HHS Public Access

Author manuscript

IEEE Trans Signal Process. Author manuscript; available in PMC 2021 January 01.

Published in final edited form as:

IEEE Trans Signal Process. 2020 ; 68: 3312–3324. doi:10.1109/tsp.2020.2997181.

Parametric Signal Estimation Using the Cumulative Distribution Transform

Abu Hasnat Mohammad Rubaiyat,
Department of Electrical and Computer Engineering, University of Virginia, Charlottesville, VA 22904 USA

Kyla M. Hallam,
U.S. Naval Research Laboratory, Washington, DC 20375 USA

Jonathan M. Nichols,
U.S. Naval Research Laboratory, Washington, DC 20375 USA

Meredith N. Hutchinson,
U.S. Naval Research Laboratory, Washington, DC 20375 USA

Shiying Li,
Department of Biomedical Engineering, University of Virginia, Charlottesville, VA 22908 USA

Gustavo K. Rohde
Department of Biomedical Engineering, University of Virginia, Charlottesville, VA 22908 USA

Abstract

We present a new method for estimating signal model parameters using the Cumulative Distribution Transform (CDT). Our approach minimizes the Wasserstein distance between measured and model signals. We derive some useful properties of the CDT and show that the resulting estimation problem, while nonlinear in the original signal domain, becomes a linear least squares problem in the transform domain. Furthermore, we discuss the properties of the estimator in the presence of noise and present a novel approach for mitigating the impact of the noise on the estimates. The proposed estimation approach is evaluated by applying it to a source localization problem and comparing its performance against traditional approaches.

Keywords

Signal parameter estimation; Wasserstein distance; CDT

I. Introduction

SIGNAL parameter estimation is at the heart of many signal processing applications that involve localisation and tracking of a source signal, e.g. radar [1], underwater acoustics [2], source localization [3] [4], seismology [5] [6], communication [7] etc. All these systems require estimation of the values of a group of parameters. In radar systems, for example, the

ar3fx@virginia.edu.

estimated time delay and Doppler stretch between transmitted and received signals are used to determine the position and speed of a target object (Fig. 1a). Fig. 1b illustrates another example where the location of an AE (acoustic emission) source is determined using the estimated time difference of arrivals (TDOAs) of the signals received by four sensors. Typical estimation techniques involve maximizing the likelihood function [1], which most often yield to non-convex optimization problems. In this paper, we propose to solve the parametric signal estimation problem by minimizing Wasserstein distance between measured and model signals. To solve the ensuing transport problem, we rely on a novel technique called the cumulative distribution transform (CDT) introduced in [8] for the purposes of simplifying the estimation process.

A. Estimation as a transport problem

We propose to solve certain signal estimation problems borrowing concepts from optimal transport theory [9]. Specifically, we are interested in the case where a strictly positive quantity $s(t)$ is undergoing a parameterized change of variables

$$s_g(t) = g_p'(t)s(g_p(t)) \tag{1}$$

where $g_p(t)$ is a one to one, differentiable function with parameters $\boldsymbol{p}\left(\text{e.g. } g_p(t) = \sum_{k=0}^{K-1} p_k t^k\right)$ and $g_p'(t) = dg_p/dt$. This signal model is particularly pertinent in physical situations where $s(t)$ represents a conserved quantity, e.g., intensity, that evolves in time or space. To emphasize this point we note that Eqn. (1) can be restated as:

$$\int_{-\infty}^{t} s_g(u)du = \int_{-\infty}^{g_p(t)} s(u)du \tag{2}$$

which underscores the conservation of $s(t)$ under the action of $g_p(t)$. Here, u is the integration variable. Indeed, Eqn. (1) is simply a Lagrangian restatement of the continuity equation from continuum mechanics where the function $g_p(t)$ transforms the independent variable according to the problem physics (see e.g., [10, 11]).

Such models are common in wave optics, for example, where Eqn. (1) is seen to operate on the squared magnitude of a wavefunction or an electric field (see e.g., Schrodinger equation [12] or paraxial wave equation [13, 14]). For example, if $s(t)$ represents the time-varying optical intensity of a beam propagating through a lossless medium, $g_p(t)$ captures the influence of the medium to yield the modified intensity $s_g(t)$ [11]. Similar physics can be observed in phase modulated acoustic signals of finite duration (i.e., “pulses”) propagating through linear elastic solids [15]. In short, the signal model used in this work is consistent with data collected from a broad and important class of physical systems.

Note that measured signals may or may not conform to the approach outlined above since they may not be strictly positive or conserve energy. These properties can be guaranteed, however, under an appropriate normalization scheme. Denoting the measured signal $z_g(t),\ t \in \Omega_z$, we can associate with this signal a positive probability density function (PDF):

$$s_g(t) = B(z_g)(t) := \frac{z_g^2(t)}{\int_{\Omega_z} z_g^2(t)dt} \quad (3)$$

where $B(.)$ is a normalization scheme that transforms raw signals into PDFs. Although this transformation is non-invertible, it guarantees strict positivity and a constant signal energy in accordance with our signal model (1). The impact of this normalization scheme on the estimation problem will be discussed in section (V). Note that $B(z_g)$ is not one-to-one function of $z_g(t)$, and thus critical phase information may not be retrieved [16]. This would impede us from performing correct estimation in case of estimating the shift of periodic signals (e.g. a pure sign wave); however, in this paper we are interested to estimate the parameters of signals that are mostly transient. Therefore our approach does not require recovering $z_g(t)$.

The goal of this work is to illustrate how relationships (1) and (2) can be leveraged to produce estimates of the parameters $\boldsymbol{p}$ of $g_p(t)$ that governs the modulation or modification of such signals. In particular, we will show how this nonlinear, generally non-convex estimation problem in the time domain can be transformed into a linear least-squares problem in the CDT domain.

B. Related works

Previously proposed methods include maximizing cross-correlation [17], l_p correlation [18], maximizing the magnitude of difference between measured and template signals [19], entropy [20] and mutual information based methods [21][22]. All these approaches assume that the only difference between the measured and template signals is the time delay, in addition to noise. In real applications, however, signals may undergo complex parameterized changes. In radar related estimation problems, for example, motion of the target object introduces linear dispersion (also called Doppler stretch) along with the time delay. In such cases the above mentioned techniques may produce erroneous estimates.

Several subspace based methods [23] [24] [25] [26] have been proposed to jointly estimate time delay and Doppler parameters. Most of the subspace methods exploit narrowband approximation of the signals so that the Doppler effect can be modeled as frequency shift, and hence, can be estimated explicitly. A *search* over a parameter space is still required for time delay estimation, which is computationally expensive. Colonnese et al. [27] proposed a generalized method of moments (GMM) for estimating shift/translation, i.e. location parameters. Applying the GMM to parameters other than shift requires a transformation function to convert it to a location parameter; for example, a transformation realizing exponential warping of the independent axis can be applied to turn a scale parameter into a location one. Moreover, similar to subspace methods the GMM based approach requires computationally expensive *search* to estimate the parameters. Joint estimation of time delay and Doppler stretch has also been studied in the literature where the estimation techniques involve maximizing the ambiguity function between the measured and the template signals [28] [29] [30]. In most cases, ambiguity function-related techniques yield non-convex optimization problems for which global minima may be difficult to produce. The proposed

method described in this paper involves minimizing Wasserstein distance which yields to a convex problem.

The Wasserstein metric is a well developed concept in the optimal transport theory [31], which measures the difference between two distributions by the optimal cost of rearranging one distribution into the other. It has been proven to be a suitable tool to model and solve problems in the areas of signal processing and machine learning [9]. Nichols et al. [4] proposed an estimator based on the Wasserstein distance for estimating the time delay, but they did not address the linear dispersion or other forms of transformation. In [32], Engquist and Froese first used this metric in the seismic inversion problems. Then, the idea of using the Wasserstein distance to identify a geophysical model from the observations was exploited in [5] and [33], where the convexity property of the Wasserstein metric in the context of model identification was utilized. In this paper, we incorporate the CDT, a new transformation technique, along with the Wasserstein distance so that the estimation problem becomes a linear least squares problem in the transform space.

In a prior work [8] the cumulative distribution transform (CDT) was introduced as a useful means of modeling and subsequently classifying observed data. The CDT is a fundamentally nonlinear mapping of the *locations* of the signal values with respect to a particular reference. Put another way, computations performed in the CDT domain alter the independent variable of the signal(s) to produce a desired effect (e.g., matching one signal to another). The advantages of the CDT include its invertibility, ease of computation, and its ability to render certain classification problems linearly separable in transform space (see again, [8]).

C. Outline and overview of contributions

The main contribution of this paper is to describe a Wasserstein distance minimization-approach to parametric signal estimation. The mathematical approach is aided by the CDT [8] to help develop a generic closed form solution to the problem. The solution is general enough to encompass a variety of mass (signal) transport phenomena. In the following sections, we briefly review the definitions of the Wasserstein distance and the CDT, derive two important lemmas to formulate an expression for the estimator, discuss details of implementation, and introduce a strategy for mitigating the influence of noise on the estimates. We will conclude with both numerical and experimental examples comparing the CDT based estimator against some standard techniques. In section VII, we present a source localization example where the location of a crack is determined on a metal plate using the estimated TDOAs of the acoustic signals received by the sensors. Our analysis shows that in most cases the accuracy is improved when using the proposed technique, but most importantly, in all cases the computational cost is *orders of magnitude* lower than competing methods.

D. Note about notation

Throughout the manuscript, we deal with real signals s,r,z etc. assuming these to be square integrable in their respective 1D domains. That is, we assume that $\int_{\Omega_s}|s(t)|^2 dt < \infty$, where Ω_s is the domain over which s is defined. In addition, we at times make use of the common

notation: $\|s\|^2 = < s, s > = \int_{\Omega_s} |s(t)|^2 dt$. Some necessary symbols used throughout the manuscript are described in Table I.

II. Estimation as a Transport Problem

First, let $g_p(t)$ be a differentiable and strictly increasing mapping (i.e. $g'_p = dg_p/dt > 0$) between $\Omega = [0, 1]$ and Ω_z, where $\boldsymbol{p}$ refers to a parameter vector. It is easy to see that g_p is one-to-one and hence invertible. For example, polynomials $g_p(t) = \sum_{k=0}^{K-1} p_k t^k$ of different degrees will be used in the estimation problem described in this paper. This polynomial is able to capture events such as time delay and dispersion in the physics of wave propagation. Moreover, such polynomial model of the transformation $g_p(t)$ is commonly used in many signal processing applications [34] [35] [36]. In applications where $g_p(t)$ is unknown, polynomial approximations are often used to model the transformation [37].

The goal in our estimation problem is to then find the parameter vector $\boldsymbol{p}$ that generated some measured, normalized data $r(t) = B(z_g(t) + \eta(t))$ where $\eta(t)$ is a noise process (see again section V). Typical estimation techniques try to solve this problem by finding the parameters of a model, e.g. $s_g(t) = g'_p(t)s(g_p(t))$ that best matches the measured signal $r(t)$ [38] [39]. Alternatively, this problem can also be stated as finding the parameter vector $\boldsymbol{p}$ such that some measure of a 'match' between $r_f(t) = f'_p(t)r(f_p(t))$ and $s(t)$ is maximized, where $f_p(t) = g_p^{-1}(t)$. In this paper, we adopt the alternative approach as it helps generating closed form solution (discussed in section IV) for our estimation problem even when $g_p(t)$ is a higher order polynomial. Here we propose to solve the signal estimation problem by finding the parameters of $g_p(t)$ such that the Wasserstein distance [9] between $r_f(t)$ and $s(t)$ is minimized:

$$W^2(r_f, s) = \inf_h \int_{\Omega_s} |h(u) - u|^2 s(u) du \tag{4}$$

where $W(.,.)$ is the Wasserstein distance between two PDFs and

$$\int_{\inf(\Omega_r)}^{h(t)} f'_p(u) r(f_p(u)) du = \int_{\inf(\Omega_s)}^{t} s(u) du \,. \tag{5}$$

Thus, we have implicitly defined a "match" as the minimum distance $h(u) - u$, for all possible $h(\cdot)$, over which the original signal values $s(t)$ must be moved in order to form $r_f(t)$. The quantity (4) features prominently in the field of optimal mass transport where the minimizer is used to define the optimal transport map $h(u)$ for moving the "mass" $s(u)$ over a distance $h(u) - u$ [9].

We note that because we are looking at 1D signals $s(t)$, there is only one h satisfying the equation above. Moreover, by Lemma III.1 (see next section) we will never need to explicitly compute h, instead, its influence is embedded in the respective CDTs of the relevant signals.

In what follows we demonstrate the benefits of defining the cost function in this manner for the parameter estimation problem.

III. The Cumulative Distribution Transform

In this section we show that using the cumulative distribution transform (CDT) [8] we can derive a solution to the signal estimation problem expressed in equation (4) above. Let $s_0(y)$, $y \in \Omega_{s_0}$ define a reference signal pattern defined on the domain Ω_{s_0} which is in general different from the signal domain Ω_s. Without loss of generality we use $s_0(y) = 1$ and $\Omega_{s_0} = [0, 1]$ in this paper. The transform of $s(t)$ is then defined to be the function $\hat{s}(y)$ that solves

$$\int_{\inf(\Omega_s)}^{\hat{s}(y)} s(u)du = \int_{\inf(\Omega_{s_0})}^{y} s_0(u)du\,. \tag{6}$$

Now define the cumulative distribution functions (CDFs) $S(t) = \int_{-\infty}^{t} s(u)du$ and $S_0(y) = \int_{-\infty}^{y} s_0(u)du$. Note that because $s(t), s_0(y) > 0$ for $t \in \Omega_s$, $y \in \Omega_{s_0}$, it follows that S, S_0 are one to one continuous maps. Furthermore, if s, s_0 are continuous, S, S_0 will be differentiable [8]. Therefore, an alternative expression for $\hat{s}$ is

$$\hat{s}(y) = S^{-1}(S_0(y))\,. \tag{7}$$

The CDT is therefore seen to inherit the domain of the reference signal. Moreover, given our particular choice of reference, $s_0(y) = 1, S_0(y) = y$ and $\hat{s}(y) = S^{-1}(y)$. That is to say, *the CDT is the inverse of the cumulative distribution function of* $s(t)$. This definition is similar to the *Quantile Function* [40] [41] in statistics, although the similarity does not hold if non-uniform reference $s_0(y)$ is used to calculate the CDT. The inverse formula can then be defined as

$$s(t) = (\hat{s}^{-1})'(t)s_0(\hat{s}^{-1}(t)) \tag{8}$$

where $\hat{s}^{-1}(t) = S(t)$. Fig. 2 illustrates the process of calculating the CDT for a normalized, strictly positive quantity $s(t)$ and uniform reference signal $s_0(y)$. Physically, the CDT is a coordinate transformation acting on the independent variable (in this case time) in such a way as to preserve the total signal energy while morphing the distribution $s_0(y)$ into the distribution $s(t)$. Note that this definition is slightly different from that used in [8] where the CDT was defined in terms of the coordinate deviation $\hat{s}(y) - y$. In summary, the CDT and inverse CDT map continuous positive PDFs to diffeomorphism, and vice versa [8].

Given these definitions, we can now describe the proposed cost function (4) in the CDT domain. The following lemma [4] helps link the Wasserstein distance between s and r, and $\|\hat{s} - \hat{r}\|_{\ell_2}^2$

Lemma III.1.

Let $\hat{s}$ *and* $\hat{r}$ *be the CDTs of s and r, respectively. We then have that* $W^2(s, r) = \|\hat{s} - \hat{r}\|_{L_2}^2$.

Proof: Given, $\hat{s}$ and $\hat{r}$ are the CDTs of s and r, respectively. That is:

$$\hat{s}'(y)s(\hat{s}(y)) = \hat{r}'(y)r(\hat{r}(y)) = s_0(y)$$

Let $h'(t)s(h(t)) = r(t)$. Plugging $t = \hat{r}(y)$ into this equation, we have $r(\hat{r}(y)) = h'(\hat{r}(y))s(h(\hat{r}(y)))$ and $\hat{s}(y) = h(\hat{r}(y))$.

$$\begin{aligned} W_2^2(s,r) &= \int_{\Omega_r} (h(u)-u)^2 r(u)du \\ &= \int_{\Omega_{s_0}} (h(\hat{r}(y)) - \hat{r}(y))^2 \hat{r}'(y) r(\hat{r}(y))dy \\ &= \int_{\Omega_{s_0}} (\hat{s}(y) - \hat{r}(y))^2 s_0(y)dy \\ &= \|\hat{s} - \hat{r}\|_{L_2}^2 \end{aligned}$$

The lemma above simply states that for 1D signals which are PDFs, the CDT naturally embeds the Wasserstein distance. As in the computations all the signals are discrete, L_2 (norm of functions on real line) will be replaced by $_2$ (norm of sequences) in what follows. Therefore, in discrete cases we have,

$$W^2(s,r) = \|\hat{s} - \hat{r}\|_{\ell_2}^2$$

In addition, we also have the following useful functional composition lemma.

Lemma III.2.

Let $\hat{r}$ and $\hat{r}_f$ be the CDTs of signals r and r_f respectively, where $r_f = f_p' r \circ f_p$ The CDT of r_f is then given by $\hat{r}_f = f_p^{-1} \circ \hat{r}$.

Proof: Consider again the reference signal $s_0(y)$ and the signal $r(t)$. The relationship between these two signals can be defined in terms of CDT as,

$$\int_{\inf(\Omega_r)}^{\hat{r}(y)} r(u)du = \int_{\inf(\Omega_{s_0})}^{y} s_0(u)du \tag{9}$$

Similarly, can we relate $r_f(t)$ and $s_0(y)$ via

$$\int_{\inf(\Omega_{r_f})}^{\hat{r}_f(y)} r_f(u)du = \int_{\inf(\Omega_{s_0})}^{y} s_0(u)du\,. \tag{10}$$

We replace $r_f(t)$ with $f_p'(t)r(f_p(t))$ (where $f_p(t) = g_p^{-1}(t)$), in which case

$$\int_{f_p^{-1}(\inf(\Omega_r))}^{\hat{r}_f(y)} f_p'(u) r(f_p(u))du = \int_{\inf(\Omega_{s_0})}^{y} s_0(u)du\,. \tag{11}$$

Applying the change of variables $f_p(u) = y$ and $dy = f'_p(u)du$ to the left hand side,

$$\int_{\inf(\Omega_r)}^{f_p(\hat{r}_f(y))} r(y)dy = \int_{\inf(\Omega_{s_0})}^{y} s_0(u)du\,. \tag{12}$$

From Eqn. (9) and (12) it can be stated,

$$\int_{\inf(\Omega_r)}^{f_p(\hat{r}_f(y))} r(y)dy = \int_{\inf(\Omega_r)}^{\hat{r}(y)} r(u)du\,. \tag{13}$$

For this statement to be true, the upper bound of the left hand side must be equal to the upper bound of the right hand side, i.e. $f_p(\hat{r}_f(y)) = \hat{r}(y)$. Since $f_p(t) = g_p^{-1}(t)$ and $g_p(t)$ are invertible, we can finally write $\hat{r}_f = f_p^{-1} \circ \hat{r} = g_p \circ \hat{r}$.

Simply stated, the CDT composition lemma says that changes along the independent variable (e.g., shifts in time $t-\tau$ or dispersions ωt) become changes in the dependent variable in transform domain (refer to Fig. 3b). Fig. 3 illustrates the relationships between the signals and the transforms for a signal undergoing parametric change. Each of the constituent CDTs transforms their respective signals into the reference signal $s_0(y)$. Similarly, the mapping $f_p(t)$ transforms $r(t)$ into $r_f(t)$.

A. Numerical Implementation of the CDT

Recall that the CDT is defined for continuous-time signals in contiguous, finite domain. Here we describe the numerical method for approximating the CDT given discrete data. As the CDT $\hat{s}(y)$ is the inverse of the CDF of $s(t)$ for a particular choice of reference signal ($s_0(y) = 1$ for $y \in [0,1]$), we need to compute the cumulative function first.

Let $\boldsymbol{s} = [s_1, s_2,...,s_N]^T$ be a N-point discrete-time signal, where $s[n] = s_n$, $\forall n = 1, 2,...,N$ is the n^{th} sample of $\boldsymbol{s}$. Then the numerical approximation of the cumulative function is given by,

$$S[n] = \sum_{i=1}^{n} s[n], \qquad n = 1, 2, \ldots, N \tag{14}$$

The CDT is then calculated by taking the inverse of the CDF using interpolation.

IV. Signal Estimation in CDT Domain

This section will demonstrate the use of CDT in estimating signal parameters. We will highlight the relevance of the transform with respect to time delay, linear dispersion, and quadratic dispersion. For each case, we will leverage lemmas III.1 and III.2 above, so that the cost function in (4) becomes

$$W^2(r_f, s) = \|f_p^{-1} \circ \hat{r} - \hat{s}\|_{\ell_2}^2 = \|g_p \circ \hat{r} - \hat{s}\|_{\ell_2}^2 . \tag{15}$$

It is evident that when $g_p(t)$ is a monotonically increasing polynomial, the problem above is simply a linear least squares problem. The advantage of signal estimation in the CDT domain is that only the function $g_p(t)$ needs to be computed before proceeding to the estimation problem. We note the following specific examples that can be derived from the above.

A. Time delay estimation

In the time delay estimation problem $g_p(t) = t - \tau$, hence the cost function (15) becomes,

$$W^2(r_f, s) = \|\hat{r} - \tau - \hat{s}\|_{\ell_2}^2 .$$

The translation value τ that minimizes the equation above is then given by:

$$\tilde{\tau} = \frac{1}{|\Omega_{s_0}|} \int_{\Omega_{s_0}} [\hat{r}(u) - \hat{s}(u)] du \tag{16}$$

Note that, as already mentioned in [4] the problem above is convex on τ, hence a global minimizer is possible and given in closed form. Furthermore, utilizing the fact that the center of mass of s can be estimated by $\mu_s = \int_{\Omega_s} t s(t) dt = \int_{\Omega_s} 1 - S(t) dt$ one can also show that $\frac{1}{|\Omega_{s_0}|} \int_{\Omega_{s_0}} \hat{s}(u) du = \mu_s$ and thus the solution of the time delay problem is also given by

$$\tilde{\tau} = \mu_r - \mu_s .$$

B. Linear Dispersion Estimation

In the linear dispersion problem we have that $g_p(t) = \omega t$, and thus

$$W^2(r_f, s) = \|\omega \hat{r} - \hat{s}\|_{\ell_2}^2 . \tag{17}$$

This problem is convex on ω and possesses closed form solution. The minimizer for the equation above (following linear least squares on ω) is

$$\tilde{\omega} = \frac{< \hat{s}, \hat{r} >}{\|\hat{r}\|_{\ell_2}^2} \tag{18}$$

where $< \cdot, \cdot >$ denotes the inner product.

C. Time delay and Linear Dispersion estimation

In the joint estimation of time delay and linear dispersion we have that $g_p(t) = \omega t - \tau$, thus $g_p \circ \hat{r} = \omega\hat{r} - \tau$. Hence, the cost function (15) becomes,

$$W^2(r_f, s) = \|\omega\hat{r} - \tau - \hat{s}\|_{\ell_2}^2 = \|\alpha\hat{r} + \beta - \hat{s}\|_{\ell_2}^2 \tag{19}$$

where, $\alpha = \omega$, and $\beta = -\tau$. Once again, this is a linear least squares problem, from which ω and τ can readily be recovered. The closed form solution to this problem is given by,

$$[\tilde{\alpha}, \tilde{\beta}]^T = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\hat{s} \tag{20}$$

where $\mathbf{X} \equiv [\overrightarrow{r}, \overrightarrow{\mathbf{1}}]$ is an $N \times 2$ matrix.

D. Quadratic dispersion estimation

In the quadratic dispersion estimation problem we have $g_p(t) = \kappa t^2$, so $g_p \circ \hat{r} = \kappa\hat{r}^2$. Consequently

$$W^2(r_f, s) = \|\kappa\hat{r}^2 - \hat{s}\|_{\ell_2}^2. \tag{21}$$

Again we have convexity with the solution being

$$\tilde{\kappa} = \frac{< \hat{s}, \hat{r}^2 >}{\|\hat{r}^2\|_{\ell_2}^2} \tag{22}$$

E. Quadratic dispersion with time delay

The quadratic dispersion with time delay can be expressed as $g_p(t) = \kappa t^2 - \tau$. Therefore the Wasserstein distance is:

$$W^2(r_f, s) = \|\kappa\hat{r}^2 - \tau - \hat{s}\|_{\ell_2}^2 = \|\alpha\hat{r}^2 + \beta - \hat{s}\|_{\ell_2}^2 \tag{23}$$

Similar to joint time delay and linear dispersion estimation described in IV-C, this problem is also convex and possesses closed form solution which is given by

$$[\tilde{\alpha}, \tilde{\beta}]^T = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\hat{s} \tag{24}$$

where $\mathbf{X} \equiv \left[\overrightarrow{\hat{r}^2}, \overrightarrow{\mathbf{1}}\right]$, $\alpha = \kappa$, and $\beta = -\tau$.

F. Higher order polynomial

For any general polynomial, i.e. $g_p(t) = \sum_{k=0}^{K-1} p_k t^k$, $g_p \circ \hat{r} = \sum_{k=0}^{K-1} p_k \hat{r}^k$. Thus, the cost function in (15) becomes,

$$W^2(r_f, s) = \| \sum_{k=0}^{K-1} p_k \hat{r}^k - \hat{s} \|_{\ell_2}^2 \tag{25}$$

which can be stated as a linear least squares problem,

$$\widetilde{\overrightarrow{p}} = \underset{\overrightarrow{p}}{\operatorname{argmin}} \| \mathbf{X}\overrightarrow{p} - \overrightarrow{\hat{s}} \|_{\ell_2}^2 \tag{26}$$

Where $\overrightarrow{p} = [p_0, p_1, \ldots, p_{K-1}]^T, \mathbf{X} \equiv \left[\overrightarrow{\mathbf{1}}, \overrightarrow{\hat{r}}, \overrightarrow{\hat{r}^2}, \ldots, \overrightarrow{\hat{r}^{K-1}}\right]$. The Hessian of (26) is $2\mathbf{X}^T\mathbf{X}$ which is positive semi-definite. Therefore the estimation problem described in equation (26) is convex. Moreover, the columns of matrix $\mathbf{X}$ are linearly independent, that means Hessian is positive definite and $\mathbf{X}^T\mathbf{X}$ is invertible. Hence, (26) possesses closed form solution which is given by,

$$\widetilde{\overrightarrow{p}} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\hat{s} \tag{27}$$

In this section, we have shown that the estimation problem, while non-linear in time domain, can be transformed into a linear least squares problem with closed form solution using CDT. In the next section we will address the influence of noise in the estimation process and a strategy to mitigate it.

V. Signal Estimation in Noise

In the previous sections, we defined the CDT, provided the relationship between the CDTs of signals related by a transformation of the independent variable, and then demonstrated linearity of the Wasserstein cost function with respect to the signal parameters that define several such transformations. These relationships were derived without explicit consideration of the corrupting noise source and how it influences the associated estimation problem.

In this section, we consider the impact of additive Gaussian noise on the CDT and on the subsequent parameter estimation. Assume the received signal is corrupted by zero mean, i.i.d Gaussian noise values, $\eta(t) \sim \mathcal{N}(0, \sigma^2)$ so that the measured data are $z_\eta(t) = z_g(t) + \eta(t)$ and

$$r(t) = B(z_\eta)(t) = \frac{\left(z_g(t) + \eta(t)\right)^2}{\|z_g(t) + \eta(t)\|_{\ell_2}^2}. \tag{28}$$

The normalization therefore results in three terms, two of which involve the signal noise. This additional signal “mass” alters the Wassserstein distance and biases the resulting signal parameter estimates. In what follows we propose a simple solution for removing the influence of the additive noise directly in the signals’ CDFs.

A. Noise corrected CDF

Using Eqn. (28) as a starting point, in expectation the effects of additive, zero-mean Gaussian noise on the CDF are modelled as (detailed in the supplementary material)

$$E[R(t)] = \frac{\mathscr{E}_z S_g(t) + \sigma^2(t - t_1)}{\mathscr{E}_z + \sigma^2(t_N - t_1)}, \quad t_1 \le t \le t_N . \tag{29}$$

Here, $S_g(t)$ and $R(t)$ are the CDFs associated with $s_g(t) = B(z_g)(t)$ and $r(t) = B(z_\eta)(t)$ respectively and the term $\mathscr{E}_z = \|z_g(t)\|^2_{\ell_2}$ is the total energy of the noise free signal. An expression for the noise corrected CDF in expectation is then obtained from Eqn. (29) as

$$\tilde{S}_g(t) = \frac{E[R(t)]\left\{\mathscr{E}_z + \sigma^2(t_N - t_1)\right\} - \sigma^2(t - t_1)}{\mathscr{E}_z} \tag{30}$$

Where $\mathscr{E}_z + \sigma^2(t_N - t_1)$ is the expected energy of noisy signal (see supplementary material). Alternatively we can define the signal-to-noise ratio $SNR = \mathscr{E}_z/\sigma^2(t_N - t_1)$ in which case (30) becomes

$$\tilde{S}_g(t) = \frac{E[R(t)][SNR + 1] - \frac{t - t_1}{t_N - t_1}}{SNR}, \quad t_1 \le t \le t_N \tag{31}$$

In short, the influence of additive, i.i.d noise is seen as the addition of a constant slope to the CDF. Moreover, under our chosen normalization scheme (3), this slope is the noise variance. Thus, a simple strategy for denoising in the CDT domain is to first estimate σ^2 using a "noise only" portion of the signal, and then apply Eqn. (30). This method is effectively filtering the signal in the CDF domain.

To illustrate, consider a Gaussian pulse subject to the coordinate transformation $g_p(t) = \omega t - \tau$ with $\omega = 2$, $\tau = 2$. The noise free input signal PDF is therefore $s(t) = A^2 \exp\left(-t^2/2b_w^2\right)$. which, after the transformation, becomes $s_g(t) = A^2\omega \exp\left(-(\omega t - \tau)^2/2b_w^2\right)$. For this example the corresponding CDFs can be determined analytically and are shown in Fig. 4 for $A = b_w = 1$. The noisy signal was taken as $z_g(t) = s_g^{1/2}(t) + \eta(t)$ where each $\eta(t) \sim \mathcal{N}\left(0, \sigma^2\right)$ with $\sigma = 0.15$. The associated CDF $R(t)$ and noise corrected version $(\tilde{S}_g(t))$ are also shown. The SNR for this example was taken as $SNR = \mathscr{E}_z/\sigma^3(t_N - t_1)) = 4$. The noise corrected $\mathrm{CDF}(\tilde{S}_g(t))$ is seen to match almost exactly the true $\mathrm{CDF}\left(S_g(t)\right)$.

B. Distribution of the CDT

Even after the expected noise is removed via (30), there will remain residual fluctuations that will impact our parameter estimates. The joint PDF of the values $\hat{r}(y)$ will dictate the degree to which the cost function (15) can be expected to produce good estimates.

Following the derivation described in the supplementary material, the distribution for the CDT values associated with each observation in the signal $z_\eta(t_k)$, $k = 1 \cdots N$ is shown to be approximated by the PDF

$$p_{\hat{R}_k}(\hat{r}_k) = \frac{e^{-\frac{(S(\hat{r}_k) - S(\hat{s}_k))^2}{2\Sigma^2(\hat{s}_k)}}}{\sqrt{2\pi\Sigma(\hat{s}_k)}} \frac{\partial S(\hat{r}_k)}{\partial \hat{r}_k}, \quad 1 \le k \le N. \tag{32}$$

where the variance

$$\Sigma^2(t_k) = \frac{\sigma^4(2k + 4\lambda_k)}{\mathscr{E}_{z_\eta}^2} \tag{33}$$

is a function of the total noisy signal energy, $\mathscr{E}_{z_\eta} = \sum_k z_\eta^2(t_k)$, and the cumulative sum of the noise-free signal, $\lambda_k = \sigma^{-2}\sum_{i=1}^{k} z^2(t_i)$. Note that both mean and variance in (32) are evaluated at the fixed, noise-free CDT values $\hat{s}_k$ (as opposed to the independent variable $\hat{r}_k$). Both can be obtained by simply interpolating the functions $S(t_k), \Sigma(t_k) \rightarrow S(\hat{s}_k), \Sigma(\hat{s}_k)$. The resulting distribution is a peaked function that is largely symmetric and centered on the noise-free CDT $\hat{s}(y)$.

As an example, consider signals for which the CDF is well-approximated by the logistic model

$$S(t) = \frac{1}{1 + e^{-at + b}}. \tag{34}$$

This is the exact CDF for the logistic distribution, however has also been used to model other CDFs [42]. For appropriate choice of *a, b* this model almost exactly captures the behavior of the CDF for signals such as those shown in Fig. 4. Using the general expression (32) with this logistic model, we can readily obtain the distribution of CDT values.

We also examined the empirical distribution of the $\hat{r}(y)$ via Monte Carlo simulation. To this end we simulated 1500 realizations of the Gaussian pulse of the previous example, using the parameter values $\omega = 2$, $\tau = 9$, $A = 1$, $b_w = 2$. The associated normalized signals $r(t)$ each consisted of $N = 200$ points sampled at $dt = 0.05s$ and where each additive noise value was taken independently from $\mathcal{N}(0, \sigma^2)$ with $\sigma^2 = 0.023$. The CDT $\hat{r}(y)$ was then estimated for each realization. The resulting empirical PDF of the CDT values is shown in Fig. 5 along with the predicted distribution given by Eqn. (32) with $a = 1.68$, $b = 7.54$ and plotted as a function of the dimensionless reference variable $y_k = k/N$, $k = 1 \cdots N$.

The analytical distribution captures precisely the behavior shown in empirically in Fig. 4, that is to say, the distributions for CDT values near $y = 0, 1$ are skewed while those near the middle of the CDT curve are approximately normally distributed. The difference between simulation and theory near $y = 1$ is due to the fact that the inverse logistic (logit)

transformation is only valid for $0 < R(t) < 1$, however our derivation (described in the supplementary material) assumed constant noisy signal energy, leaving the possibility of values $R(t) > 1$ near the end of the signal.

Nonetheless, the uncertainty in the estimate is indeed well-approximated by a Gaussian distribution as evidenced by Fig. 5 and the functional form of (32). Moreover, for signals in additive, i.i.d. Gaussian noise, the proposed cost function (4) yields a maximum likelihood estimate (MLE) of the parameters. As such, the estimate is guaranteed to reach the Cramer-Rao lower bound (CRLB) asymptotically as $N \rightarrow \infty$. The CRLB places a lower bound on the covariance of the parameter estimates and is given by

$$C(p_i, p_j) = \mathbf{F}(\overrightarrow{p})^{-1} \tag{35}$$

where $\mathbf{F}(\overrightarrow{p})$ is the Fisher Information Matrix (FIM). For signals $z_g(t)$ in additive, i.i.d. Gaussian noise the FIM is defined as [1]

$$\begin{aligned} \mathbf{F}(\overrightarrow{p}) &\equiv F_{ij} \\ &= \frac{1}{\sigma^2 \Delta t}\left(\int_{t_s}^{t_f} \frac{\partial z_g(t)}{\partial p_i} dt\right)\left(\int_{t_s}^{t_f} \frac{\partial z_g(t)}{\partial p_j} dt\right). \end{aligned} \tag{36}$$

Carrying out the integrals and subsequent inversion for the Gaussian pulse of the previous example then yields the full covariance matrix (35). We are most interested in the diagonal elements $C(p_i, p_i)$, $i = 1 \cdots P$ as these represent the variances of the associated parameter estimates. In what follows we compare our estimates to the CRLB for several of the cases described in the prior section.

VI. Impact of SNR on Quality of the Estimator

Here we explore the quality of the CDT-based estimation procedure through a series of numerical experiments. Specifically, we compare the quality of the various estimators described in section (IV) in terms of mean square error (MSE) as a function of SNR.

The signal of interest is taken as the apodized sinusoid

$$z(t) = Ae^{-(t-t_c)^2/(2b_w^2)}\sin(2\pi f t) \tag{37}$$

of width b_w and frequency f. The SNR can be well-approximated by

$$SNR = \frac{A^2\sqrt{\pi} b_w}{2\sigma^2 T} \tag{38}$$

so long as the signal length T is large enough to include the entire non-zero portion of the pulse envelope. In our examples we will take $A = b_w = f = 1$ and set $t_c = 0$. We are again assuming that the received signals are corrupted by zero mean additive Gaussian noise, $\eta(t) \sim \mathcal{N}(0, \sigma^2)$.

A. Time Delay Estimation

From Eqn. (16) one can estimate the delay as the difference in the average values of the CDTs $\hat{r}, \hat{s}$ taken over the domain $\Omega_{s_0} = [0, 1]$. Computationally, we have simply

$$\tilde{\tau} = \frac{1}{N} \sum_{i=1}^{N} (\hat{r}(u_i) - \hat{s}(u_i)) \tag{39}$$

where the CDTs are defined on the discrete grid $u_i = i/N$, $i = 1 \cdots N$. These estimates can then be compared to those obtained via the more familiar cross-correlation estimator applied in the time domain [4]. To this end we simulated 1000 realizations of the signal $z_\eta(t)$ for varying noise levels and a delay of $\tau = 0.2575$ *s*. The linear dispersion was fixed at $\omega = 1$.

To evaluate the performance of the estimator we compute mean squared error (MSE) and compare the results with cross-correlation (XC) based estimator. Although cross-correlation is known to be an MLE for delay estimates in additive Gaussian noise [21], it is a discrete estimator. To implement a continuous delay estimator, an optimization problem is designed that provides maximum likelihood estimates,

$$\tilde{\tau} = \underset{\tau}{\text{argmax}} \; - \sum_{i=0}^{N-1} z_\eta(t_i) z(t_i - \tau) \tag{40}$$

To solve this optimization problem we exploit a gradient based nonlinear programming solver *fmincon* in MATLAB [43]. As *fmincon* solves minimization problems, equation (40) is written as,

$$\tilde{\tau} = \underset{\tau}{\text{argmin}} \; - \sum_{i=0}^{N-1} z_\eta(t_i) z(t_i - \tau) \tag{41}$$

As equation (41) is a non-convex problem, this gradient based solver may get stuck in local minima. To resolve this issue, another MATLAB function *GlobalSearch* can be integrated, which repeatedly runs local solver *fmincon* with random starting point to generate global optimal solution. Therefore, in our experiments two approaches are adopted to solve this optimization problem: (i) using *fmincon* only, and (ii) using *GlobalSearch* and *fmincon* together. To compare with a subspace based method, the ESPRIT (estimation of signal parameters via rotational invariance techniques) based time delay estimation technique has been implemented following the approach described in [23]. The MSE for different delay estimators are plotted in Fig. 6. We also plot the CRLB as baseline. The plot shows that the performance of the proposed CDT based estimator is similar to the cross-correlation and ESPRIT estimators, although none of the techniques have reached the CR bound. The MLE obtained via global optimum search reaches the bound, but the estimation using local solver without *GlobalSearch* shows very poor performance.

B. Joint Estimation of Time Delay and Linear Dispersion

In this example we considered the joint estimation problem for both time delay $\tau = 0.2575s$ and linear dispersion (time scale) $\omega = 0.75$. Again, the estimation problem (19) possesses the closed form solution given in section IV-C. The MSE of the joint delay and linear dispersion estimates for different estimators are plotted in Fig. 7. For comparison, the cross-correlation and ESPRIT based estimators are used again to estimate the delay parameter only. In this case, both XC and ESPRIT estimators perform poorly as these techniques do not take linear dispersion into account. Another subspace based method, the MUSIC (multiple signal classification) algorithm discussed in [23], has been implemented to estimate time delay when both delay and dispersion are present. The time delay estimates from the ESPRIT estimator have been used in the initialization stage of this algorithm. Although this method shows very good performance in significantly high SNR (15 dB), the proposed technique outperforms it in highly noisy cases. Again, as this algorithm requires narrowband approximation of the signal, it gives incorrect estimates of the linear dispersion parameter (ω) for transient signals used in these experiments. Hence, only time delay (τ) estimates are reported for MUSIC algorithm in Fig. 7. For joint time delay and linear dispersion estimation, the another commonly used approach is to locate the peak of Wide-band Ambiguity Function (WBAF) of the received signal [28] [29] [30]. The WBAF between the measured signal $z_\eta(t)$ and the known signal $z(t)$ is given by [28],

$$A_{z_\eta, z}(\tau, \omega) = \sqrt{\omega} \int_{-\infty}^{\infty} z_\eta(t) z^*(\omega t - \tau) dt \tag{42}$$

where (*) denotes the complex conjugation which does not have any impact in our experiments as real valued signals are used. Then the joint estimates of τ and ω are given by,

$$\tilde{\tau}, \tilde{\omega} = \underset{\tau, \omega}{\operatorname{argmax}} \ |A_{z_\eta, z}(\tau, \omega)|^2$$

which can also be written as,

$$\tilde{\tau}, \tilde{\omega} = \underset{\tau, \omega}{\operatorname{argmin}} \ -|A_{z_\eta, z}(\tau, \omega)|^2 \tag{43}$$

Similar to time delay estimation discussed in VI-A, both the local and global solvers have been exploited to estimate optimum τ and ω using equation (43). While global solver performs better than the local solver, the proposed CDT based estimator yields better estimation of the delay parameter (τ) than both solvers (Fig. 7). In case of linear dispersion (ω) estimation, it does not outperform the global solver, but the results are still competitive. In both (τ and ω estimates) cases, CDT estimator outperforms local solver based estimator.

The concept of convexity can help understand the superior performance of the proposed technique over the WBAF based estimator. Fig. 8 illustrates that in CDT based estimator we are dealing with a convex problem while $-|A_{z_\eta z}(\tau, \omega)|^2$ is clearly non-convex with several local minima. Although *GlobalSearch* is designed to handle this kind of problem, it is not always accurate to find the global minimal point. It should also be noted that global solver is

computationally very expensive as it runs several local solvers. The solution provided by the CDT based estimator, on the other hand, is closed form, hence computational cost is very low. Even (b) running single local solver takes more time than the proposed estimator. The MUSIC algorithm also requires an iterative search over parameter space. Moreover, it uses the result of ESPRIT algorithm for initialization, which contributes to the computational cost. As a result, MUSIC based estimation is also a computationally expensive technique relative to the proposed approach. Fig. 9 shows the average times taken by CDT, MUSIC, and WBAF based estimators to jointly estimate the time delay (τ) and the linear dispersion (ω) parameters.

C. Quadratic Dispersion & Delay

As a final illustration we consider the problem of jointly estimating both the quadratic dispersion coefficient (κ) and time delay (τ), i.e. $g_p(t) = \kappa t^2 - \tau$ with $\kappa = 0.5$ and $\tau = 1.2575s$. As discussed in section VI-C, the proposed estimation approach described by equation (23) possesses closed form solution which is given by equation (24). Fig. 10 shows the MSE of the estimates of time delay and quadratic dispersion coefficients jointly estimated using the proposed approach. The plots show that CDT based estimator could not reach CR bounds. But the performance of time delay estimation using proposed estimator is better than the cross-correlation based estimator, as cross-correlation does not correct the effect of quadratic dispersion.

VII. Application: Source Localization

As we have mentioned, the estimation approach we have described is appropriate for signals undergoing an invertible transformation of the independent variable, i.e, $z(t) \rightarrow g_p'(t)z(g_p(t))$. This is a reasonable model in situations where the signal becomes distorted as it propagates through a medium.

One such situation is the propagation of acoustic signals in solids. Fig. 11 shows a metal plate with crack emanating from the end of a horizontal “slot”. As the crack propagates it gives off acoustic emissions, loosely defined as a spatially localized release of energy. The result is a short elastic wave “pulse”, similar to those used in the preceding numerical examples. By measuring these pulses at different locations on the plate and estimating time difference of arrival one can in principle localize the source i.e., the crack tip. In this experiment we used four fiber-optic strain sensors arranged in a diamond pattern, (see Fig. 11) [44]. Sample time series from an acoustic emission event are also shown in Fig 11.

The challenge is that such signals are difficult to detect and are affected by more than just a time delay during propagation. For example, dispersion is known to influence such pulses during transit [45]. By including dispersion in the model we hypothesize an improved ability to estimate the time delay. Moreover, because this estimation problem reduces to linear least-squares in the CDT domain, the inclusion of this additional term incurs no computational penalty (see again Fig. 9) Using the estimation procedure outlined herein, we estimated time delay of arrival among the four sensors. Based 10 on the obtained delays, we then used the source localization algorithm described in [46] to estimate the location of the crack tip. Fig. 12 shows the results of these estimates for six different data sets. Specifically,

we show the cost function associated with the localization algorithm for a typical realization, along with the estimated minimum which should denote the location of the acoustic emission (i.e., the crack tip). The localization algorithm depends on delay estimates among the four sensors shown as black numbers. To obtain the required delay estimates we used the delay-only estimator (section VI-A) as well as the joint delay and linear dispersion estimator (section VI-B).

For each of the six data sets, the addition of linear dispersion in the signal model yielded a more well-defined cost function and provided modest improvement in localizing the source of the emission. In fact, one of the location estimates obtained using the “delay only” approach placed the crack tip at the edge of the plate. The source localization results are compared to those obtained using the cross-correlation and the WBAF gives off acoustic emission pulses which can be measured at different locations on the plate (right). The time delay of arrival between recorded pulses can then provide the location of the crack tip. based estimators. Fig. 12 shows that the proposed CDT based technique outperforms both the estimators in estimating the location of the crack tip.

VIII. Summary

We considered a class of signal estimation problems for which a positive valued signal is altered by a transformation of the independent variable. Such transformations are common in wave propagation, where the energy of a signal or field is modified by the medium through which it travels, but is ultimately conserved.

We proposed using the Wasserstein distance between the modified received signal and the model signal as a cost function for estimating the parameters that govern the transformation. The idea is to select those parameters that minimize the amount of work it takes to transform the model signal into the received signal. It was then shown that by using the cumulative distribution transform (CDT) the Wasserstein distance becomes a linear, convex function of the desired parameters and possesses a closed form (least squares) solution.

A series of numerical experiments were then conducted to assess the quality of the estimator. The CDT was found empirically to be approximately Gaussian distributed, hence the Wasserstein estimator is an approximate MLE for this class of problem. Indeed, the proposed estimator performs well in comparison to other approaches in terms of estimator MSE. Moreover, because the estimator is linear in CDT space, the computational cost is orders of magnitude lower than competing methods.

The noise model used in the aforementioned experiments is assumed to be zero mean, i.i.d. Gaussian. While not considered explicitly, we expect other noise models to yield similar results. The reason is that the CDF is the summation of random variables, and with enough such variables the central limit theorem is expected to be applicable. Thus, even under other additive noise models, our CDT distribution will likely remain Gaussian (see Eqn. 32); this is a topic of ongoing work.

The numerical experiments also demonstrated the proposed noise reduction method, which works by subtracting the i.i.d. noise cumulative distribution function (CDF) from the total noisy signal CDF prior to effecting the transformation into the CDT domain.

Finally, the estimator was used to localize the source of acoustic emissions in a thin metal plate. It was shown that including linear dispersion in the signal model offered modest improvement in the ability to localize the source without incurring a computational penalty.

IX. Conclusion

In this paper, we proposed a parametric signal estimation approach by minimizing Wasserstein distance between measured and model signals. This approach, aided by the use of the cumulative distribution transform [8], was shown to produce generic closed form solution to the estimation problem. Several numerical experiments showed that the proposed approach not only performs well in comparison to existing methods but also is significantly more computationally efficient compared to the competing methods.

In short, by using the CDT and the Wasserstein cost, one can easily and accurately estimate the parameters that govern the modification of signal energy during propagation. Future efforts will be aimed at gaining a deeper understanding of the statistical properties of the CDT and applying this technique in experiments focused on identifying properties of the medium through which a signal is traveling.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The work of AHMR, SL, and GKR was supported in part by NIH award R01 GM130825.

References

[1]. McDonough RN and Whalen AD, Detection of Signals in Noise, 2nd ed. San Diego: Academic Press, 1995.

[2]. Hassab JC, “On the weighting of time delay measurements for improved direction finding,” Journal of the Acoustical Society of America, vol. 78, no. 5, pp. 1658–1663, 1985.

[3]. Amar A, Leus G, and Friedlander B, “Emitter localization given time delay and frequency shift measurements,” IEEE Transactions on Aerospace and Electronic Systems, vol. 48, no. 2, pp. 1826–1837, 2012.

[4]. Nichols JM, Hutchinson MN, Menkart N, Cranch GA, and Rohde GK, “Time delay estimation via Wasserstein distance minimization,” Signal Processing Letters, vol. 26, no. 6, pp. 908–912, 2019.

[5]. Engquist B, Froese BD, and Yang Y, “Optimal transport for seismic full waveform inversion,” arXiv preprint arXiv:1602.01540, 2016.

[6]. Chen J, Chen Y, Wu H, and Yang D, “The quadratic wasserstein metric for earthquake location,” Journal of Computational Physics, vol. 373, pp. 188–209, 2018.

[7]. Proakis JG, “Digital communications,” 1983.

[8]. Park SR, Kolouri S, Kundu S, and Rohde GK, “The cumulative distribution transform and linear pattern recognition,” Applied and computational harmonic analysis, vol. 45, pp. 616–641, 2018.

[9]. Kolouri S, Park SR, Thorpe M, Slepcev D, and Rohde GK, "Optimal mass transport: Signal processing and machine-learning applications," IEEE Signal Processing Magazine, vol. 34, no. 4, pp. 43–59, 2017. [PubMed: 29962824]

[10]. Benamou J-D and Brenier Y, "A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem," Numerische Mathematik, vol. 84, pp. 375–393, 2000.

[11]. Nichols JM, Emerson TH, and Rohde GK, "A transport model for broadening of a linearly polarized, coherent beam due to inhomogeneities in a turbulent atmosphere," Journal of Modern Optics, vol. 66, no. 8, pp. 835–849, 2019.

[12]. Gosse L and Runborg O, "Finite moment problems and applications to multiphase computations in geometric optics," Communications in Mathematical Sciences, vol. 3, no. 3, pp. 373–392.

[13]. Rubinstein J and Wolansky G, "Geometrical optics and optimal transport," Journal of the Optical Society of America – A, vol. 32, no. 10, pp. 1817–1823, 2017.

[14]. Nichols JM, Emerson TH, Cattell L, Park S, Kanaev A, Bucholtz F, Watnik A, Doster T, and Rohde GK, "A transport-based model for turbulence-corrupted imagery," Applied Optics, vol. 57, no. 16, pp. 4524–4536, 2018. [PubMed: 29877400]

[15]. Achenbach JD, Wave propagation in elastic solids, ser. Applied Mathematics and Mechanics Series. Amsterdam: North-Holland Publishing Company, 1973.

[16]. Engquist B and Yang Y, "Seismic inversion and the data normalization for optimal transport," arXiv preprint arXiv:1810.08686, 2018.

[17]. Jacovitti G and Scarano G, "Discrete time techniques for time delay estimation," IEEE Transactions on Signal Processing, vol. 41, no. 2, pp. 525–533, 1993.

[18]. Jiang X, Zeng WJ, So HC, Rajan S, and Kirubarajan T, "Robust matched filtering in ℓ_p-space," IEEE Transactions on Signal Processing, vol. 63, no. 23, pp. 6184–6199, 2015.

[19]. Chen J, Benesty J, and Huang Y, "Performance of GCC- and AMDF-based time-delay estimation in practical reverberant environments," EURASIP Journal on Applied Signal Processing, vol. 1, pp. 25–36, 2005.

[20]. Benesty J, Huang Y, and Chen J, "Time delay estimation via minimum entropy," IEEE Signal Processing Letters, vol. 14, no. 3, pp. 157–160, 2007.

[21]. Rohde GK, Bucholtz F, and Nichols JM, "Maximum empirical likelihood estimation of time delay in independently and identically distributed noise," IET Signal Processing, vol. 8, no. 7, pp. 720–728, 2014.

[22]. Mars NJI and Arragon GWV, "Time delay estimation in nonlinear systems," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 29, no. 3, pp. 619–621, 1981.

[23]. Jakobsson A, Swindlehurst AL, and Stoica P, "Subspace-based estimation of time delays and Doppler shifts," IEEE Transactions on Signal Processing, vol. 46, no. 9, pp. 2472–2483, 1998.

[24]. Friedlander B, "High resolution Doppler and delay estimation," in 2013 Asilomar Conference on Signals, Systems and Computers IEEE, 2013, pp. 181–185.

[25]. Gu J-F, Moghaddasi J, and Wu K, "Delay and Doppler shift estimation for OFDM-based radar-radio (RadCom) system," in 2015 IEEE international wireless symposium (IWS 2015). IEEE, 2015, pp. 1–4.

[26]. Tsatsanis MK and Giannakis GB, "Subspace methods for blind estimation of time-varying fir channels," IEEE Transactions on Signal Processing, vol. 45, no. 12, pp. 3084–3093, 1997.

[27]. Colonnese S, Rinauro S, and Scarano G, "Generalized method of moments estimation of location parameters: Application to blind phase acquisition," IEEE Transactions on Signal Processing, vol. 58, no. 9, pp. 4735–4749, 2010.

[28]. Jin Q, Wong KM, and Luo ZQ, "The estimation of time delay and doppler stretch of wideband signals," IEEE Transactions on Signal Processing, vol. 43, no. 4, pp. 904–916, 1995.

[29]. Niu XX, Ching PC, and Chan YT, "Wavelet based approach for joint time delay and doppler stretch measurements," IEEE Transactions on Aerospace and Electronic Systems, vol. 35, no. 3, pp. 1111–1119, 1999.

[30]. Tao R, Zhang WQ, and Chen EQ, "Two-stage method for joint time delay and doppler shift estimation," IET Radar, Sonar & Navigation, vol. 2, no. 1, pp. 71–77, 2008.

[31]. Villani C, Topics in optimal transportation. American Mathematical Soc, 2003, no. 58.

[32]. Engquist B and Froese BD, “Application of the wasserstein metric to seismic signals,” arXiv preprint arXiv:1311.4581, 2013.

[33]. Yang Y, Engquist B, Sun J, and Hamfeldt BF, “Application of optimal transport and the quadratic Wasserstein metric to full-waveform inversion,” Geophysics, vol. 83, no. 1, pp. R43–R62, 2018.

[34]. Hassanien A, Vorobyov SA, and Gershman AB, “Moving target parameters estimation in noncoherent mimo radar systems,” IEEE Transactions on Signal Processing, vol. 60, no. 5, pp. 2354–2361, 2012.

[35]. Nichols J, Currie M, Bucholtz F, and Link W, “Bayesian estimation of weak material dispersion: theory and experiment,” Optics express, vol. 18, no. 3, pp. 2076–2089, 2010. [PubMed: 20174036]

[36]. Bohara VA and Ting SH, “Theoretical analysis of OFDM signals in nonlinear polynomial models,” in 2007 6th International Conference on Information, Communications & Signal Processing IEEE, 2007, pp. 1–5.

[37]. Wikipedia, “Stone–Weierstrass theorem,” https://en.wikipedia.org/wiki/Stone-Weierstrass_theorem, accessed: 2020-03-22.

[38]. Kay SM, Fundamentals of Statistical Signal Processing: Estimation Theory. New Jersey: Prentice Hall, 1993.

[39]. Trees HV, Bell K, and Tian Z, “Detection, estimation, and modulation theory, part I,” 1968.

[40]. Wikipedia, “Quantile function,” https://en.wikipedia.org/wiki/Quantile_function, accessed: 2020-03-22.

[41]. Gilchrist W, Statistical modelling with quantile functions. CRC Press, 2000.

[42]. Bowling SR, Khasawneh MT, Kaewkuekool S, and Cho BR, “A logistic approximation to the cumulative normal distribution,” Journal of Industrial Engineering and Management, vol. 2, no. 1, pp. 114–127, 2009.

[43]. Optimization Toolbox MATLAB version 9.4 (R2018a). Natick, Massachusetts: The MathWorks Inc., 2018.

[44]. Hart JD, Hutchinson MN, Williams CRS, Nichols JM, Cranch GA, Rubaiyat AHM, and Rohde GK, “Fiber laser acoustic emission sensing based crack detection via cumulative distribution transform,” in Submitted to 2020 Optical Networking and Communication Conference.

[45]. Aggelis DG and Matikas TE, “Effect of plate wave dispersion on the acoustic emission parameters in metals,” Computers and Structures, vol. 98-99, 2012.

[46]. Gustafsson F and Gunnarsson F, “Positioning using time-difference of arrival measurements,” in 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP’03)., vol. 6. IEEE, 2003, pp. VI–553.

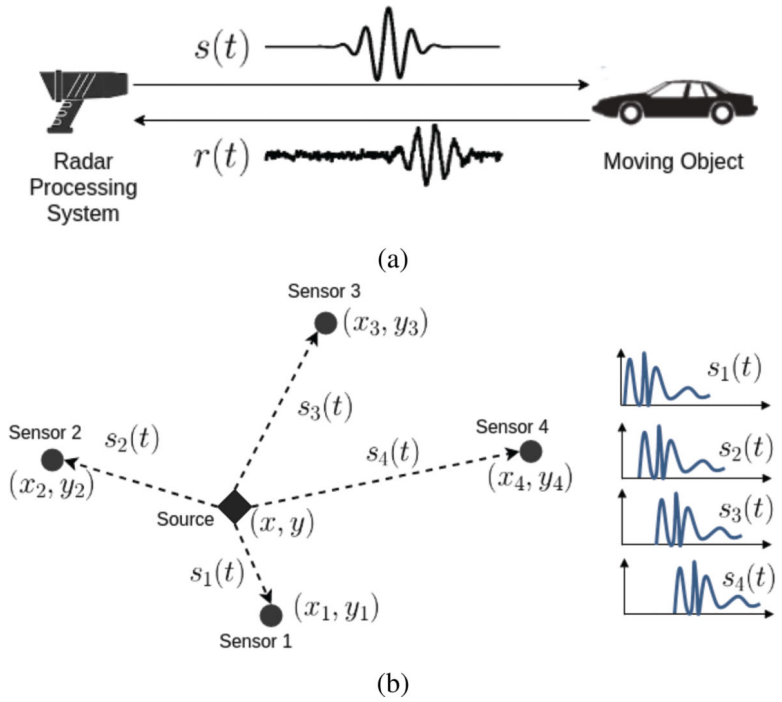


Fig. 1:
Estimation in signal processing applications: (a) radar system, and (b) source localization.

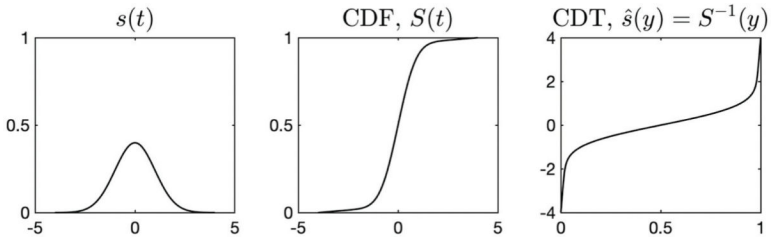


Fig. 2:

Steps of calculating the CDT of a signal $s(t)$, given the uniform reference signal $s_0(y)$.

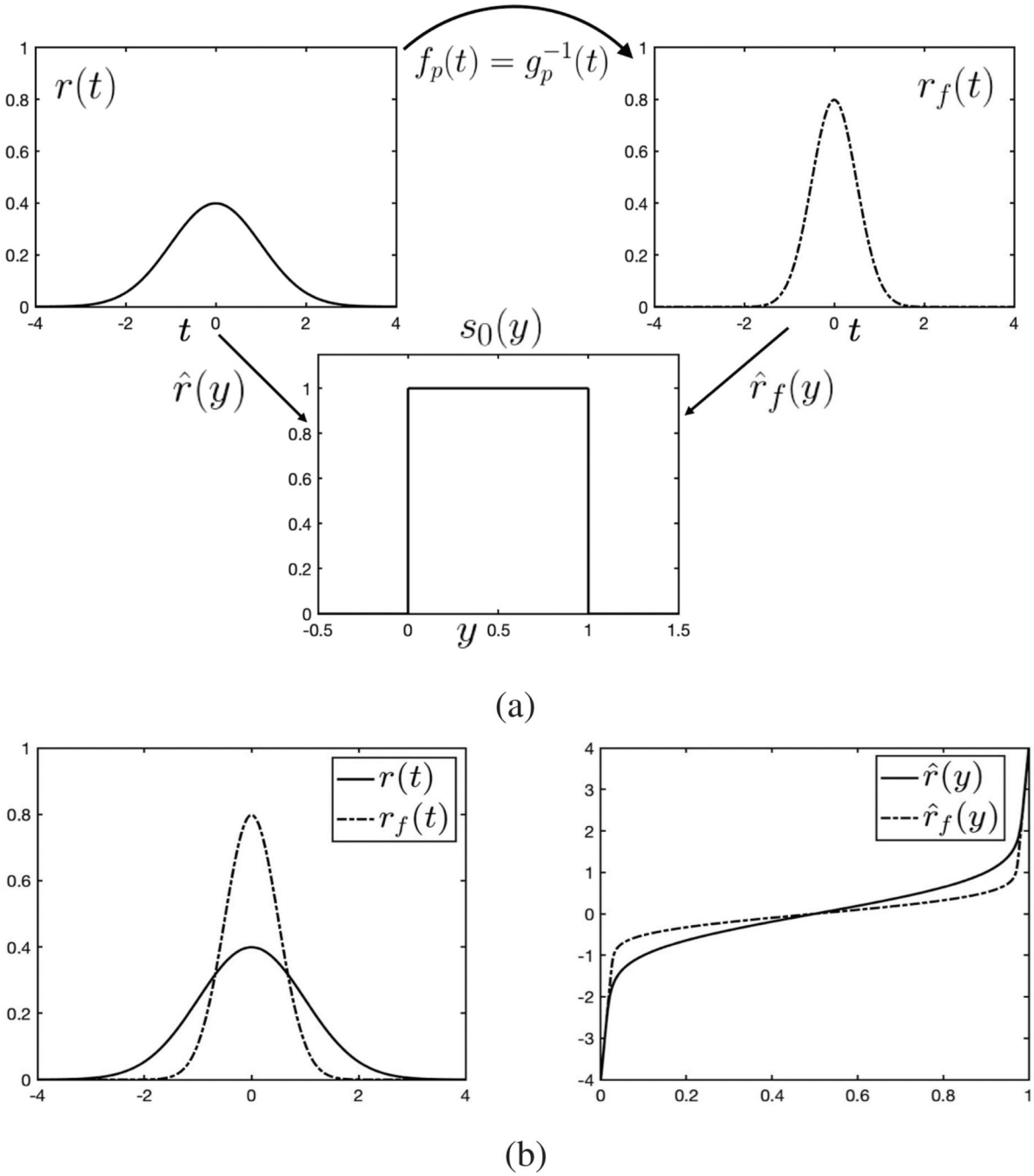


Fig. 3:

Example case relating the different signals and transforms used in the CDT. Let s_0 be the reference signal, r be the measured signal, and r_f be the manipulated signal. The transforms and their directions are also given. Plots in (b) show that the transformation (linear dispersion in this case) along the independent axis in signal space becomes a transformation along the dependent axis in CDT space.

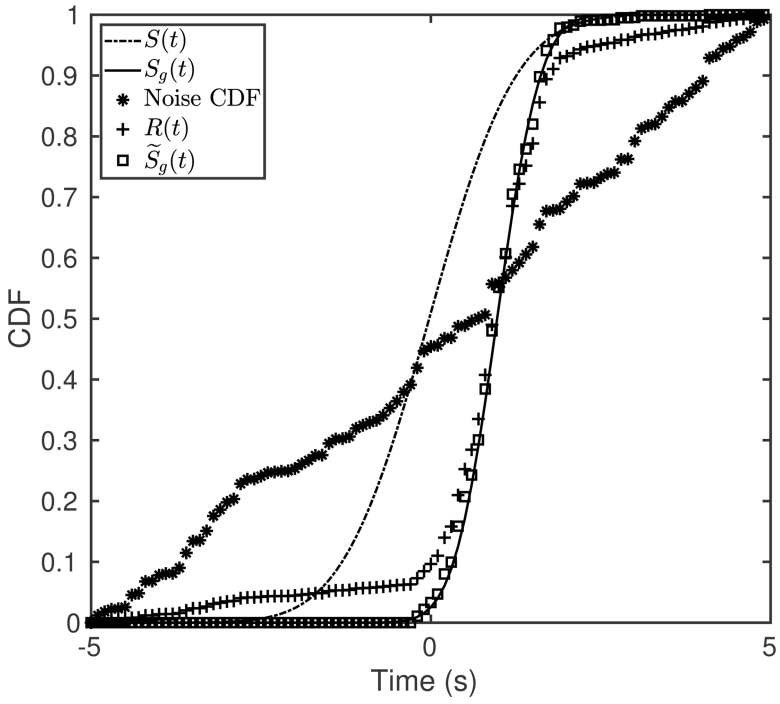
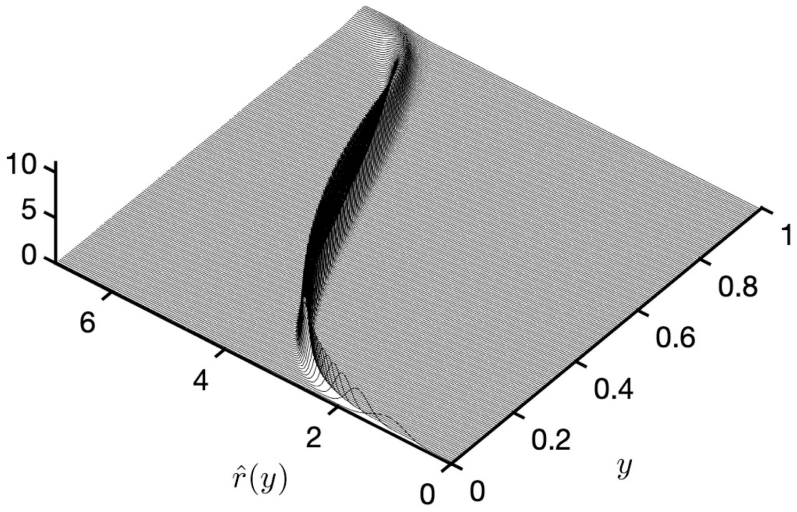
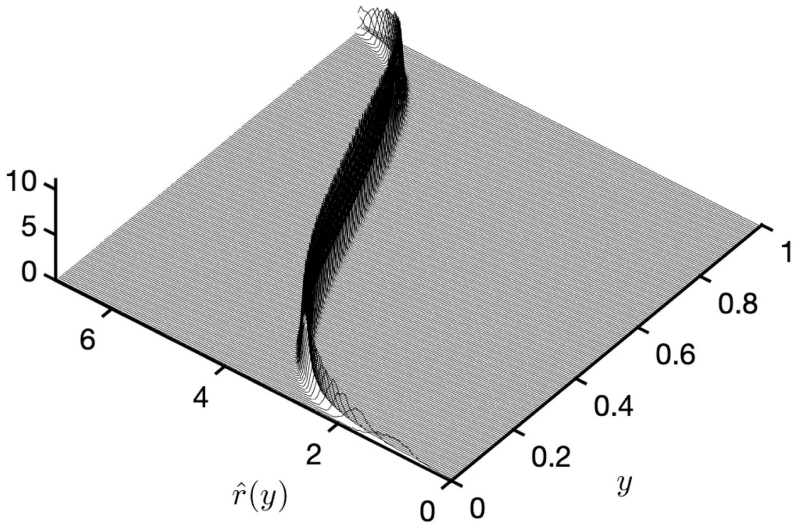


Fig. 4:

CDFs associated with a Gaussian pulse before, $S(t)$, and after, $S_g(t)$, transformation by $g_p(t)$. Also shown are the noisy, $R(t)$, and noise corrected, $\tilde{S}_g(t)$, CDFs. The true and corrected CDFs match almost exactly, even for this relativelyhigh ($SNR = 4$) level of noise.



(a) Theoretical distributions (given a logistic CDF)



(b) Empirically obtained distributions

Fig. 5:

(a) Theoretical and (b) empirically obtained distributions for the CDT $\hat{r}(y)$ given a logistic CDF (Eqn. (34)) consistent with the example of Fig. 4. The distributions appear peaked with a single maximum and are well-approximated by a Gaussian function.

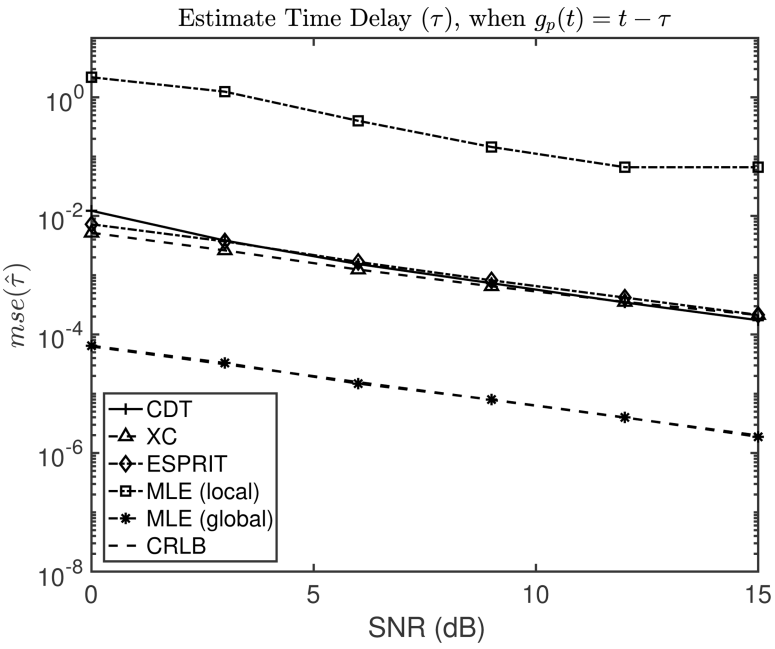


Fig. 6:

MSE in the delay estimates as a function of SNR. For comparison, the estimates produced using MLE, XC, ESPRIT, and the CRLB are also shown. In case of MLE, results for both local and global solvers are plotted.

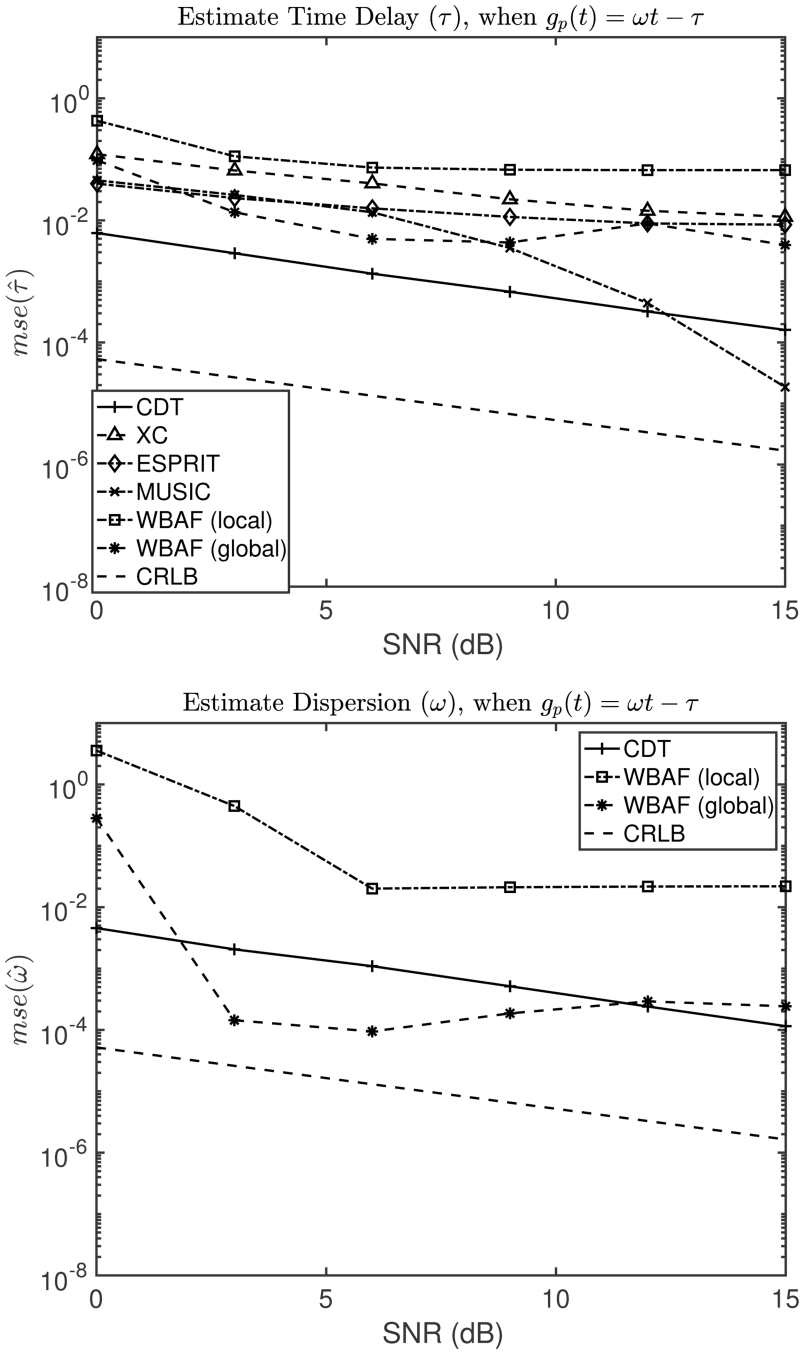


Fig. 7:

MSE in the joint time delay (top) and linear dispersion (bottom) estimates obtained via linear least squares in the CDT domain as compared to the CRLB, XC, ESPRIT, MUSIC, and WBAF (using both local and global solvers). In case of XC, ESPRIT, and MUSIC algorithms, only delay estimates are reported.

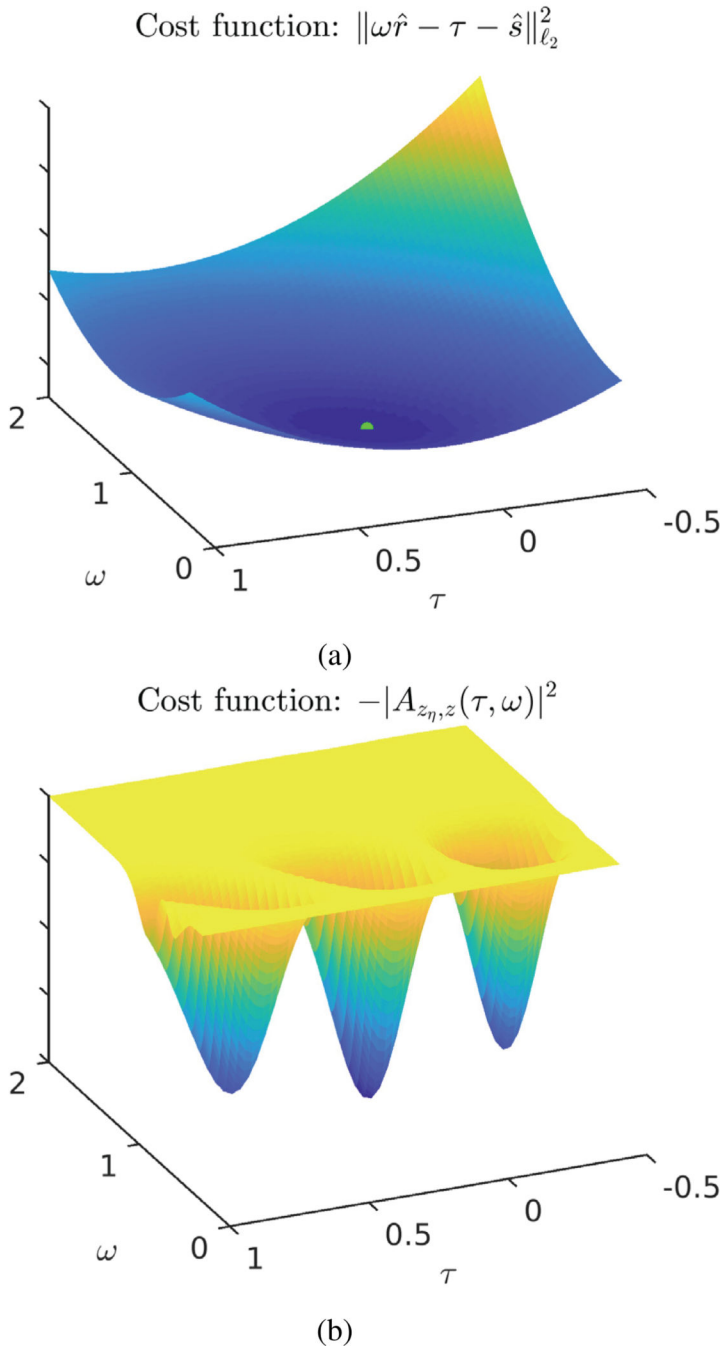


Fig. 8:

Cost functions associated with (a) proposed CDT based estimator (green dot shows the global minimum point), and (b) joint time delay and linear dispersion estimation using WBAF.

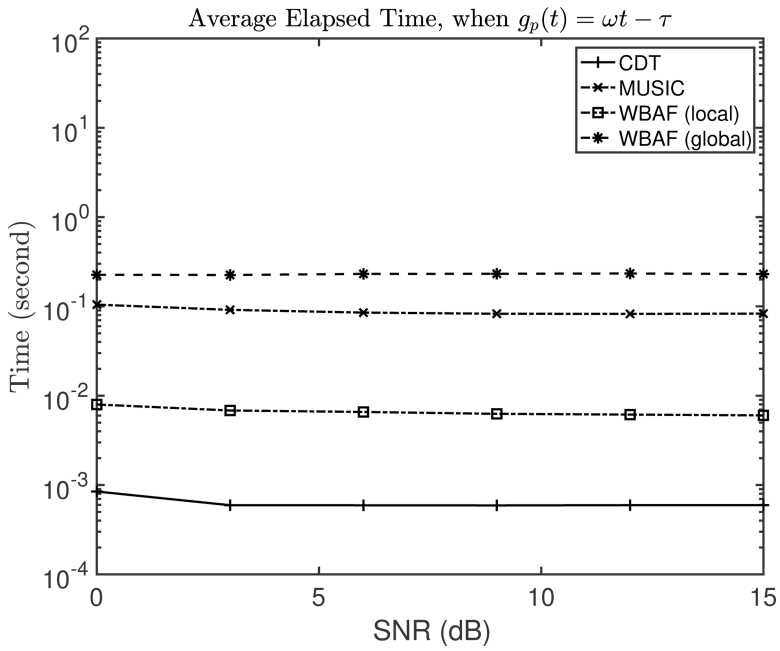


Fig. 9:

Average elapsed time for CDT, MUSIC, and WBAF based estimators. Experiments were run using MATLAB version: 9.4.0 (R2018a) on a computer with an Intel Xeon(R) CPU E5-2630 v3 processor running at 2.40 GHz using 32 GB of RAM.

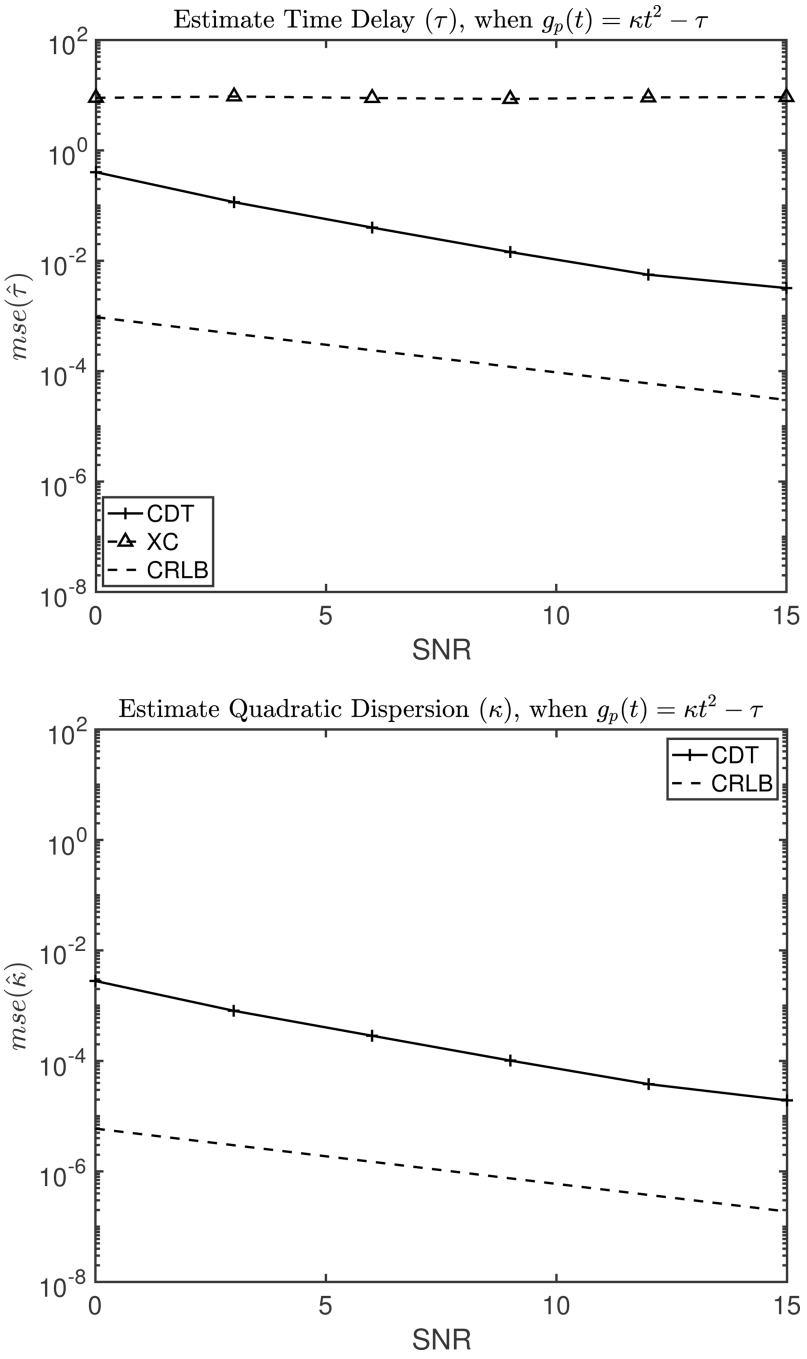


Fig. 10:
MSE associated with the joint estimates of time delay (top) and quadratic dispersion (bottom) parameters

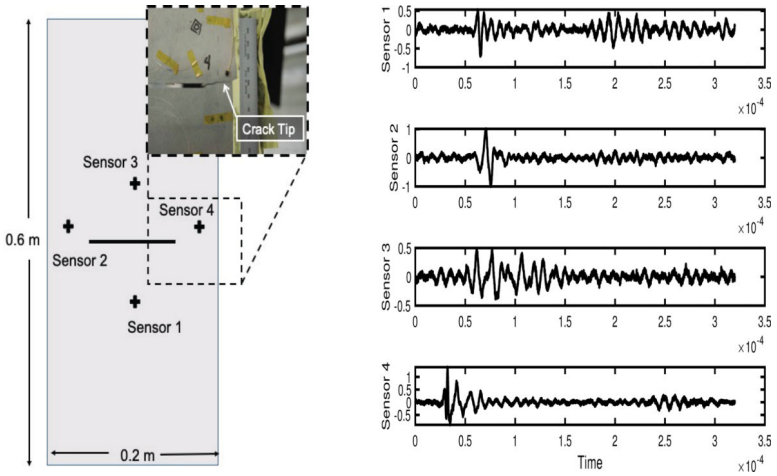


Fig. 11:

(left) Slotted aluminum plate with a crack emanating from the right end of the slot. As the crack propagates it gives off acoustic emission pulses which can be measured at different locations on the plate (right). The time delay of arrival between recorded pulses can then provide the location of the crack tip.

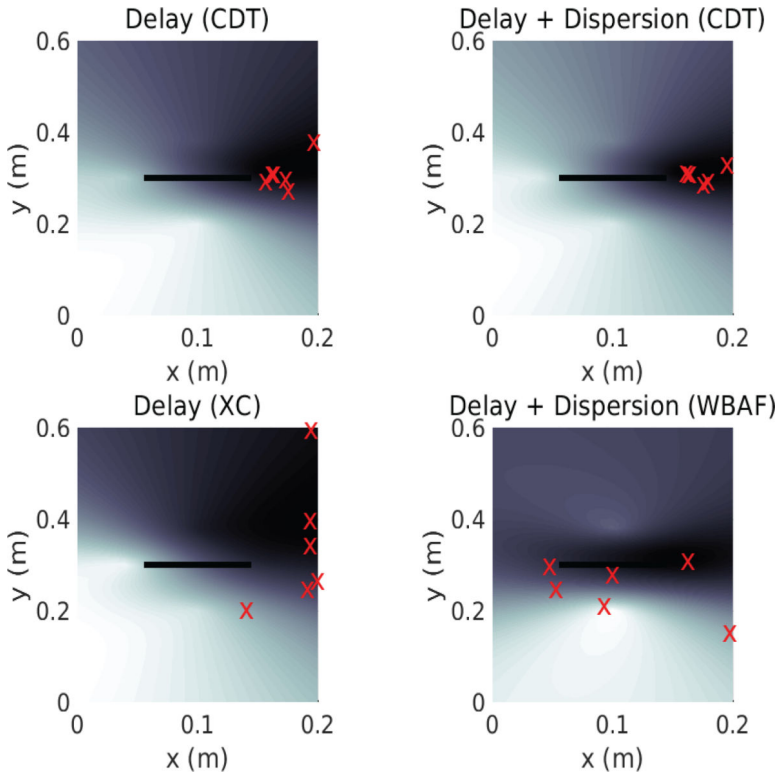


Fig. 12:

Typical cost function for the source localization problem superimposed on the physical plate dimensions. In this case the estimated source location provides the crack tip location (denoted 'X'). Four fiber-optic strain sensors record the data and the location estimate is based on the delay estimates among the sensors as obtained via the cross-correlation (XC), WBAF and CDT. Shown are the locations that rely on the delay-only estimators (left) and the delay + linear dispersion estimator (right).

TABLE I:

Description of symbols

Symbols	Description
s	Normalized, strictly positive signal
S	Cumulative distribution function (CDF) of s
s_0	Reference density function
$\hat{s}$	Cumulative distribution transform (CDT) of s
s_g	Generated from s under the action of g_p
g_p	One-to-one, continuous function with parameter $\boldsymbol{p}$
z_g	Measured signal (not normalized) in absence of noise
r	$B(z_g + \eta)$; normalized measured signal with noise (η)
B	Normalization scheme to transform raw signals into PDFs
r_f	Generated from r under the action of f_p
f_p	g_p^{-1}
$W(s_1, s_2)$	Wasserstein distance between density functions s_1 and s_2