_{k\in\mathcal{K}} p_k\left|\mathbf{C}_{\boldsymbol{yy}}^{(k)}\right|^{-1}}, \tag{16}$$

because

$$f(\boldsymbol{y}=\boldsymbol{y}_0=\mathbf{0}|\tau_k) = \frac{1}{\pi^M\left|\mathbf{C}_{\boldsymbol{yy}}^{(k)}\right|}. \tag{17}$$

When the random amplitude is modeled as $\alpha \sim \mathcal{CN}(0,\sigma_\alpha^2)$, the probability distribution of the measurement $\boldsymbol{y}$ conditioned on time delay τ is

$$f(\boldsymbol{y}|\tau) = \frac{1}{\pi^M\left|\mathbf{C}_{\boldsymbol{yy}|\tau}\right|} e^{-\boldsymbol{y}^{\mathrm{H}}\mathbf{C}_{\boldsymbol{yy}|\tau}^{-1}\boldsymbol{y}}, \tag{18}$$

where $\mathbf{C}_{\boldsymbol{yy}|\tau} = \boldsymbol{\Phi}\left(\sigma_\alpha^2\boldsymbol{\psi}(\tau)\boldsymbol{\psi}^{\mathrm{H}}(\tau)+\mathbf{C}_{\boldsymbol{nn}}\right)\boldsymbol{\Phi}^{\mathrm{H}}$. Hence, the differential entropy of the measurement $\boldsymbol{y}$ conditioned on the time delay τ is

$$\begin{aligned}
h(\boldsymbol{y}|\tau) &\approx -\iint f(\boldsymbol{y},\tau)\log f(\boldsymbol{y}_0|\tau)\,d\boldsymbol{y}\,d\tau\\
&= -\int \log f(\boldsymbol{y}_0|\tau) f(\tau)\,d\tau\\
&\approx \sum_{k\in\mathcal{K}} p_k \log\left|\mathbf{C}_{\boldsymbol{yy}}^{(k)}\right| + M\log\pi,
\end{aligned} \tag{19}$$

where the first approximation is based on a first-order Taylor series expansion of $\log f(\boldsymbol{y}|\tau)$ around the mean value $\boldsymbol{y}_0 = \mathbf{0}$. By taking the gradient of the resulting approximation in (19) with respect to the sensing matrix $\boldsymbol{\Phi}$, we have

$$\nabla_{\boldsymbol{\Phi}} h(\boldsymbol{y}|\tau) \approx \sum_{k\in\mathcal{K}} p_k\left[\mathbf{C}_{\boldsymbol{yy}}^{(k)}\right]^{-1}\boldsymbol{\Phi}\left(\sigma_\alpha^2\boldsymbol{\psi}(\tau_k)\boldsymbol{\psi}^{\mathrm{H}}(\tau_k)+\mathbf{C}_{\boldsymbol{nn}}\right). \tag{20}$$

Substituting (16) and (20) into (11), we can obtain the approximate gradient of the mutual information with respect to the sensing matrix directly. If we further assume that the complex additive Gaussian noise is white ($\mathbf{C}_{\boldsymbol{nn}} = \sigma_{\boldsymbol{n}}^2\mathbf{I}$), the approximate

$$\nabla_{\boldsymbol{\Phi}} I(\boldsymbol{y};\tau) \approx \frac{|\alpha|^2\sum_{k\in\mathcal{K}} p_k e^{-|\alpha|^2\left[\bar{\boldsymbol{\psi}}-\boldsymbol{\psi}(\tau_k)\right]^{\mathrm{H}}\boldsymbol{\Phi}^{\mathrm{H}}\mathbf{C}_{\boldsymbol{yy}}^{-1}\boldsymbol{\Phi}\left[\bar{\boldsymbol{\psi}}-\boldsymbol{\psi}(\tau_k)\right]}\nabla_{\boldsymbol{\Phi}}\left\{\left[\bar{\boldsymbol{\psi}}-\boldsymbol{\psi}(\tau_k)\right]^{\mathrm{H}}\boldsymbol{\Phi}^{\mathrm{H}}\mathbf{C}_{\boldsymbol{yy}}^{-1}\boldsymbol{\Phi}\left[\bar{\boldsymbol{\psi}}-\boldsymbol{\psi}(\tau_k)\right]\right\}}{\sum_{k\in\mathcal{K}} p_k e^{-|\alpha|^2\left[\bar{\boldsymbol{\psi}}-\boldsymbol{\psi}(\tau_k)\right]^{\mathrm{H}}\boldsymbol{\Phi}^{\mathrm{H}}\mathbf{C}_{\boldsymbol{yy}}^{-1}\boldsymbol{\Phi}\left[\bar{\boldsymbol{\psi}}-\boldsymbol{\psi}(\tau_k)\right]}}, \tag{14}$$

$$\nabla_{\boldsymbol{\Phi}} I(\boldsymbol{y};\tau) \approx \frac{\frac{|\alpha|^2}{\sigma_{\boldsymbol{n}}^2}\sum_{k\in\mathcal{K}} p_k e^{-\frac{|\alpha|^2}{\sigma_{\boldsymbol{n}}^2}\left[\bar{\boldsymbol{\psi}}-\boldsymbol{\psi}(\tau_k)\right]^{\mathrm{H}}\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{\Phi}\left[\bar{\boldsymbol{\psi}}-\boldsymbol{\psi}(\tau_k)\right]}\boldsymbol{\Phi}\left[\bar{\boldsymbol{\psi}}-\boldsymbol{\psi}(\tau_k)\right]\left[\bar{\boldsymbol{\psi}}-\boldsymbol{\psi}(\tau_k)\right]^{\mathrm{H}}}{\sum_{k\in\mathcal{K}} p_k e^{-\frac{|\alpha|^2}{\sigma_{\boldsymbol{n}}^2}\left[\bar{\boldsymbol{\psi}}-\boldsymbol{\psi}(\tau_k)\right]^{\mathrm{H}}\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{\Phi}\left[\bar{\boldsymbol{\psi}}-\boldsymbol{\psi}(\tau_k)\right]}}, \tag{15}$$

gradient can be simplified as

$$\nabla_{\boldsymbol{\Phi}} I(\boldsymbol{y};\tau) \approx \frac{\sum_{k\in\mathcal{K}} p_k \left|\frac{\mathbf{C}_{\boldsymbol{yy}}^{(k)}}{\sigma_{\boldsymbol{n}}^2}\right|^{-1} \left[\frac{\mathbf{C}_{\boldsymbol{yy}}^{(k)}}{\sigma_{\boldsymbol{n}}^2}\right]^{-1} \boldsymbol{\Phi}\left(\frac{\sigma_\alpha^2}{\sigma_{\boldsymbol{n}}^2}\boldsymbol{\psi}(\tau_k)\boldsymbol{\psi}^{\mathrm{H}}(\tau_k)+\mathbf{I}\right)}{\sum_{k\in\mathcal{K}} p_k \left|\frac{\mathbf{C}_{\boldsymbol{yy}}^{(k)}}{\sigma_{\boldsymbol{n}}^2}\right|^{-1}} - \sum_{k\in\mathcal{K}} p_k \left[\frac{\mathbf{C}_{\boldsymbol{yy}}^{(k)}}{\sigma_{\boldsymbol{n}}^2}\right]^{-1} \boldsymbol{\Phi}\left(\frac{\sigma_\alpha^2}{\sigma_{\boldsymbol{n}}^2}\boldsymbol{\psi}(\tau_k)\boldsymbol{\psi}^{\mathrm{H}}(\tau_k)+\mathbf{I}\right), \quad (21)$$

where $\frac{\sigma_\alpha^2}{\sigma_{\boldsymbol{n}}^2}$ denotes the input SNR of the target signal, and $\frac{\mathbf{C}_{\boldsymbol{yy}}^{(k)}}{\sigma_{\boldsymbol{n}}^2} = \boldsymbol{\Phi}\left(\frac{\sigma_\alpha^2}{\sigma_{\boldsymbol{n}}^2}\boldsymbol{\psi}(\tau_k)\boldsymbol{\psi}^{\mathrm{H}}(\tau_k)+\mathbf{I}\right)\boldsymbol{\Phi}^{\mathrm{H}}$.

With the approximate gradient of the mutual information with respect to the sensing kernel ((15) in fixed amplitude case or (21) in random amplitude case), we can optimize the compressive sensing kernel by using the gradient-based optimization method. The detailed optimization process and convergence criterion can be found in [17]. From (15) and (21), we can see that the computational complexity of calculating the gradient is $\mathcal{O}(KML^2)$ for both the fixed-amplitude case and the random-amplitude case because $M \ll L$. Hence, the overall computational complexity of the proposed information-theoretic-based compressive sensing kernel optimization for time delay estimation is $\mathcal{O}(GKML^2)$, where G denotes the number of iterations.

IV. Bayesian Cramér–Rao Bound

In this section, we derive the Bayesian CRB for time delay estimation under compressive sensing, which will be used to compare and verify time delay estimation performance for both optimized and random measurement kernels. When the probability distribution of time delay is a known prior, we can calculate the Bayesian CRB. The Bayesian version of the Fisher information J_B comprises two parts [21]

$$J_B = J_D + J_P, \quad (22)$$

where J_D and J_P denote the information gained from the data and the information due to the prior knowledge, respectively.

A. Fixed Amplitude Case

In the fixed-amplitude case, the probability distribution of the measurement $\boldsymbol{y}$ conditioned on time delay τ is

$$f(\boldsymbol{y}|\tau) = \frac{1}{\pi^M |\mathbf{C}_{\boldsymbol{yy}}|} e^{-\left[\boldsymbol{y}-\alpha\boldsymbol{\Phi}\boldsymbol{\psi}(\tau)\right]^{\mathrm{H}} \mathbf{C}_{\boldsymbol{yy}}^{-1}\left[\boldsymbol{y}-\alpha\boldsymbol{\Phi}\boldsymbol{\psi}(\tau)\right]}. \quad (23)$$

Differentiating the log-likelihood function once produces

$$\frac{d\log f(\boldsymbol{y}|\tau)}{d\tau} = \alpha^* \frac{d\boldsymbol{\psi}^{\mathrm{H}}(\tau)}{d\tau}\boldsymbol{\Phi}^{\mathrm{H}}\mathbf{C}_{\boldsymbol{yy}}^{-1}\left[\boldsymbol{y}-\alpha\boldsymbol{\Phi}\boldsymbol{\psi}(\tau)\right] + \alpha\left[\boldsymbol{y}-\alpha\boldsymbol{\Phi}\boldsymbol{\psi}(\tau)\right]^{\mathrm{H}}\mathbf{C}_{\boldsymbol{yy}}^{-1}\boldsymbol{\Phi}\frac{d\boldsymbol{\psi}(\tau)}{d\tau}, \quad (24)$$

where $(\cdot)^*$ denotes the complex conjugate. Note that the required regularity condition is satisfied

$$\mathrm{E}_{\boldsymbol{y}|\tau}\left\{\frac{d\log f(\boldsymbol{y}|\tau)}{d\tau}\right\} = 0, \quad (25)$$

because $\mathrm{E}_{\boldsymbol{y}|\tau}\{\boldsymbol{y}\} = \alpha\boldsymbol{\Phi}\boldsymbol{\psi}(\tau)$. The second derivative is

$$\frac{d^2\log f(\boldsymbol{y}|\tau)}{d\tau^2} = \alpha^* \frac{d^2\boldsymbol{\psi}^{\mathrm{H}}(\tau)}{d\tau^2}\boldsymbol{\Phi}^{\mathrm{H}}\mathbf{C}_{\boldsymbol{yy}}^{-1}\left[\boldsymbol{y}-\alpha\boldsymbol{\Phi}\boldsymbol{\psi}(\tau)\right] + \alpha\left[\boldsymbol{y}-\alpha\boldsymbol{\Phi}\boldsymbol{\psi}(\tau)\right]^{\mathrm{H}}\mathbf{C}_{\boldsymbol{yy}}^{-1}\boldsymbol{\Phi}\frac{d^2\boldsymbol{\psi}(\tau)}{d\tau^2} - 2|\alpha|^2\frac{d\boldsymbol{\psi}^{\mathrm{H}}(\tau)}{d\tau}\boldsymbol{\Phi}^{\mathrm{H}}\mathbf{C}_{\boldsymbol{yy}}^{-1}\boldsymbol{\Phi}\frac{d\boldsymbol{\psi}(\tau)}{d\tau}. \quad (26)$$

Taking the expected value yields the information gained from the data according to

$$J_D = \mathrm{E}_{\boldsymbol{y},\tau}\left\{-\frac{d^2\log f(\boldsymbol{y}|\tau)}{d\tau^2}\right\} = \mathrm{E}_\tau\left\{J_F(\tau)\right\}, \quad (27)$$

which is the expected value of the standard Fisher information

$$J_F(\tau) = \mathrm{E}_{\boldsymbol{y}|\tau}\left\{-\frac{d^2\log f(\boldsymbol{y}|\tau)}{d\tau^2}\right\} = 2|\alpha|^2\frac{d\boldsymbol{\psi}^{\mathrm{H}}(\tau)}{d\tau}\boldsymbol{\Phi}^{\mathrm{H}}\mathbf{C}_{\boldsymbol{yy}}^{-1}\boldsymbol{\Phi}\frac{d\boldsymbol{\psi}(\tau)}{d\tau} \quad (28)$$

for a specific time delay τ. On the other hand, the Fisher information due strictly to prior knowledge is

$$J_P = \mathrm{E}_\tau\left\{-\frac{d^2\log f_\tau(\tau)}{d\tau^2}\right\}, \quad (29)$$

where the prior pdf of the time delay parameter is $f_\tau(\tau)$. The Fisher information due to prior information is zero for the uniform distribution, and equal to the reciprocal of variance for the normal distribution. The corresponding Bayesian CRB is

$$\mathrm{BCRB} = J_B^{-1} = \frac{1}{2|\alpha|^2\mathrm{E}_\tau\left\{\frac{d\boldsymbol{\psi}^{\mathrm{H}}(\tau)}{d\tau}\boldsymbol{\Phi}^{\mathrm{H}}\mathbf{C}_{\boldsymbol{yy}}^{-1}\boldsymbol{\Phi}\frac{d\boldsymbol{\psi}(\tau)}{d\tau}\right\} + J_P}. \quad (30)$$

With the assumption of white noise ($\mathbf{C}_{\boldsymbol{nn}} = \sigma_{\boldsymbol{n}}^2\mathbf{I}$) and row orthonormalization of the sensing kernel ($\boldsymbol{\Phi}\boldsymbol{\Phi}^{\mathrm{H}} = \mathbf{I}$), J_D in (27) can be simplified as [34]

$$J_D = 2\frac{|\alpha|^2}{\sigma_{\boldsymbol{n}}^2}\mathrm{E}_\tau\left\{\frac{d\left[\boldsymbol{\Phi}\boldsymbol{\psi}(\tau)\right]^{\mathrm{H}}}{d\tau}\frac{d\left[\boldsymbol{\Phi}\boldsymbol{\psi}(\tau)\right]}{d\tau}\right\}, \quad (31)$$

and the corresponding Bayesian CRB in (30) is simplified as

$$\mathrm{BCRB} = J_B^{-1} = \frac{1}{2\frac{|\alpha|^2}{\sigma_{\boldsymbol{n}}^2}\mathrm{E}_\tau\left\{\frac{d[\boldsymbol{\Phi}\boldsymbol{\psi}(\tau)]^{\mathrm{H}}}{d\tau}\frac{d[\boldsymbol{\Phi}\boldsymbol{\psi}(\tau)]}{d\tau}\right\} + J_P}. \quad (32)$$

In many cases, the expectation of $J_F(\tau)$ over τ, is not analytically tractable. In this case, one approach sometimes used is the so-called expected CRB (ECRB) [22] by taking the expectation of the conditional CRB as

$$\mathrm{ECRB} = \mathrm{E}_\tau\left\{J_F^{-1}(\tau)\right\}, \quad (33)$$

which can be evaluated using a Monte Carlo simulation.

The expression in (32) is interesting to interpret. First, the Bayesian CRB depends on the input SNR through the term $\frac{|\alpha|^2}{\sigma_{\boldsymbol{n}}^2}$. For very high input SNR, the first denominator term will dominate, and the Bayesian CRB will decrease linearly with input SNR. This behavior is typical of nonlinear parameter estimation problems operating in the asymptotic region. Second, the term inside the curly braces is essentially the square of the ℓ_2 norm of the derivative of the measurements with respect to the delay parameter, averaged over the prior pdf of the parameter. This rate of change evaluated as a derivative will generally increase linearly with increased waveform and/or kernel bandwidth. Because the Bayesian CRB depends on the inner product of these derivatives, performance will generally vary with the square of the signal bandwidth. Specific values of the derivative terms will depend on the specific interactions of the measurement kernel with the waveform, and on the specific prior pdf of τ. However, the general behavior is consistent with the traditional CRB for time delay estimation - in the asymptotic region, performance improves linearly with SNR and quadratically with signal bandwidth. In the compressive case, the measurement kernel can be used to impress or encode τ-dependent fluctuations on the measurements that vary more rapidly with τ than can be observed due to the waveform alone. Of course, these fluctuations must be unique over all possible values of τ in order for τ to be uniquely recoverable. Random measurement kernels are likely to provide this unique mapping from τ to measurement, but optimized kernels could introduce unique encodings that are enhanced for the most likely τ values (based on the prior pdf).

On the other hand, the square of the ℓ_2 norm of these derivatives will also depend on the dimensionality of the measurement kernel $\boldsymbol{\Phi}$. As more samples are taken (more rows in $\boldsymbol{\Phi}$), more fluctuation points can be measured and will contribute to the inner product inside the curly braces. Viewed another way, the number of measurements controls the number of rows in $\boldsymbol{\Phi}$, which then controls the number of terms in the inner product. Fewer measurements mean that less energy is collected, which shows up as an SNR loss, as expected in RF compressive sensing [32], [33], [35]. So in general, there is a tradeoff between SNR loss and resolution enhancement. For fixed sample rate, increased waveform bandwidth will improve the Bayesian CRB by a factor of the *square* of the bandwidth increase, but will linearly degrade the Bayesian CRB due to SNR loss. The SNR loss, however, degrades approximately linearly with the compression ratio. These conclusions apply when the input SNR is sufficient to operate in the high-performing asymptotic region where estimation performance can be considered to be *resolution limited* rather than *SNR limited*. In this asymptotic scenario, it may be beneficial to accept some SNR loss due to compression in exchange for improved native resolution of the measurements.

B. Random Amplitude Case

From (18), the log-likelihood function conditioned on the time delay τ is

$$\log f(\boldsymbol{y}|\tau) = -\boldsymbol{y}^{\mathrm{H}}\mathbf{C}_{\boldsymbol{yy}|\tau}^{-1}\boldsymbol{y} - \log\left|\mathbf{C}_{\boldsymbol{yy}|\tau}\right| - M\log\pi, \quad (34)$$

where the third term is a constant independent of τ. Hence, we have the first derivative as

$$\begin{aligned}&\frac{d\log f(\boldsymbol{y}|\tau)}{d\tau}\\ &= \boldsymbol{y}^{\mathrm{H}}\mathbf{C}_{\boldsymbol{yy}|\tau}^{-1}\frac{d\mathbf{C}_{\boldsymbol{yy}|\tau}}{d\tau}\mathbf{C}_{\boldsymbol{yy}|\tau}^{-1}\boldsymbol{y}\\ &- \sigma_\alpha^2\left[\boldsymbol{\psi}^{\mathrm{H}}(\tau)\boldsymbol{\Phi}^{\mathrm{H}}\mathbf{C}_{\boldsymbol{yy}|\tau}^{-1}\boldsymbol{\Phi}\frac{d\boldsymbol{\psi}(\tau)}{d\tau} + \frac{d\boldsymbol{\psi}^{\mathrm{H}}(\tau)}{d\tau}\boldsymbol{\Phi}^{\mathrm{H}}\mathbf{C}_{\boldsymbol{yy}|\tau}^{-1}\boldsymbol{\Phi}\boldsymbol{\psi}(\tau)\right],\end{aligned} \quad (35)$$

where $\frac{d\mathbf{C}_{\boldsymbol{yy}|\tau}}{d\tau} = \sigma_\alpha^2\boldsymbol{\Phi}\left(\frac{d\boldsymbol{\psi}(\tau)}{d\tau}\boldsymbol{\psi}^{\mathrm{H}}(\tau) + \boldsymbol{\psi}(\tau)\frac{d\boldsymbol{\psi}^{\mathrm{H}}(\tau)}{d\tau}\right)\boldsymbol{\Phi}^{\mathrm{H}}$. Here, $\mathrm{Tr}[\,\cdot\,]$ denotes the trace of a matrix. Note that the required regularity condition

$$\mathrm{E}_{\boldsymbol{y}|\tau}\left\{\frac{d\log f(\boldsymbol{y}|\tau)}{d\tau}\right\} = 0, \quad (36)$$

is satisfied because

$$\begin{aligned}&\mathrm{E}_{\boldsymbol{y}|\tau}\left\{\boldsymbol{y}^{\mathrm{H}}\mathbf{C}_{\boldsymbol{yy}|\tau}^{-1}\frac{d\mathbf{C}_{\boldsymbol{yy}|\tau}}{d\tau}\mathbf{C}_{\boldsymbol{yy}|\tau}^{-1}\boldsymbol{y}\right\}\\ &= \sigma_\alpha^2\mathrm{Tr}\left[\mathbf{C}_{\boldsymbol{yy}|\tau}^{-1}\boldsymbol{\Phi}\left(\frac{d\boldsymbol{\psi}(\tau)}{d\tau}\boldsymbol{\psi}^{\mathrm{H}}(\tau) + \boldsymbol{\psi}(\tau)\frac{d\boldsymbol{\psi}^{\mathrm{H}}(\tau)}{d\tau}\right)\right.\\ &\quad\left.\times\ \boldsymbol{\Phi}^{\mathrm{H}}\mathbf{C}_{\boldsymbol{yy}|\tau}^{-1}\mathrm{E}_{\boldsymbol{y}|\tau}\left\{\boldsymbol{y}\boldsymbol{y}^{\mathrm{H}}\right\}\right]\\ &= \sigma_\alpha^2\left[\boldsymbol{\psi}^{\mathrm{H}}(\tau)\boldsymbol{\Phi}^{\mathrm{H}}\mathbf{C}_{\boldsymbol{yy}|\tau}^{-1}\boldsymbol{\Phi}\frac{d\boldsymbol{\psi}(\tau)}{d\tau} + \frac{d\boldsymbol{\psi}^{\mathrm{H}}(\tau)}{d\tau}\boldsymbol{\Phi}^{\mathrm{H}}\mathbf{C}_{\boldsymbol{yy}|\tau}^{-1}\boldsymbol{\Phi}\boldsymbol{\psi}(\tau)\right].\end{aligned} \quad (37)$$

The second derivative is

$$\begin{aligned}\frac{d^2\log f(\boldsymbol{y}|\tau)}{d\tau^2} &= \mathbf{y}^{\mathrm{H}}\mathbf{C}_{\mathbf{yy}|\tau}^{-1}\frac{d^2\mathbf{C}_{\mathbf{yy}|\tau}}{d\tau^2}\mathbf{C}_{\mathbf{yy}|\tau}^{-1}\mathbf{y}\\ &- 2\mathbf{y}^{\mathrm{H}}\mathbf{C}_{\mathbf{yy}|\tau}^{-1}\frac{d\mathbf{C}_{\mathbf{yy}|\tau}}{d\tau}\mathbf{C}_{\mathbf{yy}|\tau}^{-1}\frac{d\mathbf{C}_{\mathbf{yy}|\tau}}{d\tau}\mathbf{C}_{\mathbf{yy}|\tau}^{-1}\mathbf{y}\\ &- \sigma_\alpha^2\left[\frac{d^2\boldsymbol{\psi}^{\mathrm{H}}(\tau)}{d\tau^2}\boldsymbol{\Phi}^{\mathrm{H}}\mathbf{C}_{\mathbf{yy}|\tau}^{-1}\boldsymbol{\Phi}\boldsymbol{\psi}(\tau) + 2\frac{d\boldsymbol{\psi}^{\mathrm{H}}(\tau)}{d\tau}\boldsymbol{\Phi}^{\mathrm{H}}\mathbf{C}_{\mathbf{yy}|\tau}^{-1}\right.\\ &\times\boldsymbol{\Phi}\frac{d\boldsymbol{\psi}(\tau)}{d\tau} + \boldsymbol{\psi}^{\mathrm{H}}(\tau)\boldsymbol{\Phi}^{\mathrm{H}}\mathbf{C}_{\mathbf{yy}|\tau}^{-1}\boldsymbol{\Phi}\frac{d^2\boldsymbol{\psi}(\tau)}{d\tau^2}\\ &- \boldsymbol{\psi}^{\mathrm{H}}(\tau)\boldsymbol{\Phi}^{\mathrm{H}}\mathbf{C}_{\mathbf{yy}|\tau}^{-1}\frac{d\mathbf{C}_{\mathbf{yy}|\tau}}{d\tau}\mathbf{C}_{\mathbf{yy}|\tau}^{-1}\boldsymbol{\Phi}\frac{d\boldsymbol{\psi}(\tau)}{d\tau}\\ &\left.- \frac{d\boldsymbol{\psi}^{\mathrm{H}}(\tau)}{d\tau}\boldsymbol{\Phi}^{\mathrm{H}}\mathbf{C}_{\mathbf{yy}|\tau}^{-1}\frac{d\mathbf{C}_{\mathbf{yy}|\tau}}{d\tau}\mathbf{C}_{\mathbf{yy}|\tau}^{-1}\boldsymbol{\Phi}\boldsymbol{\psi}(\tau)\right],\end{aligned} \quad (38)$$

where $\frac{d^2\mathbf{C}_{\boldsymbol{yy}|\tau}}{d\tau^2} = \sigma_\alpha^2\boldsymbol{\Phi}\left(\frac{d^2\boldsymbol{\psi}(\tau)}{d\tau^2}\boldsymbol{\psi}^{\mathrm{H}}(\tau) + 2\frac{d\boldsymbol{\psi}(\tau)}{d\tau}\frac{d\boldsymbol{\psi}^{\mathrm{H}}(\tau)}{d\tau} + \boldsymbol{\psi}(\tau)\frac{d^2\boldsymbol{\psi}^{\mathrm{H}}(\tau)}{d\tau^2}\right)\boldsymbol{\Phi}^{\mathrm{H}}$. Taking the expected value yields

$$\begin{aligned}J_D &= \mathrm{E}_\tau\left\{J_F(\tau)\right\}\\ &= -2\sigma_\alpha^2\mathrm{E}_\tau\left\{\Re\left[\boldsymbol{\psi}^{\mathrm{H}}(\tau)\boldsymbol{\Phi}^{\mathrm{H}}\frac{d\mathbf{C}_{\boldsymbol{yy}|\tau}^{-1}}{d\tau}\boldsymbol{\Phi}\frac{d\boldsymbol{\psi}(\tau)}{d\tau}\right]\right\},\end{aligned} \quad (39)$$

because the Fisher information for a specific time delay τ is

$$J_F(\tau) = \mathrm{E}_{\boldsymbol{y}|\tau}\left\{-\frac{d^2 \log f(\boldsymbol{y}|\tau)}{d\tau^2}\right\} = 2\sigma_\alpha^2 \Re\left[\boldsymbol{\psi}^{\mathrm{H}}(\tau)\boldsymbol{\Phi}^{\mathrm{H}}\mathbf{C}_{\boldsymbol{yy}|\tau}^{-1}\frac{d\mathbf{C}_{\boldsymbol{yy}|\tau}}{d\tau}\mathbf{C}_{\boldsymbol{yy}|\tau}^{-1}\boldsymbol{\Phi}\frac{d\boldsymbol{\psi}(\tau)}{d\tau}\right], \tag{40}$$

where $\Re[\cdot]$ denotes the real part of a complex number. With the priori Fisher information J_P in (29), the corresponding Bayesian CRB is

$$\mathrm{BCRB} = J_B^{-1} = \frac{1}{-2\sigma_\alpha^2 \mathrm{E}_\tau\left\{\Re\left[[\boldsymbol{\Phi}\boldsymbol{\psi}(\tau)]^{\mathrm{H}}\frac{d\mathbf{C}_{\boldsymbol{yy}|\tau}^{-1}}{d\tau}\frac{d[\boldsymbol{\Phi}\boldsymbol{\psi}(\tau)]}{d\tau}\right]\right\} + J_P}. \tag{41}$$

For the random-amplitude case, the signal information is expressed in the covariance matrix of the measurement pdf rather than the mean; therefore, the Bayesian CRB in (41) is different from the one in (30).

V. Simulation

In this section, we compare time delay estimation performance of the proposed optimal sensing kernel with that of the Gaussian random sensing kernel. We also study the behavior of compressive time delay estimation with respect to derived bounds and demonstrate the potentially favorable resolution/SNR loss tradeoff enabled by CS. In our simulations, the A/D converter is assumed to be operating at a fixed sampling rate of $B = 1$ MHz. The radar transmitter transmits an LFM waveform with bandwidth $CR \times B$ MHz, where CR is the compression ratio of the compressive sensing radar receiver. When the compression ratio is five ($CR = 5$), the transmit waveform bandwidth is 5×1 MHz $= 5$ MHz. When the compression ratio is $CR = 10$, the transmit waveform bandwidth is 10×1 MHz $= 10$ MHz. The CR is the important factor, whether the sampling rate or signal bandwidth is held fixed. We prefer to hold sample rate fixed in order to focus the perspective on performance gains achievable by generated waveforms with higher bandwidths than the A/D converter is capable of acquiring at Nyquist. Meanwhile, a 1-MHz waveform is also adopted for the baseline Nyquist result, such that the transmit waveform bandwidth is matched to the receiver sampling rate. Therefore, the Nyquist reference for these simulations is a constant sampling rate of 1 MHz. We also note that although we use an LFM waveform in these simulations for its relative ease in varying the waveform bandwidth as needed for comparisons, the behavior and main conclusions apply to other waveforms as well.

We define the relationship between signal amplitude and our input SNR metric according to $\alpha = \sqrt{\mathrm{SNR}} \times \sigma_{\boldsymbol{n}}$ in the fixed-amplitude case and $\alpha \sim \mathcal{CN}(0, \mathrm{SNR} \times \sigma_{\boldsymbol{n}}^2)$ in the random-amplitude case. The time delay τ is assumed to be between 0 μs and 8 μs. We uniformly discretize the pdf of time delay into $K = 1{,}000$ bins for compressive sensing kernel optimization regardless of SNR. Namely, the number of components in the Gaussian mixture in (5) is 1,000. The optimized kernel must be re-calculated for every value of compression ratio and input SNR; however, description of the detailed gradient-based compressive sensing kernel optimization process and the corresponding convergence criterion can be found in [17]. It is worth noting that when the exact input SNR is not available in practical applications for kernel optimization, we prefer to use CS kernels optimized at high SNR in the asymptotic performance region rather than those optimized at low SNR of threshold region. Our studies have indicated that, within the asymptotic region, performance is not particularly sensitive to mismatched SNR during the kernel optimization.

In Fig. 3, we compare the structure of the initial Gaussian random compressive sensing kernels ($CR = 5$) with the optimized sensing kernels, where the optimized sensing kernels are optimized with $\tau \sim \mathcal{N}\left(\frac{4}{B}, \left(\frac{1}{B}\right)^2\right)$ at SNR $= 30$ dB. The optimized sensing kernels are more focused in both the time and frequency domains and have strong components at both ends of their frequency spectra, which aids resolution performance. In the MSE performance comparison of different sensing kernels, $N_{MC} = 1{,}000$ Monte-Carlo trials are performed for each data point (SNR). The MSE is defined as

$$\mathrm{MSE} \triangleq \frac{1}{N_{MC}}\sum_{q=1}^{N_{MC}}\left(\hat{\tau}_q - \tau_q\right)^2, \tag{42}$$

where $\hat{\tau}_q$ and τ_q are the MMSE estimate and true time delay, respectively, for the qth Monte-Carlo trial.

Before comparing the MSE performance of different cases, we first analyze pairwise distances, average output SNR, and the approximate distributions of squared error for different sensing kernels. In Fig. 4, we present the pairwise signal distance maps for different sensing kernels, where the pairwise distance between compressed signals obtained for time delays τ_1 and τ_2 is defined as

$$\mathrm{d}(\tau_1, \tau_2) \triangleq \left\|\boldsymbol{\Phi}\left(\boldsymbol{\psi}(\tau_1) - \boldsymbol{\psi}(\tau_2)\right)\right\|, \forall\, \tau_1, \tau_2. \tag{43}$$

Here, $\|\cdot\|$ denotes the ℓ_2-norm of a vector. The distance map for the random sensing kernel is generated by averaging the pairwise distances over many realizations of the random sensing kernel; therefore, for the random kernel case we show expected distances, i.e., $\mathrm{E}\{\mathrm{d}(\tau_1, \tau_2)\}$. The distances should be as large as possible, implying that different values of τ will map to significantly different signal vectors, thus enabling resolution and accurate estimation. The main diagonals of the distance maps show the distance between signal vectors for the same time delay, which is zero. As the time delay separation increases, the distance between resulting signal vectors generally increases, which enables resolution and better estimation accuracy. The result in Fig. 4(a) shows that Nyquist sampling has the worst resolution, as indicated by the broadest null along the diagonal. The compressive cases enable more transmit bandwidth, so they have better native resolution and narrower diagonals than the Nyquist case. This is the fundamental resolution improvement we are seeking to achieve through a compressive solution that doesn't require corresponding A/D hardware upgrades. When range sidelobes cause two significantly different values of τ to

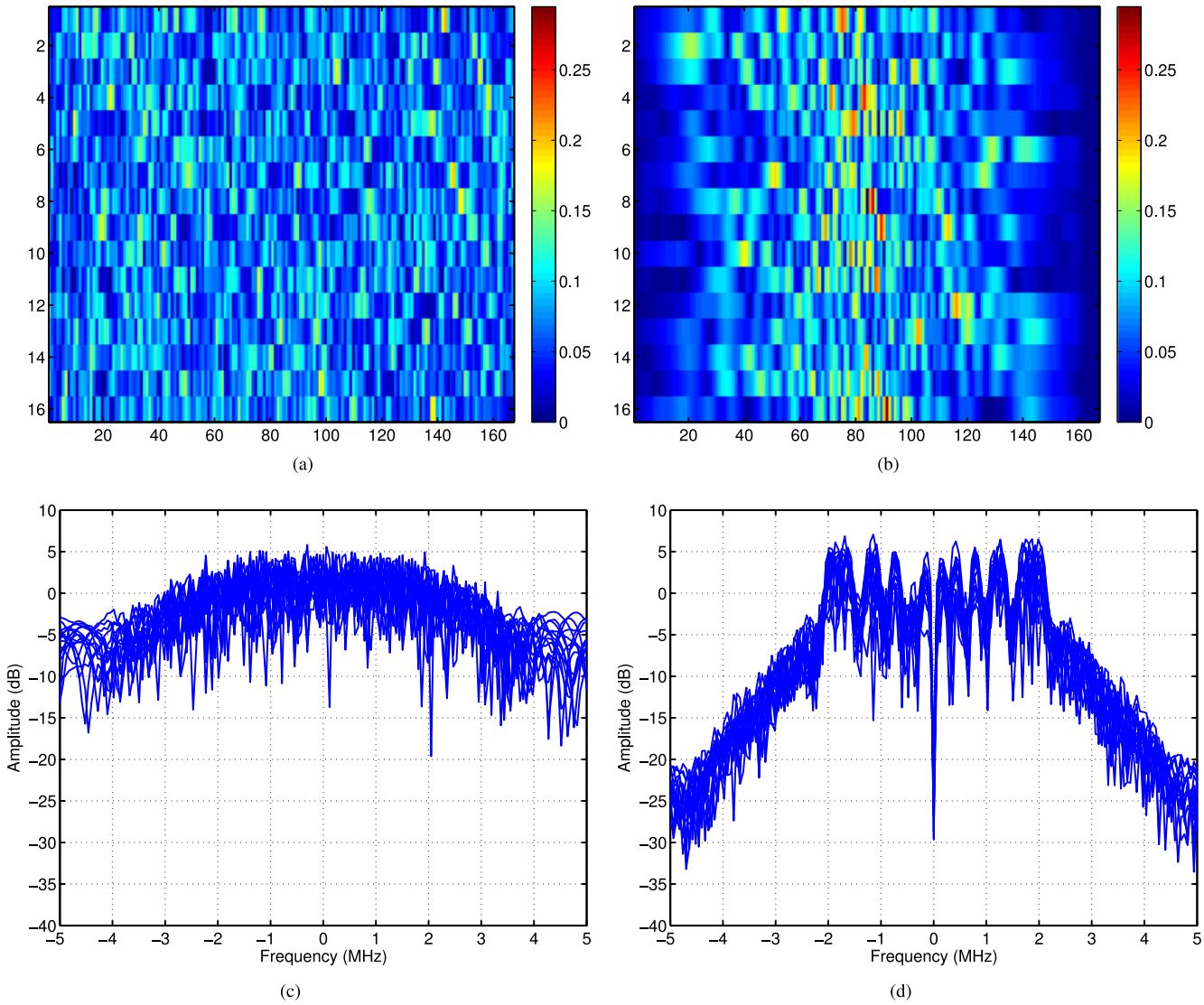


Fig. 3. Compressive sensing kernels ($CR = 5$) comparison. (a) Time response of the random (Gaussian) sensing kernels; (b) Time response of the optimized sensing kernels, where the sensing kernels are optimized with $\tau \sim \mathcal{N}\left(\frac{4}{B}, (\frac{1}{B})^2\right)$ at SNR $= 30$ dB; (c) Frequency response of the random sensing kernels; (d) Frequency response of the optimized sensing kernels.

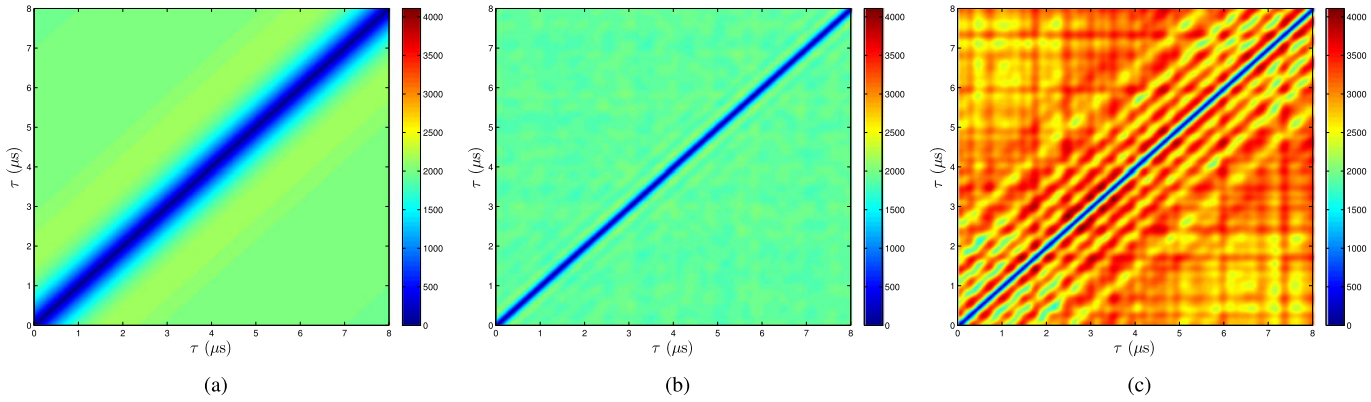
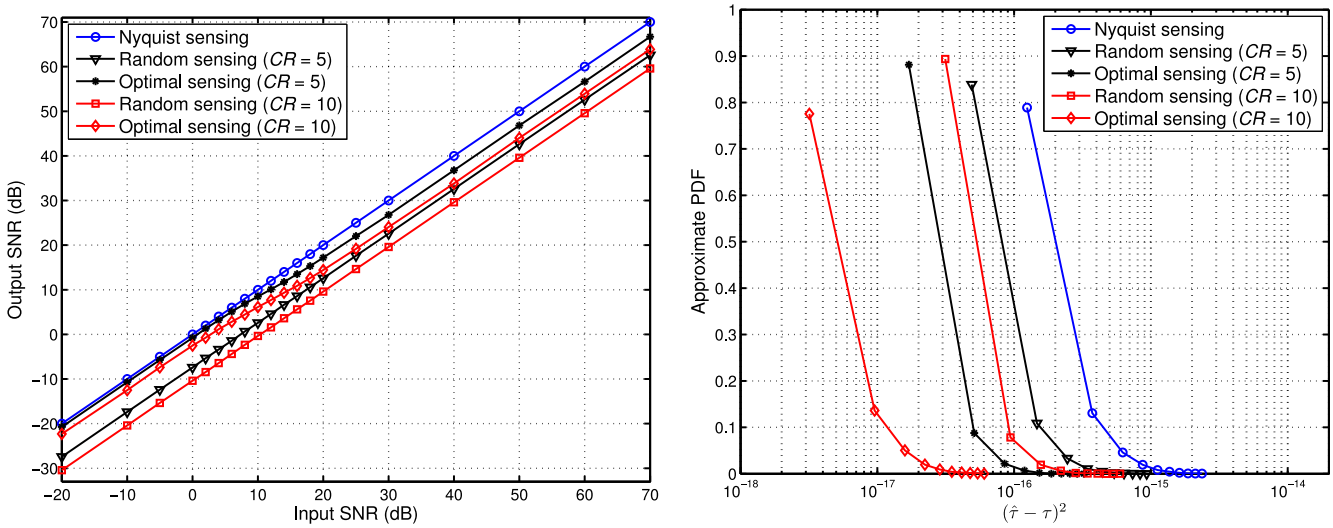


Fig. 4. Pairwise distances comparison of different sensing kernels. (a) Nyquist sensing kernel with bandwidth 1 MHz; (b) Random sensing kernels with bandwidth 5 MHz; (c) Optimal sensing kernel with bandwidth 5 MHz, where the sensing kernels are optimized with $\tau \sim \mathcal{N}\left(\frac{4}{B}, (\frac{1}{B})^2\right)$ in the fixed-amplitude case at SNR $= 30$ dB.



(a) (b)

Fig. 5. Performance comparison of different sensing kernels with fixed amplitude α and normal distributed time delay τ. (a) Output SNR versus input SNR; (b) The approximate pdf of squared error of time delay estimate at SNR $= 30$ dB.

induce similar signal vectors, we see localized drops in the distance maps. These are clearly seen in the map for the optimized kernel, but they are also present for any one realization of a random sensing kernel (in fact, they are much worse for a single random measurement kernel). When averaged over multiple random kernels, these randomly located range sidelobes average out to a constant level as seen in Fig. 4(b). Finally, we point out that the distance map for the optimized kernel depends on SNR because the optimized kernel design depends on SNR.

Finally, in comparing the distance map for the optimized kernel versus the averaged distance map for random kernels, we see that they have approximately the same resolution (same diagonal width), but the optimized kernel produces larger distances (and, hence, better resolution) away from the diagonal, especially near the center of the map where the prior pdf of τ indicates more likely values. At the tails of the delay parameter's prior distribution, corresponding to the corners of the distance map, the optimized kernel allows the distances to decrease because these values are less likely to occur. We can interpret this behavior as an optimal projection of observed waveforms into compressed space according to input SNR and the parameter's prior pdf. Overall, this distance behavior produces better MSE results, as we will show shortly. Although the random kernel has good resolution, the high sidelobes of any one kernel realization degrade its average performance.

In Fig. 5(a), we compare the average output SNR of different sensing kernels compared to the input SNR for the fixed-amplitude case, where the average output SNR is calculated as

$$\text{Output SNR} \triangleq \frac{\mathrm{E}\left\{\left\|\alpha\boldsymbol{\Phi}\boldsymbol{\psi}(\tau)\right\|^2\right\}}{\mathrm{E}\left\{\left\|\boldsymbol{\Phi}\boldsymbol{n}\right\|^2\right\}}. \tag{44}$$

This metric evaluates SNR loss due to compressive sampling. It can be seen from the figure that there is no SNR loss for Nyquist sampling (the output SNR equals the input SNR for all input values). As the theoretical analysis predicts, there is a fixed 7 dB ($\approx 10\log_{10} 5$) SNR loss for random sensing kernels with ($CR = 5$), and a fixed 10 dB ($= 10\log_{10} 10$) SNR loss for $CR = 10$. In other words, random kernels have an average SNR loss equal to the compression ratio. The optimized sensing kernel has different behavior. At low input SNR (e.g., below 0 dB), the optimized kernel clearly attempts to retain SNR, and the SNR loss compared to Nyquist sampling is minimal. At high input SNR (e.g., above 20 dB), the optimized kernel admits a compromise in SNR loss in order to improve resolution and discriminating capability across different values of τ. We see this behavior change from an input SNR of about 0 dB to an input SNR of about 20 dB. This interval of SNR $\in [0\text{ dB}, 20\text{ dB}]$ is the so-called *threshold region* typical of nonlinear estimation problems. Below the threshold region, the estimation problem is noise limited and large errors can result. Above the threshold region is the asymptotic region where the system is resolution limited, and a good compressive solution may improve performance.

Although the average output SNR of the optimized CS kernels is worse than the output SNR of the Nyquist case, the optimized CS case will still have smaller estimation MSE in the asymptotic region, as seen in Fig. 5(a). In Fig. 5(b), the approximate pdf of squared estimation error (numerically estimated over 1,000 trials) at an input SNR of 30 dB shows that the compressive system has better estimation performance. These distributions show that the average error will clearly be lower for the compressive systems, and the worst-case errors occur infrequently. At low input SNR (not shown), however, these conclusions do not hold, as the system is noise limited and SNR loss due to compression is detrimental to performance. The following MSE results demonstrate this behavior.

A. MSE Results for Fixed Amplitude

In the fixed-amplitude case, we consider two different distributions on time delay τ: a uniform distribution $\mathcal{U}\left[0, \frac{8}{B}\right]$

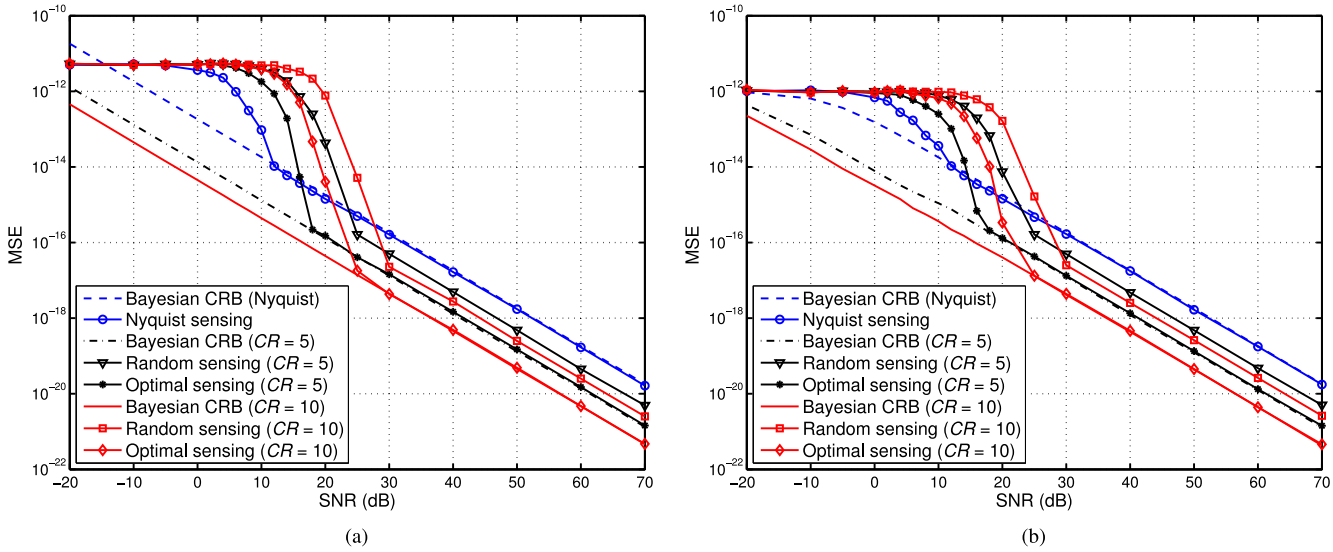


Fig. 6. MSE performance comparison and the corresponding CRBs, where the amplitude α is fixed. (a) The distribution of time delay τ is uniform $\tau \sim \mathcal{U}\left[0, \frac{8}{B}\right]$; (b) The distribution of time delay τ is normal $\tau \sim \mathcal{N}\left(\frac{4}{B}, (\frac{1}{B})^2\right)$.

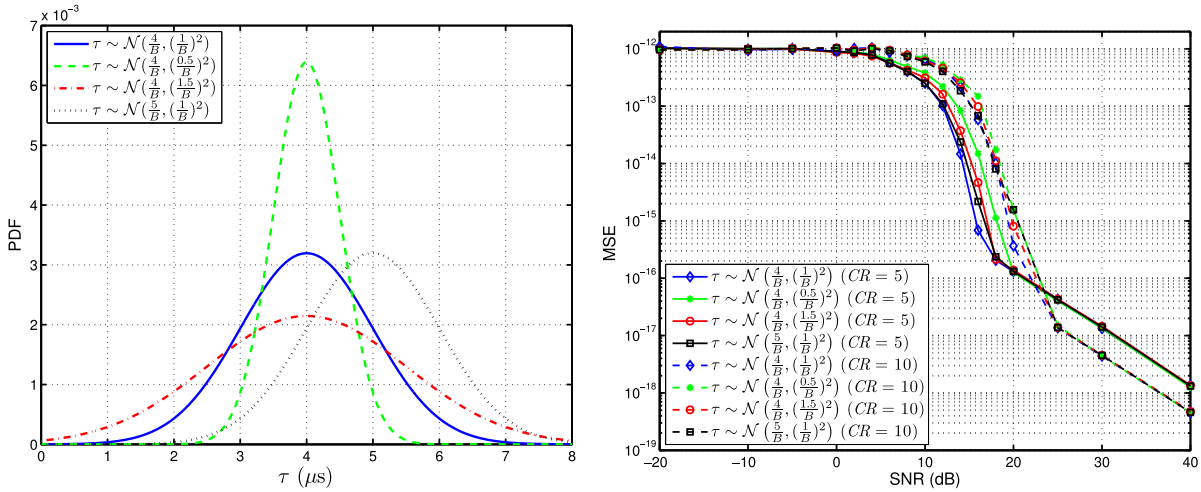
and a normal distribution $\mathcal{N}\left(\frac{4}{B}, (\frac{1}{B})^2\right)$. From Fig. 6, it is apparent that the proposed optimal sensing kernel outperforms the random sensing kernels, especially at intermediate and high input SNR. The impact of SNR loss due to compression is observed in the fact that the Nyquist kernel achieves the asymptotic region at lower input SNR than required by the compressive kernels. Because SNR is lost in the compression, additional input SNR is required in order to make up for this loss and to reach the desired asymptotic performance region. Note that for a uniform prior pdf, the prior is uninformative with $J_P = 0$, which leads the Bayesian CRB not to converge to a constant level at low SNR. For the normal prior, the Bayesian CRB converges to a constant level at low SNR with the aid of prior information $J_P = \frac{1}{\sigma_\tau^2} = B^2$.

Fig. 6 shows that the compressive schemes require more input SNR than Nyquist sampling in order to achieve Bayesian CRB performance, but once the input SNR is sufficient (e.g., above 20 dB), performance of the compressively sampled high-bandwidth waveforms is improved over the low-bandwidth Nyquist sampling. For example, we can see that for the same input SNR, the compressive schemes outperform Nyquist sampling, as long as the input SNR is at least about 20–25 dB. We note that performance in the asymptotic region is consistent with the Bayesian CRB predictions.

Unlike random projections, the optimized sensing kernel depends on *a priori* knowledge of the parameter to be estimated, as the approximate gradient $\nabla_\Phi I(\boldsymbol{y};\tau)$ ((15) or (21)) shows. Hence, it is important to study the sensitivity of the information-theoretic-based sensing kernel optimization with respect to mismatch in the knowledge of the *a priori* distribution. For the following results, we kept the actual distribution used by MMSE estimation the same as before ($\tau \sim \mathcal{N}\left(\frac{4}{B}, (\frac{1}{B})^2\right)$), while changing the *a priori* distribution for the sensing kernel optimization to different distributions shown in Fig. 7(a). From Fig. 7(b), the MSE performance difference for mismatched *a priori* distribution only occurs in the threshold region, i.e., the interval of SNR $\in [0\text{ dB}, 20\text{ dB}]$. More specifically, when the *a priori* distribution cannot cover the actual distribution, e.g., $\tau \sim \mathcal{N}\left(\frac{4}{B}, (\frac{0.5}{B})^2\right)$, the sensing kernel optimized from the *a priori* distribution results in larger estimation error. From this observation we conclude that the information-theoretic-based sensing kernel is a little sensitive to the *a priori* distribution of the parameter to be estimated; therefore, the *a priori* distribution should try its best to cover more possibilities in practical applications while remaining tight enough to provide some benefit. By comparing different compression ratios ($CR = 5$ versus $CR = 10$), we can see that the compressive sensing kernel with fewer rows ($CR = 10$) is more robust to the *a priori* distribution mismatch than the sensing kernel with more rows ($CR = 5$). Nevertheless, the optimized sensing kernel still significantly outperforms the random sensing kernel by comparing Fig. 7(b) with Fig. 6(b).

B. Random Amplitude Case

In the random-amplitude case, the time delay τ is assumed to follow a normal distribution $\mathcal{N}\left(\frac{4}{B}, (\frac{1}{B})^2\right)$. From Fig. 8, we can see that the proposed optimal sensing kernel outperforms both Nyquist and random compressive sensing at high SNR, which is same as the previous results in the fixed-amplitude case. However, compared to the fixed-amplitude case, the *threshold region* appears at much higher input SNR and the performance advantage of the optimal sensing kernel is not as significant (though still an order of magnitude improvement at the highest SNR values). In addition, we also observe that there is a gap between the MSE curve and the corresponding Bayesian CRB curve in the asymptotic performance region. That is to say, the derived Bayesian CRB (41) is not as tight as it was in the fixed-amplitude case, at least not within the input SNR range evaluated here.



(a) (b)

Fig. 7. Sensitivity analysis of the proposed sensing kernel optimization method. The actual time delay follows a normal distribution as $\tau \sim \mathcal{N}\left(\frac{4}{B}, (\frac{1}{B})^2\right)$, and the amplitude α is fixed. (a) The *a priori* distribution of the time delay for the sensing kernel optimization; (b) MSE performance comparison with respect to mismatched *a priori* distribution of the time delay.

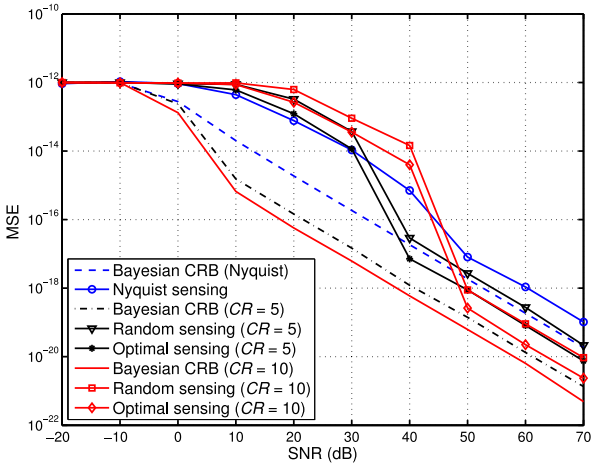


Fig. 8. MSE performance comparison and the corresponding Bayesian CRBs. The distribution of time delay τ is normal, and the amplitude α is random.

VI. Conclusion

In this paper, we have evaluated the fundamental behavior of time delay estimation in a sub-Nyquist receiver and applied a sensing kernel optimization method previously proposed for radar target profiling in [16], [17]. The optimization method is applied by quantizing the prior pdf of the time delay parameter, which leads to a Gaussian mixture distribution and eventually to a gradient-based search process for an optimized compressive sensing kernel. The results demonstrate the inherent and potentially beneficial design tradeoff between SNR loss and system resolution enabled by compressive sampling. This tradeoff is likely a losing proposition when low input SNR is expected, but compressive implementations may result in improved performance when SNR is not the limiting factor. At high SNR, the increased native resolution of a compressive implementation enables compressive high-resolution radar imaging that may outweigh the inherent SNR loss that is a disadvantage of compressive systems.

Further, we show how *a priori* knowledge can be used to enhance the performance of compressive systems by properly trading between SNR loss and resolution. We believe the basic conclusions observed in the time delay application are valid for other compressive parameter estimation problems. If these compressive sampling structures enable, for example, increased integration time or increased aperture, then higher native resolution can be exploited at high SNR [20] (which has always been the motivation for structures such as sparse arrays [36], even before the recent focus on CS). These basic conclusions include: 1) the threshold input SNR of a nonlinear estimation problem will increase due to the unavoidable SNR loss of compressive RF systems, 2) in the asymptotic performance region, the benefit of additional resolution enabled by compressive sampling can outweigh the disadvantage of reduced SNR, and 3) SNR loss can be mitigated and estimation performance can be improved by optimizing the CS measurement kernel according to the prior pdf of the parameter and expected SNR of the scenario.

Acknowledgment

The authors would like to thank the associate editor Prof. Mark Plumbley and the anonymous reviewers for their helpful comments and suggestions.

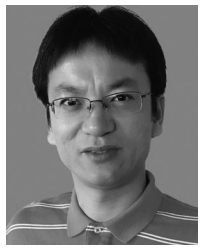
References

[1] M. Wax and A. Leshem, "Joint estimation of time delays and directions of arrival of multiple reflections of a known signal," *IEEE Trans. Signal Process.*, vol. 45, no. 10, pp. 2477–2484, Oct. 1997.

[2] J. Riba, J. Sala, and G. Vázquez, "Conditional maximum likelihood timing recovery: Estimators and bounds," *IEEE Trans. Signal Process.*, vol. 49, no. 4, pp. 835–850, Apr. 2001.

[3] A. Dogandzi and A. Nehorai, "Cramér–Rao bounds for estimating range, velocity, and direction with an active array," *IEEE Trans. Signal Process.*, vol. 49, no. 6, pp. 1122–1137, Jun. 2001.

[4] B. M. Sadler, N. Liu, and Z. Xu, "Ziv–Zakai bounds on time delay estimation in unknown convolutive random channels," *IEEE Trans. Signal Process.*, vol. 58, no. 5, pp. 2729–2745, May 2010.
[5] K. Gedalyahu and Y. C. Eldar, "Time-delay estimation from low-rate samples: A union of subspaces approach," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3017–3031, Jun. 2010.
[6] H. S. Mir and K. T. Wong, "Low-rate sampling technique for range-windowed radar/sonar using nonlinear frequency modulation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 51, no. 3, pp. 1972–1979, Jul. 2015.
[7] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
[8] E. J. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.
[9] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.
[10] M. F. Duarte *et al.*, "Single-pixel imaging via compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 83–91, Mar. 2008.
[11] R. Baraniuk and P. Steeghs, "Compressive radar imaging," in *Proc. IEEE Radar Conf.*, Waltham, USA, MA, Apr. 2007, pp. 128–133.
[12] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York, NY, USA: Wiley, 2006.
[13] M. A. Neifeld, A. Ashok, and P. K. Baheti, "Task-specific information for imaging system analysis," *J. Opt. Soc. Amer. A*, vol. 24, no. 12, pp. B25–B41, Dec. 2007.
[14] W. R. Carson, M. Chen, M. R. D. Rodrigues, R. Calderbank, and L. Carin, "Communications-inspired projection design with application to compressive sensing," *SIAM J. Imag. Sci.*, vol. 5, no. 4, pp. 1185–1212, 2012.
[15] J. Duarte-Carvajalino, G. Yu, L. Carin, and G. Sapiro, "Task-driven adaptive statistical compressive sensing of Gaussian mixture models," *IEEE Trans. Signal Process.*, vol. 61, no. 3, pp. 585–600, Feb. 2013.
[16] Y. Gu and N. A. Goodman, "Compressed sensing kernel design for radar range profiling," in *Proc. IEEE Radar Conf.*, Ottawa, On, Canada, Apr./May, 2013, pp. 1–5.
[17] Y. Gu, N. A. Goodman, and A. Ashok, "Radar target profiling and recognition based on TSI-optimized compressive sensing kernel," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3194–3207, Jun. 2014.
[18] Y. Gu and N. A. Goodman, "Time domain CS kernel design for mitigation of wall reflections in urban radar," in *Proc. IEEE Sens. Array Multichannel Signal Process. Workshop*, A Coruña, Spain, Jun. 2014, pp. 493–496.
[19] N. A. Goodman, Y. Gu, and J. Bae, "Measurement kernel design for HRR imaging of urban objects," in *Compressive Sensing for Urban Radar*. Boca Raton, FL, USA: CRC Press, 2014, pp. 197–229.
[20] Y. Gu, Y. D. Zhang, and N. A. Goodman, "Optimized compressive sensing-based direction-of-arrival estimation in massive MIMO," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, New Orleans, LA, USA, Mar. 2017, pp. 3181–3185.
[21] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part IV: Optimum Array Processing*. New York, NY, USA: Wiley, 2002.
[22] H. L. Van Trees and K. L. Bell, *Bayesian Bounds for Parameter Estimation and Nonlinear Filtering/Tracking*. New York, NY, USA: Wiley, 2007.
[23] B. Babadi, N. Kalouptsidis, and V. Tarokh, "Asymptotic achievability of the Cramér–Rao bound for noisy compressive sampling," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 1233–1236, Mar. 2009.
[24] R. Niazadeh, M. Babaie-Zadeh, and C. Jutten, "On the achievability of Cramér–Rao bound in noisy compressed sensing," *IEEE Trans. Signal Process.*, vol. 60, no. 1, pp. 518–526, Jan. 2012.
[25] P. Pakrooh, L. L. Scharf, A. Pezeshki, and Y. Chi, "Analysis of Fisher information and the Cramer–Rao bound for nonlinear parameter estimation after compressed sensing," in *Proc. IEEE Int.Conf. Acoust., Speech Signal Process.*, Vancouver, BC, Canada, May 2013, pp. 6630–6634.
[26] D. Ramasamy, S. Venkateswaran, and U. Madhow, "Compressive parameter estimation in AWGN," *IEEE Trans. Signal Process.*, vol. 62, no. 8, pp. 2012–2027, Apr. 2014.
[27] D. Guo, S. Shamai (Shitz), and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, Apr. 2005.
[28] J. Ziv and M. Zakai, "Some lower bounds on signal parameter estimation," *IEEE Trans. Inf. Theory*, vol. IT-15, no. 3, pp. 386–391, May 1969.
[29] S. Bellini and G. Tartara, "Bounds on error in signal parameter estimation," *IEEE Trans. Commun.*, vol. COM-22, no. 3, pp. 340–342, Mar. 1974.
[30] K. Bell, Y. Steinberg, Y. Ephraim, and H. L. Van Trees, "Extended Ziv–Zakai lower bound for vector parameter estimation," *IEEE Trans. Inf. Theory*, vol. 43, no. 2, pp. 624–637, Mar. 1997.
[31] E. Arias-Castro and Y. C. Eldar, "Noise folding in compressed sensing," *IEEE Signal Process. Lett.*, vol. 18, no. 8, pp. 478–481, Aug. 2011.
[32] B. Pollock and N. A. Goodman, "An examination of the effects of sub-Nyquist sampling on SNR," *Proc. SPIE*, vol. 8365, Jun. 2012, Art. no. 836504.
[33] M. A. Davenport, J. N. Laska, J. R. Treichler, and R. G. Baraniuk, "The pros and cons of compressive sensing for wideband signal acquisition: Noise folding versus dynamic range," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4628–4642, Sep. 2012.
[34] Y. Gu and N. A. Goodman, "Compressive sensing kernel optimization for time delay estimation," in *Proc. IEEE Radar Conf.*, Cincinnati, OH, USA, May 2014, pp. 1209–1213.
[35] N. A. Goodman and L. C. Potter, "Pitfalls and possibilities of radar compressive sensing," *Appl. Opt.*, vol. 54, no. 8, pp. C1–C13, Mar. 2015.
[36] Z. Shi, C. Zhou, Y. Gu, N. A. Goodman, and F. Qu, "Source estimation using coprime array: A sparse reconstruction perspective," *IEEE Sensors J.*, vol. 17, no. 3, pp. 755–765, Feb. 2017.



Yujie Gu (M'10–SM'16) received the B.E. degree in mechanical engineering from Harbin Institute of Technology, Harbin, China, in 2001, the M.S. degree in control engineering from Sichuan University, Chengdu, China, in 2004, and the Ph.D. degree in electrical engineering from Zhejiang University, Hangzhou, China, in 2008.

After graduation, he was an R&D Engineer at CETC 51, Shanghai, China. From 2009 to 2010, he was in the Department of Electrical and Computer Engineering, Concordia University, Montréal, QC, Canada. From 2010 to 2011, he was in the School of Engineering, Bar-Ilan University, Ramat-Gan, Israel. He worked as an Associate Professor in the Shanghai Advanced Research Institute, Chinese Academy of Sciences, Shanghai, China, in 2012. After that, he was in the School of Electrical and Computer Engineering and the Advanced Radar Research Center, University of Oklahoma, Norman, OK, USA, and the Department of Computer Science, Georgia State University, Atlanta, GA, USA, and is currently with the Department of Electrical and Computer Engineering, Temple University, Philadelphia, PA, USA. His research interests include statistical and array signal processing. He was awarded the ReSMiQ Postdoctoral Scholarship from Microsystems Strategic Alliance of Québec, Montréal, QC, Canada in 2009. He is an Associate member of the Sensor Array and Multichannel Technical Committee of the IEEE Signal Processing Society. He serves on the Editorial Boards of *Signal Processing*, *Electronics Letters*, *IET Signal Processing*, *Circuits, Systems and Signal Processing* and the *Journal on Advances in Signal Processing*. He has been member of the Technical Program Committee of several international conferences, including IEEE ICASSP, SAM, and CoSeRa.

Nathan A. Goodman (S'98–M'02–SM'07) received the B.S., M.S., and Ph.D. degrees in electrical engineering from the University of Kansas, Lawrence, KS, USA, in 1995, 1997, and 2002, respectively.

From 1996 to 1998, he was an RF Systems Engineer in Texas Instruments, Dallas, TX, USA. From 1998 to 2002, he was a Graduate Research Assistant in the Radar Systems and Remote Sensing Laboratory, University of Kansas. He was a faculty member in the department of electrical communication engineering, University of Arizona, Tucson, AZ, USA, from 2002 to 2011, and is currently a Professor in the School of Electrical and Computer Engineering and the Director of Research for the Advanced Radar Research Center at the University of Oklahoma, Norman, OK, USA. His research interests include radar and array signal processing. He was awarded the Madison A. and Lila Self Graduate Fellowship from the University of Kansas, Lawrence, KS, USA, in 1998. He was also awarded the IEEE 2001 International Geoscience and Remote Sensing Symposium Interactive Session Prize Paper Award. He has served as Technical Co-Chair for the 2011 IEEE Radar Conference, finance chair for the 2012 SAM workshop, and an Associate Editor-in-Chief for Elsevier's *Digital Signal Processing* journal. He is currently an Associate Editor for the IEEE TRANSACTIONS ON AEROSPACE AND ELECTRONIC SYSTEMS and General Co-Chair for the 2018 IEEE Radar Conference. He also serves as a Co-Chair for the NATO SET-227 research task group on Cognitive Radar, a Lecturer for the SET-216 lecture series on Cognition and Radar Sensing, and as a member of the IEEE AES Radar Systems Panel.