Linear Estimation of Time Warped Signals

Rami Hagege and Joseph M. Francos *

Abstract

We introduce a novel methodology for estimating the time-axis deformation between two observations on a time-warped signal. Since the problem of estimating the warping function is non-linear, existing methods iteratively minimize some metric between the observation and an hypothesized deformed template. Assuming the family of possible deformations the signal may undergo, admits a finite dimensional representation, we show that there is a *nonlinear* mapping from the space of observations to a low dimensional linear space, such that in this space the problem of estimating the parametric model of the warping function is solved by a *linear* system of equations. We call the family of estimators derived based on this representation, Linear Warping Estimators (LWE). The new representation of the problem enables an analytic analysis of the behavior of the solution in the presence of model mismatches, which is prohibitive when iterative methods are employed. The ability to achieve this major simplification both in the solution and in analyzing its performance results from the representation of the problem in a new coordinate system which is natural to the properties of the problem, instead of representing it in the standard coordinate system imposed by the sampling mechanism. The proposed solution is *unique and exact*, as it provides a closed form expression for evaluating each of the parameters of the warping model using *only* measurements of the amplitude information of the observed and reference signals. The solution is applicable to any elastic warping *regardless of its magnitude.* We analyze the behavior of the LWE in the presence of noise, and obtain a minimum variance unbiased estimator for the model parameters, by finding an optimal set of nonlinear operators for mapping the original problem into a low dimensional linear space.

Keywords: Elastic deformations, dynamic time warping, non-linear functionals, linear estimation, unbiased minimum variance estimation

*R. Hagege and J. M. Francos are with the Department of Electrical and Computer Engineering, Ben-Gurion University, Beer-Sheva 84105, Israel. Tel: +972 8 6461842, FAX: +972 8 6472949, email: hagege,francos@ee.bgu.ac.il

1 Introduction

Registration is the procedure of bringing two or more observations on the same signal to a common coordinate system. These signals are usually referred to as the template (or reference) signal, and the observed signal. The difficulty of the registration problem results from its most basic characteristic: although the template is known, the variability associated with the object, due to the warping of the time axis is unknown *a-priori*, and only the group of actions causing this variability in the observation can be defined, based for example on the physical characteristics of the problem. This huge variability in the object signature (for any single object) due to the tremendous set of possible deformations that may relate the template and the observed signature, makes any detection and recognition problem ill-defined unless this variability is implicitly or explicitly taken into account.

The fundamental settings of the problem are provided in [1]. There are two key elements in a deformable template representation: A typical element (the template); and a family of transformations and deformations, $\mathcal{G}$, which when applied to the typical element produce other elements. The family of deformations considered in this paper is extremely wide: we consider homeomorphisms having a continuous and differentiable inverse, where the derivative of the inverse is also continuous, and admits a finite dimensional representation.

Thus each template is associated with its orbit, induced by the group action on the template. Hence, given the template, to be denoted by g, and measurements of an observed signal denoted by h, registration becomes the procedure of finding the group element ϕ, that minimizes some metric between the observation and the hypothesized deformed template $g(\phi)$. In the absence of noise, the solution to the problem is obtained by applying each of the deformations in the group to the template, followed by comparing the result to the observed realization. However, as the number of such possible deformations is infinite, this direct approach is computationally prohibitive. Hence, more sophisticated methods are essential.

In principle, two possible methodologies for estimating the warping function may be considered: explicit and implicit. An implicit method is one that finds some map ψ such that, ideally, $\psi(h, g, \phi) = 0$. All registration methods based on minimizing some metric, are implementations of this basic idea, including the Dynamic Time Warping algorithm (DTW), [8], [9].

The common principle in the implementation of all the implicit methods is the definition of a cost function penalizing both the 'distance' between a deformed version of the template and the observation, and a measure of the 'size' of the deformation. The aim is then to find the deformation that minimizes the cost. More specifically, let $d(h, g)$ be some metric on the function space that contains h and g. A solution to the deformation estimation problem is given by $\hat{\phi} = \arg\min_{\phi \in \mathcal{G}} \left(d(h(\mathbf{x}), g(\phi(\mathbf{x}))) + D(I, \phi) \right)$, where $D(I, \phi)$ is a regularization term specifying some *a-priori* knowledge about the distance of ϕ from the identity [1, 3]. In principle, in order to find the global minimum of $d(h(\mathbf{x}), g(\phi(\mathbf{x}))) + D(I, \phi)$ one has to check each and every element of $\mathcal{G}$, which is usually impossible. Nevertheless, application of some optimization procedure allows for finding a local minimum of this type of cost function, (see *e.g.*, [7], [5], [14]). Unfortunately, in general, there is no systematic way to obtain the global minimum. This type of approach is applied, for example, in [6] (using a very similar setting of the problem to the one discussed in this paper) in the framework of word recognition in speech. More specifically, in [6] word recognition takes place by comparing the pronounced word with a set of word templates. This word is assumed to be obtained from one of the prototype words by a local change of speed in the pronunciation, which results in a monotone time-warping plus an additive observation noise. For each template the best time-warping function is obtained by minimizing a functional of the above form that penalizes both its matching error (due to noise) and its departure form the identity. Different hypotheses are then compared on the basis of the minimum value of this functional, and the pronounced word is recognized as the template for which this minimum value is the smallest.

On the other hand, in an explicit solution one obtains a map H (or an operator) such that the unknown deformation can be expressed by $\phi = H(h, g)$. Obviously, an explicit solution is preferable due to many reasons. These include, computational complexity as optimization is avoided, and more importantly uniqueness of the solution. As indicated above, the equality $\psi(h, g, \phi) = 0$ may have more than a single solution. On the other hand the explicit solution is always global in nature, since no local minimization operations are involved. Many such global methods exist both for 1-D and 2-D signals (see, *e.g.*, [17] and the references therein) however their scope is restricted to a relatively small family of transformations. Thus, in the case of images for example, there are explicit methods for handling translation only, rotation only, or global scale (moderate factor) only, but they turn into combined explicit/implicit methods for the combined transformation of rotation, scaling and translation, [18]. Translation estimation is conveniently carried out in the Fourier domain based on the phase shift of the Fourier transforms of the two images to be registered, by employing the normalized phase-correlation algorithm, *e.g.*, [19].

Since currently, no explicit methods for estimating an elastic time warping are known, the dynamic programming based DTW and its modifications (see, *e.g.*, [11]) became the standard state-of the art tool in estimating time warped functions, and in registering signals whose time-axes are warped. The DTW provides the best piece-wise linear approximation of the deformation function on a discrete grid, with respect to the defined metric. Yet, this solution is also obtained by iteratively minimizing a metric, and hence is computationally demanding. In speech recognition, DTW and its extension to the stochastic case in the form of Hidden Markov Model (HMM), have become the standard tool for accommodating different durations and pronunciations of the same phoneme or word, by the same, or by different speakers [8, 9, 10]. DTW is extensively used in indexing of time series databases, [12, 13], where the need is to find the best match to a query time series, from a large collection of possible candidates.

More recently, in the context of decomposing signals into sparse linear combinations of template signals drawn from a large finite dictionary, [15] points to the inherent difficulty of dictionary based approaches in handling real signals that are subject to continuous deformations of the time axis, where identification of the warping parameters, is critical for finding the correct matches in the dictionary. Therefore, the common practice is to construct a dictionary that represents the undeformed template functions and the transformations of each such template on some discretization of the deformations parameter space. However, in general, in order to faithfully represent signals using the dictionary, a very fine sampling of the parameter space is required, which leads to a very large and ill-conditioned dictionary. The method proposed in this paper provides a key to how problems that involve the estimation of time-warped signals, that are currently solved by optimization techniques, can be solved analytically and linearly. Moreover, the new representation of the problem enables an analytic analysis of the behavior of the solution in the presence of model mismatches, which is prohibitive when iterative methods are employed. The ability to achieve this major simplification both in the solution and in analyzing its performance results from the representation of the problem in a new coordinate system which is natural to the properties of the problem, instead of representing it in the standard coordinate system imposed by the sampling mechanism.

Our goal in this paper is to find an explicit global operator $H(h, g)$ such that for every pair (h, g) for which $h(x) = g(\phi(x))$, $\phi \in \mathcal{G}$ where $\mathcal{G}$ is the group of homeomorphic warps admitting a finite dimensional representation, we have $\phi = H(f, g)$. To the best of our knowledge there is no other method that is both explicit and is capable of recovering such general class of warping functions. The center of the proposed solution is a method that reduces the original high dimensional problem of evaluating the orbit created by applying the set of all possible homeomorphic transformations in the group to the template, into a problem of analyzing a function in a low dimensional Euclidian space. In general, an explicit modeling of the homeomorphisms group is impossible. Nevertheless we show in this paper that in cases where the set of deformations, $\mathcal{G}$, admits a finite dimensional representation, there is a mapping from the space of observations to a low dimensional linear space. In this setting, the problem of estimating the parametric model of the warping function is solved by a *linear* system of equations in the low dimensional Euclidian space. The proposed solution is *unique and exact*, as it provides a closed form expression for evaluating each of the parameters of the warping model using *only* measurements of the amplitude information of the observed and reference signals. The solution is applicable to any elastic warping *regardless of its magnitude*. We call the family of estimators derived based on this representation, Linear Warping Estimators (LWE). Due to their low memory and computational requirements, LWE greatly simplify the solution to any application in which estimation of time-warped signals is involved, some of which were mentioned above. In fact, when long observations on time-warped signals are considered, the computational and memory requirements of the implicit optimization methods, are prohibitive, leaving the LWE as the only practical option for solving this problem as we illustrate using some numerical examples in Section 4.

The structure of the paper is as follows: In Section 2 we rigorously define the problem of estimating the homeomorphic deformation in the absence of observation noise, its setting, and derive the algorithmic solution for the parameters of the warping function. To simplify the notation and the accompanying discussion we present the solution for the case where the observed signals are one-dimensional. The derivation for higher dimensions follows along similar lines, [20]. In Section 3 we rigorously analyze the structure of the space of non-linear operators applied to the observed warped signals in order to map the original nonlinear problem into a linear problem in the deformation model parameters. In addition to the huge variability in the object signature due to the unknown deformations, the observations are also noisy, in general. In Section 4 we analyze the behavior of the proposed solution for estimating the deforming function in the presence of noise, and obtain minimum variance unbiased estimator for the model parameters. Assuming further that the observation noise is Gaussian, a maximum likelihood estimator is also derived for the high SNR regime.

2 Explicit Evaluation of Time Warps

In this section we shall briefly set the mathematical framework we adopt in order to formalize the analysis of the deformation estimation problem. This framework enables accurate representation and analysis of our problem, leading to rigorous criteria on the existence and uniqueness of the solution, and under some mild restrictions to be explained below to the derivation of an explicit solution.

2.1 Problem Statement

We note that due to the inherent physical properties of the problem, it is natural to model and solve it in the continuum. Inherently, the mapping ϕ of $\mathbb{R}$ into itself is of a continuous nature, as is the physical phenomenon of geometric deformation of real-life objects it represents. Thus, if we impose a discrete model (*e.g.*, $x \in \mathbb{Z}$), we find that, in general, the natural ϕ to consider is incompatible (as for "almost all" $x \in \mathbb{Z}$, $\phi(x) \notin \mathbb{Z}$). Thus, in contrast with existing methods such as DTW, the problem and its solution are formulated in the continuum, while the sampling and quantization effects that accompany the digital implementation of the method, are handled as noise contributions.

Let M denote the space of compact support, bounded, and Lebesgue measurable (or more simply, integrable) functions from $\mathbb{R}$ to itself. Let $\mathcal{G}$ be a group representing the set of elastic deformations the function may undergo. In this paper it is assumed that $\mathcal{G}$ is the group of homeomorphisms such that each element of the group has a continuous and differentiable inverse, where the derivative of the inverse is also continuous. $\mathcal{G}$ is said to act as a transformation group on M if there is a mapping $\mathcal{G} \times M \to M$, denoted by $(\phi, m) \mapsto m \circ \phi = m(\phi(\mathbf{x}))$ such that $(m \circ \phi_1) \circ \phi_2 = m \circ (\phi_1 \circ \phi_2)$ for every $\phi_1, \phi_2 \in \mathcal{G}$ and $m \in M$; and if $m \circ e = m$ for all $m \in M$, where e is the identity element of $\mathcal{G}$.

For a given $m \in M$, the set $\{m \circ \phi : \phi \in \mathcal{G}\}$ is called the orbit of m. It is the entire set of possible observations on the object – the result of applying to it any of the deformations in the group. The stabilizer of the function $m \in M$ with respect to the group $\mathcal{G}$ is the set of group elements $\phi \in \mathcal{G}$ such that $m \circ \phi = m$, *i.e.*, the set of group elements that map m to itself.

Thus the group $\mathcal{G}$ naturally defines an equivalence relation on M in terms of the orbits of M induced by the action of $\mathcal{G}$: Any two functions h and g are equivalent if they are on the same orbit, *i.e.*, if there exists some $\phi \in \mathcal{G}$ such that $g \circ \phi = h$.

In the framework of the present paper it is assumed that M is in fact a subset of the set of compact support, bounded, and Lebesgue measurable functions from $\mathbb{R}$ to itself, such that for all functions in M the stabilizer is trivial and includes only e, the identity element of $\mathcal{G}$. Thus, uniqueness of the solution to the defined problem is guaranteed in the sense that if $h, g \in M$ such that they are on the same orbit, then there exists a *single* ϕ such that $g \circ \phi = h$.

Thus, given two functions $h, g \in M$ such that

$$h(x) = g(\phi(x)) \ . \tag{1}$$

our problem is to determine $\phi(x)$.

Next, let $C(X)$ denote the set of continuous real-valued functions of X onto itself, where the norm is the standard L_2 norm. By the above assumption every $\phi^{-1}, (\phi^{-1})' \in C(X)$. Since $C(X)$ is a normed separable space, [21], there exists a countable set of basis functions $\{e_i\} \subset C(X)$, such that for every $\phi \in \mathcal{G}$,

$$(\phi^{-1})'(x) = \sum_i a_i e_i(x) \ . \tag{2}$$

In other words, it is assumed that every element in the group and its derivative can be represented as a convergent series of basis functions of the separable space $C(X)$. Our goal then, is to obtain the expansion of $\phi^{-1}(x)$ with respect to the basis functions $\{e_i(x)\}$. In practice, the series (2) is replaced by a finite sum.

More specifically, in all the analysis that follows, the amplitude value 0 represents no object. Without limiting the generality of the derivation it is assumed that the support of $g(x)$ is $[a, b]$ and we want to model warping within the template but with fixed ends; more accurately we are looking for a time warp $\phi(x)$ such that $\phi(a) = a$ and $\phi(b) = b$ yet within the interval the warp may be elastic. Hence, the finite-dimensional model of the inverse warp is given by $\phi^{-1}(x) = x + \sum_{i=1}^{P} a_i E_i(x)$ with $E_i(a) = E_i(b) = 0 \ \forall i$ and $e_i(x) = \frac{dE_i(x)}{dt}$. Let $\mathbf{a} = [a_1, \ldots, a_P]^T$ be the vector of the deformation parameters.

Remark 1. We consider time warping functions that can be modeled as a subset of some finite dimensional linear space of differentiable functions, and we choose to model the inverse warping. Note that if we consider the two functions $g(x)$ and $h(x)$ in the relation $h(x) = g(\phi(x))$ as having the same role then this relation is equivalent to the relation $h(\phi^{-1}(x)) = g(x)$ and therefore modeling of the warping or its inverse are equivalent. When the roles are different, for example when $g(x)$ is some known template function, while $h(x)$

is the observation, it becomes more natural to model $\phi^{-1}(x)$ instead of modeling $\phi(x)$. This is because the time warp $\phi(x)$ represents a map from the coordinate system where $h(x)$ is measured to the coordinate system of $g(x)$. However $h(x)$ is subject to an unknown wrappings and therefore the meaning of a fixed map defined on its coordinate system is ambiguous. On the other hand, $\phi^{-1}(x)$ models a map from the coordinate system of $g(x)$ which is fixed, as $g(x)$ is the reference template.

2.2 The Fundamental Solution

Let W be the space of bounded measurable functions (operators) from $\mathbb{R}$ into itself.

Lemma 1. *Let $h(x), g(x) \in M$ be two functions such that $h(x) = g(\phi(x))$ with $\frac{d\phi^{-1}(x)}{dx} = \sum_{i=1}^{P} a_i e_i(x)$. Then, every $w(y) \in W$ provides a single linear constraint on the elements of* $\mathbf{a}$ *in the form $\int_{-\infty}^{\infty} w\left(h(x)\right) dx = \sum_{i=1}^{P} a_i \int_{-\infty}^{\infty} e_i(x) w\left(g(x)\right) dx$.*

Proof: Let $z = \phi(x)$. Then $\phi^{-1}(z) = x$. Using a change of variables

$$\int_{-\infty}^{\infty} w\left(h(x)\right) dx = \int_{-\infty}^{\infty} w\left(g(\phi(x))\right) dx = \int_{-\infty}^{\infty} \left(\phi^{-1}(z)\right)' w\left(g(z)\right) dz = \sum_{i=1}^{P} a_i \int_{-\infty}^{\infty} e_i(z) w\left(g(z)\right) dz \tag{3}$$

Repeating this procedure by applying a family of Lebesgue measurable, left-hand compositions $\{w_i\}_{i=1}^{N} \in W$, to the known relation $h(x) = g(\phi(x))$, and rewriting it in a matrix form we obtain

$$\begin{bmatrix} \int w_1(h(x)) \\ \vdots \\ \int w_N(h(x)) \end{bmatrix} = \begin{bmatrix} \int e_1(x)\ w_1(g(x)) & \dots & \int e_P(x)\ w_1(g(x)) \\ \vdots & \ddots & \vdots \\ \int e_1(x)\ w_N(g(x)) & \dots & \int e_P(x)\ w_N(g(x)) \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_P \end{bmatrix} \tag{4}$$

In the sequel, let $\mathbf{w}$ denote a vector whose elements are functions, and $\mathbf{w}_k$ denotes the function in the $k'th$ row of $\mathbf{w}$. (Hence, when such a vector is left multiplied by a matrix of scalars, the elements of the resulting vector are linear combinations of functions.)

Thus let $\mathbf{e} = [e_1(x), \dots, e_P(x)]^T$ and let $\mathbf{w} = [w_1(x), \dots, w_N(x)]^T$. Also, let

$$\mathbf{G}(g, \mathbf{e}, \mathbf{w}) = \begin{bmatrix} \int e_1(x)\ w_1(g(x)) & \dots & \int e_P(x)\ w_1(g(x)) \\ \vdots & \ddots & \vdots \\ \int e_1(x)\ w_N(g(x)) & \dots & \int e_P(x)\ w_N(g(x)) \end{bmatrix} \tag{5}$$

and let

$$\mathbf{h}(h, \mathbf{w}) = \begin{bmatrix} \int w_1(h(x)) \\ \vdots \\ \int w_N(h(x)) \end{bmatrix} \tag{6}$$

Rewriting (4) we have

$$\mathbf{h}(h, \mathbf{w}) = \mathbf{G}(g, \mathbf{e}, \mathbf{w})\mathbf{a} \tag{7}$$

We have just proved the following theorem:

Theorem 1. *Let $\phi(x)$ be an element of the group of homeomorphisms such that $\frac{d\phi^{-1}(x)}{dx} = \sum_{i=1}^{P} a_i e_i(x)$. Let $h(x), g(x) \in M$ be two functions such that $h(x) = g(\phi(x))$. Then, given h and g, the deformation $\phi(x)$ can be uniquely determined if there exists a set of functions $\{w_N\}_{i=1}^{N} \in W$, such that the matrix $\mathbf{G}(g, \mathbf{e}, \mathbf{w})$ is full rank. Then,*

$$\mathbf{a} = [\mathbf{G}(g, \mathbf{e}, \mathbf{w})]^{-1}\mathbf{h}(h, \mathbf{w}) \tag{8}$$

In Section 3 we elaborate on the meaning of $[\mathbf{G}(g, \mathbf{e}, \mathbf{w})]^{-1}$ for $N > P$. Also, note from (8) that we have found an explicit solution for the unknown deformation parameters $\mathbf{a}$. In fact, there could be an infinite number of choices of the functions in the set $\{w_i\}_{i=1}^{N}$ leading to the same solution for $\mathbf{a}$. Section 3 is devoted to a detailed analysis of the relations between the solutions obtained by different choices of the set of functions $\{w_i\}_{i=1}^{N}$.

Remark 2. Note that the elements of the matrix $\mathbf{G}(g, \mathbf{e}, \mathbf{w})$ depend only on the template and its coordinate system and thus have to be evaluated only *once*. In fact $\mathbf{G}(g, \mathbf{e}, \mathbf{w})$ represents all the information in the template, required for finding the warping parameters. Thus $\mathbf{G}(g, \mathbf{e}, \mathbf{w})$ forms a **"sufficient representation"** of the template (similarly to the notion of sufficient statistics), so that the template itself is not needed in order to uniquely determine the warping function once $\mathbf{G}(g, \mathbf{e}, \mathbf{w})$ has been evaluated.

Remark 3. The application of a set $\{w_i\}_{i=1}^{N}$ to $g(x)$ yielding $\mathbf{G}(g, \mathbf{e}, \mathbf{w})$ is in fact a projection from the space of compact support, bounded, and measurable functions to the space of $N \times P$ matrices. The following theorem states that the subset of functions $g \in M$, for which there exists a set $\{w_i\}_{i=1}^{N}$ such that $\mathbf{G}(g, \mathbf{e}, \mathbf{w})$ is full-rank, is dense in M in the supremum norm. Hence, for every g, or for an infinitesimal modification of it, the matrix $\mathbf{G}(g, \mathbf{e}, \mathbf{w})$ is invertible.

Theorem 2. *Let $g \in M$. Then for every $\epsilon > 0$, there exist some function $g^1 \in M$ such that $||g - g^1||_\infty < \epsilon$, and a set $\{w_i\}_{i=1}^{N}$ such that $\mathbf{G}(g^1, \mathbf{e}, \mathbf{w})$ is full rank.*

Proof. The proof follows the same lines as in the case where the transformation is affine, [2], Theorem 2, and hence omitted. □

Remark 4. The practical implication of Remark 3 is that for a function $g \in M$, there exists a set $\{w_i\}_{i=1}^{P}$ such that $\mathbf{G}(g, \mathbf{e}, \mathbf{w})$ is full-rank, if and only if the range of g contains at least P distinct values.

3 The Structure of the Space of Solutions

Lemma 1 implies that given a geometric transformation model expressed by the transformation group $\mathcal{G}$ and a function $g(x) \in M$, each $w \in W$ yields a linear constraint on the parameters $\{a_i\}_{i=1}^{P}$ defining the transformation $h(x) = g(\phi(x))$ such that the constraints depend on the template $g(x)$ but *not* on the deformation $\phi \in \mathcal{G}$. In fact, $\{a_i\}_{i=1}^{P}$, the set of parameters defining the space of possible deformations $\mathcal{G}$, defines an P dimensional *coordinate system* on $\mathcal{G}$. Since the dimension of the space of parameters is P, we can find at most P independent linear constraints on the model parameters. Therefore, in the coordinate system whose axes are the $\{a_i\}$'s, we may plot the linear constraint imposed by each $w \in W$. For example, consider the geometric transformation model $\phi^{-1}(x) = ae_1(x) + be_2(x)$. Applying Lemma 1 to $h(x) = g(\phi(x))$, we find that each $w \in W$ produces a single constraint of the form $\int_{-\infty}^{\infty} w(h(x))dx = a \int_{-\infty}^{\infty} e_1(x)w(g(x))dx + b \int_{-\infty}^{\infty} e_2(x)w(g(x))dx$ and that the slopes imposed in the (a, b) domain by each $w \in W$ are independent of the geometric transformation parameters. Obviously, the intersection point of the linear constraints in the (a, b) domain provide the deformation parameters.

Since the dimension of W is much larger than P, in most of the cases we have linearly independent w's which nevertheless create exactly the same linear constraints. At first it may seem that these w's provide redundant information. However, the linear independence of the w's implies linear independence of the related functionals. This linear independence suggests that constraints which coincide on the orbit, but not outside it, can be applied to achieve robustness of the linear constraints when model mismatch occurs due to noise for example. A different interpretation of this observation is the following one: As shown in the previous section, in the deterministic case all choices of the set of functions $\{w_i\}_{i=1}^{P}$ are equally optimal as long as $\mathbf{G}(g, \mathbf{e}, \mathbf{w})$ is full rank. Obviously, in the presence of noise different choices of the set $\{w_i\}_{i=1}^{P}$ shall yield different performance in estimating the deformation model parameter vector $\mathbf{a}$. Thus, in order to enable optimal selection of the set $\{w_i\}_{i=1}^{P}$, we must first analyze the structure of the function space W.

Thus, the goal of this section is to decompose the space W into a direct sum of subspaces representing "particular solutions" and "homogenous solutions" (borrowing the terminology from the classical linear theory as explained below).

Lemma 2. *For any matrix* $\mathbf{A}$,$\mathbf{G}(g, \mathbf{e}, \mathbf{Aw}) = \mathbf{AG}(g, \mathbf{e}, \mathbf{w})$ *and* $\mathbf{h}(h, \mathbf{Aw}) = \mathbf{Ah}(h, \mathbf{w})$

Proof.

$$[\mathbf{AG}(g, \mathbf{e}, \mathbf{w})]_{i,j} = \sum_{k=1}^{N} a_{i,k} \int_{-\infty}^{\infty} e_j(x) w_k(g(x)) dx = \int_{-\infty}^{\infty} e_j(x) \left(\sum_{k=1}^{N} a_{i,k} w_k \right) (g(x)) dx = [\mathbf{G}(g, \mathbf{e}, \mathbf{Aw})]_{i,j}$$

and similarly for proving that $\mathbf{h}(h, \mathbf{Aw}) = \mathbf{Ah}(h, \mathbf{w})$. □

Lemma 3. *For any matrix* $\mathbf{B}$,*we have* $\mathbf{G}(g, \mathbf{Be}, \mathbf{w}) = \mathbf{G}(g, \mathbf{e}, \mathbf{w})\mathbf{B}^T$

Proof.

$$\left[\mathbf{G}(g, \mathbf{e}, \mathbf{w})\mathbf{B}^T\right]_{i,j} = \sum_{k=1}^{N} b_{j,k} \int_{-\infty}^{\infty} e_k(x) w_i(g(x)) dx = \int_{-\infty}^{\infty} \left(\sum_{k=1}^{N} b_{j,k} e_k(x) \right) w_i(g(x)) dx = [\mathbf{G}(g, \mathbf{Be}, \mathbf{w})]_{i,j}$$

□

Corollary 1. *For any matrices* $\mathbf{A}$ *and* $\mathbf{B}$, *we have* $\mathbf{G}(g, \mathbf{Be}, \mathbf{Aw}) = \mathbf{AG}(g, \mathbf{e}, \mathbf{w})\mathbf{B}^T$

Thus, choosing different bases either for the functions that are the elements of $\mathbf{w}$ or $\mathbf{e}$ we obtain different constraints on the model parameters. Yet, as the corollary implies, these are no more than different representations of the same information on the constraints. In the following, we shall take advantage of this property in order to obtain more convenient representations of the constraints.

Definition 1. Let $g \in M$. A P dimensional vector of functions $\mathbf{w}$ is *independent with respect* to $\mathbf{e}$ if $\mathbf{G}(g, \mathbf{e}, \mathbf{w}) = \mathbf{I}$.

Lemma 4. *Fix* $\mathbf{e}$. *For a set of functions* $\{w_i\}_{i=1}^{P} \in W$ *such that* $\mathbf{G}(g, \mathbf{e}, \mathbf{w})$ *is invertible there exists a corresponding set of functions* $\{\tilde{w}_i\}_{i=1}^{P} \in W$ *where* $\tilde{w}_i \in span\left(\{w_i\}_{i=1}^{P}\right)$ *such that* $\mathbf{G}(g, \mathbf{e}, \tilde{\mathbf{w}}) = I$.

Proof. Set $\mathbf{A} = \mathbf{G}^{-1}(g, \mathbf{e}, \mathbf{w})$. Then using Lemma 2, $\mathbf{I} = \mathbf{AG}(g, \mathbf{e}, \mathbf{w}) = \mathbf{G}(g, \mathbf{e}, \mathbf{Aw}) = \mathbf{G}(g, \mathbf{e}, \tilde{\mathbf{w}})$ where $\tilde{\mathbf{w}} = \mathbf{Aw}$. □

We therefore conclude that starting with any arbitrary choice of left-hand compositions $\{w_i\}_{i=1}^P$, such that the corresponding $\mathbf{G}(g, \mathbf{e}, \mathbf{w})$ is invertible, there exists an "equivalent" set of left-hand compositions $\{\tilde{w}_i\}_{i=1}^P$ which is independent with respect to $\mathbf{e}$.

Obviously, there is a dual to Lemma 4, where instead of linearly transforming $\mathbf{w}$, we could change the basis functions of the warping:

Lemma 5. *Fix* $\mathbf{w}$. *For a set of functions* $\{e_i\}_{i=1}^P$ *spanning the space of warping functions, such that* $\mathbf{G}(g, \mathbf{e}, \mathbf{w})$ *is invertible there exists a corresponding set of functions* $\{\tilde{e}_i\}_{i=1}^P$ *where* $\tilde{e}_i \in span\left(\{e_i\}_{i=1}^P\right)$ *such that* $\mathbf{G}(g, \tilde{\mathbf{e}}, \mathbf{w}) = I$.

Note however, that in the following analyzes we fix the basis functions $\{e_i\}_{i=1}^P$ representing the space of time warps. We next decompose W into a direct sum of P linear subspaces of particular solutions, each one providing the constraints on a single model parameter a_k, and an additional subspace spanning the constraints on the homogenous solutions, as explained below.

Definition 2. *Let* w *be some function in* W. *Fix* $\{e_i\}_{i=1}^P$ *to be the set of basis functions spanning the space of time warps, and* g *is the template function. Define the operator* $D_{g,\mathbf{e}}(w) = \left[\int\limits_{\infty}^{\infty} e_1(x)w(g(x))dx \cdots \int\limits_{\infty}^{\infty} e_P(x)w(g(x))dx\right]^T$. *Hence, for fixed* $g, \mathbf{e}$, *we have that* $D_{g,\mathbf{e}}(w)$ *is a linear map from* W *to* $\mathbb{R}^P$.

Thus, rewriting (3) using these notations we have $\int\limits_{-\infty}^{\infty} w\left(h(x)\right)dx = \sum\limits_{k=1}^{P} a_k[D_{g,\mathbf{e}}(w)](k)$. As indicated above, $\{a_i\}_{i=1}^P$ the set of parameters defining the space of possible deformations $\mathcal{G}$, can be interpreted as defining an P dimensional coordinate system on $\mathcal{G}$. In this setting, the elements of the vector $D_{g,\mathbf{e}}(w)$ are the coefficients of the linear constraint expressed in the coordinate system defined by $\{a_i\}_{i=1}^P$. In other words, we map each element $w \in W$ into the coefficients of the linear constraint on the elements of $\mathbf{a}$.

Let $\mathbf{1}_i = [0, ..., 0, 1, 0, ...0]^T$ be the P dimensional vector with 1 in the $i'th$ entry only, and let $\mathbf{0} = [0, ..., 0, 0, 0, ...0]^T$.

Definition 3. *The space* $W_i = \left\{w \in W | \int e_j(x)w(g(x))dx = 0 \ \ i \neq j\right\}$ *is the space of constraints on* e_i *only. In other words* $W_i = [D_{g,\mathbf{e}}(w)]^{-1}(span(\mathbf{1}_i)))$, *and every function* $w \in W_i$ *is a particular solution for* a_i. *Also, define* $W_0 = Ker\left(D_{g,\mathbf{e}}\right) = [D_{g,\mathbf{e}}(w)]^{-1}(\mathbf{0})$. *Thus* W_0 *is the space of homogeneous constraints with respect to the template* g *and the set of basis functions* $\{e_i\}_{i=1}^P$.

Lemma 6. *For every $i \neq j$ we have $W_i \cap W_j = W_0$.*

Proof. Let $v \in W_i \cap W_j$ where $i \neq j$. Then $D_{g,\mathbf{e}}(v) = \alpha \mathbf{1}_i = \beta \mathbf{1}_j$ which holds only if $\alpha = \beta = 0$. Since $D_{g,\mathbf{e}}(v) = \mathbf{0}$, by definition $v \in W_0$ □

All the constraints imposed by functions in W_0 are actually null constraints. In other words the information they give us is of the form $0 = 0$. These constraints are clearly redundant in the exact (deterministic) cases. Nevertheless, by using such constraints we extend the space of possible solutions. As we show in the following chapters, the key to achieving optimality and stability in estimating the deformation model lies in the behavior of solutions that are identical in the absence of noise, but which are very different in the presence of deviation from the exact models.

For any nonzero constraint $w \in W_i$ we can find a_i explicitly: Since $\int w\,(f(x))dx = a_i[D_{g,\mathbf{e}}(w)](i)$ we have that $a_i = \frac{\int w(f(x))dx}{[D_{g,\mathbf{e}}(w)](i)}$.

Consider next the subset of W such that $w \in [D_{g,\mathbf{e}}(w)]^{-1}(\mathbf{1}_i)$. Each of the functions in this subset is a particular solution for a_i. Thus for *every* $w \in [D_{g,\mathbf{e}}(w)]^{-1}(\mathbf{1}_i)$, we have $a_i = \int w\,(f(x))dx$. We therefore call $[D_{g,\mathbf{e}}(w)]^{-1}(\mathbf{1}_i)$ *the space of particular solutions* for a_i. Note however that any function $w \in [D_{g,\mathbf{e}}(w)]^{-1}(\mathbf{1}_i)$ yields a different functional on M. Yet in the absence of noise, they all provide the same solution for a_i in the form $a_i = \int w\,(f(x))dx$.

Definition 4. *Let $\alpha(x)$ be some real function defined on M and let $w_p \in W$ be some operator. Then, for some $g(x) \in M$ and for any operator $w \in W$ such that $\int_{-\infty}^{\infty} \alpha(x)w(g(x))dx \neq 0$, we define an operator $Q_w^\alpha : W \to W$ such that*

$$Q_w^\alpha[w_p(z)] = w_p(z) - \frac{1}{\int_{-\infty}^{\infty} \alpha(x)w(g(x))dx}\left(\int_{-\infty}^{\infty} \alpha(x)w_p(g(x))dx\right)w(z) \tag{9}$$

The operator Q_w^α is a linear operator. The next lemma shows that it is also a projection operator.

Lemma 7. *For any $w \in W$ such that $\int_{-\infty}^{\infty} \alpha(x)w(g(x))dx \neq 0$, and for every $w_p \in W$ the operator $Q_w^\alpha(w_p)$ has the property that $\int_{-\infty}^{\infty} \alpha(x)Q_w^\alpha[w_p(g(x))]dx = 0$. Moreover, for every $w_p \in W$ we have that $Q_w^\alpha(w_p)$ is a projection operator.*

Proof. The lemma follows from direct calculations, and verifying that $Q_w^\alpha(Q_w^\alpha(w_p)) = Q_w^\alpha(w_p)$. □

Therefore $Q_w^\alpha(w_p)$ is called the projection operator on the space independent of $\alpha(x)$, and we denote by $Q_w^\alpha(W)$ the subset of operators in W that are independent of $\alpha(x)$.

Note from (9) that in order to project $w_p \in W$ to the space independent of $\alpha(x)$ using $w \in W$, we actually find an element $w_1 \in span(w)$ where $w_1(z) = \frac{\int_{-\infty}^{\infty} \alpha(x) w_p(g(x))dx}{\int_{-\infty}^{\infty} \alpha(x) w(g(x))dx} w(z)$ such that $\int_{-\infty}^{\infty} \alpha(x) w_p(g(x))dx = \int_{-\infty}^{\infty} \alpha(x) w_1(g(x))dx$ and then subtract this w_1 from w_p.

Next, considering the special case where $\alpha(x) = e_i(x)$ we have that for some $w_p \in W$ and for any operator $w \in W$ for which $\int_{-\infty}^{\infty} e_i(x) w(g(x)dx \neq 0$ the operator $Q_w^{e_i}$ is such that $\int_{-\infty}^{\infty} e_i(x) Q_w^{e_i}(w_p(g(x)))dx = 0$. Thus $Q_w^{e_i}(W)$ is the space of constraints on all the parameters except a_i, as for a_i these constraints are always null. Note the similarity of this procedure to the Gram-Schmidt procedure.

Theorem 3. *Let $\{\tilde{w}_i\}_{i=1}^P$ be an independent set as defined in Definition 1. Then for every i, the space of constraints on e_i, denoted by W_i, has the direct sum representation $W_i = W_0 \oplus span(\tilde{w}_i)$, where W_0 is the space of homogeneous constraints and $span(\tilde{w}_i)$ (i.e., the space of particular solutions for a_i) is a 1-D space spanned by the function $\tilde{w}_i$.*

Proof. Let $w \in W_i$. Next, we evaluate the projection of this $w \in W_i$ on the space independent of a_i, *i.e.*, on $Q_{\tilde{w}_i}^{e_i}(w)$ and verify that indeed $Q_{\tilde{w}_i}^{e_i}(w)$ imposes a homogenous constraint: For every j

$$\begin{aligned}
\int e_j(z) Q_{\tilde{w}_i}^{e_i}(w)[g(z)]dz &= \int e_j(z) \left(w(g(z)) - \frac{\int_{-\infty}^{\infty} e_i(x) w(g(x))dx}{\int_{-\infty}^{\infty} e_i(x) \tilde{w}_i(g(x))dx} \tilde{w}_i(g(z)) \right) dz \\
&= \int e_j(z) w(g(z))dz - \frac{\int_{-\infty}^{\infty} e_i(x) w(g(x))dx}{\int_{-\infty}^{\infty} e_i(x) \tilde{w}_i(g(x))dx} \int e_j(z) \tilde{w}_i(g(z))dz \\
&= 0 \qquad (10)
\end{aligned}$$

as both $\int e_j(z)w(g(z))dz = 0$ and $\int e_j(z)\tilde{w}_i(g(z))dz = 0$ since both $w, \tilde{w}_i \in W_i$, which is the space of constraints on e_i only. The above argument holds for every $j \neq i$. Evaluating (9) in the case where $i = j$ yields that $\int e_i(z)Q^{e_i}_{\tilde{w}_i}(w)[g(z)]dz = 0$. Hence, $D_{g,\mathbf{e}}(Q^{e_i}_{\tilde{w}_i}(w)) = \mathbf{0}$. In other words application of $Q^{e_i}_{\tilde{w}_i}(w)$, with $w \in W_i$, to $g(x)$ yields a P dimensional null vector. Hence, $Q^{e_i}_{\tilde{w}_i}(W_i) \subset W_0$. On the other hand, as every $w \in W_0$ yields a zero constraint with respect to every e_k, while $Q^{e_i}_{\tilde{w}_i}(W)$ only yields zero constraints on e_i, it is clear that $W_0 \subset Q^{e_i}_{\tilde{w}_i}(W_i)$. Hence, $W_0 = Q^{e_i}_{\tilde{w}_i}(W_i)$. Finally, by rewriting (9) we have that for any $w \in W$ we obtain $w(z) = Q^{e_i}_{\tilde{w}_i}[w(z)] + \frac{\int_{-\infty}^{\infty} e_i(x)w(g(x))dx}{\int_{-\infty}^{\infty} e_i(x)\tilde{w}_i(g(x))dx}\tilde{w}_i(z)$. Since all the operators $Q^{e_i}_{\tilde{w}_i}(W_i)$ provide null constraints on e_i while by definition $\tilde{w}_i$ yields a non-zero constraint on e_i, we conclude that $Q^{e_i}_{\tilde{w}_i}(W_i) \cap \text{span}(\tilde{w}_i) = 0$, and hence that $W_i = Q^{e_i}_{\tilde{w}_i}(W_i) \oplus \text{span}(\tilde{w}_i)$ □

Corollary 2. *Every operator $w \in W_i$ can be uniquely decomposed into the direct sum representation $w = w_0 + a\tilde{w}_i$ where $\{\tilde{w}_i\}_{i=1}^P$ is an independent set and w_0 is an operator, not necessarily unique, in W_0.*

The aim of the next discussion is to show that the foregoing conclusion regarding operators in W_i can be generalized to the entire space W. More specifically, we will show that any operator $w \in W$ admits a unique direct sum representation: $w = w_0 + \sum_{i=1}^{P} b_i\tilde{w}_i$ where $w_0 \in W_0$, and $\{\tilde{w}_i\}_{i=1}^P$ is an independent set of particular solutions for the elements of $\mathbf{a}$.

Definition 5. *Let $\{w_i\}_{i=1}^P$ be some set of operators, and let $w \in W_i$ be some other operator. Define the operator $M_{\{w_i\}_{i=1}^P} : W \to W$*

$$M_{\{w_i\}_{i=1}^P}[w(z)] = \sum_{i=1}^{P} \frac{\int_{-\infty}^{\infty} e_i(x)w(g(x))dx}{\int_{-\infty}^{\infty} e_i(x)w_i(g(x))dx} w_i(z). \tag{11}$$

In the special case where $\{\tilde{w}_i\}_{i=1}^P$ is an independent set (11) reduces to

$$M_{\{\tilde{w}_i\}_{i=1}^P}[w(z)] = \sum_{i=1}^{P} \left(\int_{-\infty}^{\infty} e_i(x)w(g(x))dx \right) \tilde{w}_i(z). \tag{12}$$

As we show next, given some independent set $\{\tilde{w}_i\}_{i=1}^P$ one can employ the operator $M_{\{\tilde{w}_i\}_{i=1}^P}$ in order to decompose any constraint into its components on the above independent set.

Theorem 4. *For any independent set $\{\tilde{w}_i\}_{i=1}^P$, some function $g(x) \in M$, and some $w \in W$, the map $L_{\{\tilde{w}_i\}_{i=1}^P}[w(z)] = w(z) - M_{\{\tilde{w}_i\}_{i=1}^P}[w(z)]$ is a projection of w on W_0.*

Proof. Let us evaluate the k-th element of the vector $D_{g,\{e_i\}_{i=1}^P}\left(M_{\{\tilde{w}_i\}_{i=1}^P}(w)\right)$:

$$\begin{aligned}
\left[D_{g,\{e_i\}_{i=1}^P}\left(M_{\{\tilde{w}_i\}_{i=1}^P}(w)\right)\right]_k &= \left[D_{g,\{e_i\}_{i=1}^P}\left(\sum_{i=1}^P\left(\int_{-\infty}^{\infty} e_i(x)w(g(x))dx\right)\ \tilde{w}_i(z)\right)\right]_k \\
&= \int_{-\infty}^{\infty} e_k(x)\sum_{i=1}^P\left(\int_{-\infty}^{\infty} e_i(x)w(g(x))dx\right)\ \tilde{w}_i(g(x))dx \\
&= \sum_{i=1}^P\left(\int_{-\infty}^{\infty} e_i(x)w(g(x))dx\right)\int_{-\infty}^{\infty} e_k(x)\tilde{w}_i(g(x))dx \\
&= \int_{-\infty}^{\infty} e_k(x)w(g(x))dx = \left[D_{g,\{e_i\}_{i=1}^P}(w)\right]_k \qquad (13)
\end{aligned}$$

where the forth equality is due to the fact that $\{\tilde{w}_i\}_{i=1}^P$ is an independent set. Hence, $D_{g,\{e_i\}_{i=1}^P}\left(w - M_{\{\tilde{w}_i\}_{i=1}^P}(w)\right) = \mathbf{0}$ and consequently $L_{\{\tilde{w}_i\}_{i=1}^P}[w(z)] = w(z) - M_{\{\tilde{w}_i\}_{i=1}^P}[w(z)] \in \mathrm{Ker}(D_{g,\{e_i\}_{i=1}^P})$. Since by definition $W_0 = \mathrm{Ker}(D_{g,\{e_i\}_{i=1}^P})$, we have that $L_{\{\tilde{w}_i\}_{i=1}^P}(w)(z) \in W_0$. On the other hand $w(z) - L_{\{\tilde{w}_i\}_{i=1}^P}[w(z)]$ is by definition $M_{\{\tilde{w}_i\}_{i=1}^P}[w(z)]$ which by (12) is a linear combination of non-zero constraints, and hence has a null intersection with W_0. □

In other words Theorem 4 asserts that $L_{\{\tilde{w}_i\}_{i=1}^N}$ is a projection operator on W_0 as for all $w \in W$, $L_{\{\tilde{w}_i\}_{i=1}^N}(L_{\{\tilde{w}_i\}_{i=1}^N}(w)) = L_{\{\tilde{w}_i\}_{i=1}^N}(w)$, $L_{\{\tilde{w}_i\}_{i=1}^N}(W) = W_0$. Finally, since for every $w \in W_0$ and for every i, $\int_{-\infty}^{\infty} e_i(x)w(g(x))dx = 0$, we conclude that $M_{\{\tilde{w}_i\}_{i=1}^P}(w) = 0$ and hence $L_{\{\tilde{w}_i\}_{i=1}^N}$ is the identity map on W_0.

Theorem 5. *Let the subspace defined by span$(\tilde{w}_i)$ be denoted by $\tilde{W}_i$. Fix g and the basis functions in $\mathbf{e}$ and assume there exists some set of operators $\{w_i\}_{i=1}^P \in W$ such that the corresponding matrix $\mathbf{G}(g,\mathbf{e},\mathbf{w})$ is full rank. Then W admits the following direct sum representation $W = \tilde{W}_1 \oplus \tilde{W}_2 \oplus \cdots \oplus \tilde{W}_P \oplus W_0$.*

Proof. Choose some $w \in W$. By the assumption that $\mathbf{G}(g,\mathbf{e},\mathbf{w})$ is full rank, there exists an independent set $\{\tilde{w}_i\}_{i=1}^P \in W$. Using Theorem 4 we have that $w(z) = M_{\{\tilde{w}_i\}_{i=1}^P}[w(z)] + L_{\{\tilde{w}_i\}_{i=1}^P}[w(z)]$ and that $M_{\{\tilde{w}_i\}_{i=1}^P}[w(z)] + L_{\{\tilde{w}_i\}_{i=1}^P}[w(z)] \in \tilde{W}_1 \oplus \tilde{W}_2 \oplus \cdots \oplus \tilde{W}_P \oplus W_0$. Hence, $W \subset \tilde{W}_1 \oplus \tilde{W}_2 \oplus \cdots \oplus \tilde{W}_P \oplus W_0$. On the other hand, $\tilde{W}_1 \oplus \tilde{W}_2 \oplus \cdots \oplus \tilde{W}_P \oplus W_0 \subset W$ □

We have thus shown that any constraint $w \in W$ can be represented by $w = w_0 + \sum_{i=1}^P b_i\tilde{w}_i$ with $w_0 \in W_0$. Therefore there is a multiplicity of constraints that are equivalent, although their homogeneous components are different.

Finally, we address the practical problem of constructively finding the elements of the decomposition. Suppose we are given some finite dimensional subset $W_M \subset W$, expressed in terms of a basis $W_M = span\left(\{w_i\}_{i=1}^M\right)$, and suppose that the matrix $\mathbf{G}(g, \mathbf{e}, \{w_i\}_{i=1}^M)$ is full rank. Hence, the geometric deformation can be estimated using the operators in W_M. Assume that we wish to describe the decomposition of W_M into particular and homogenous constraints. Taking a subset of $\{w_i\}_{i=1}^M$ such that $\mathbf{G}(g, \mathbf{e}, \mathbf{w})$ is invertible and constructing an independent set $\{\tilde{w}_i\}_{i=1}^P$ we obtain a particular solution for each one of the parameters in $\mathbf{a}$. All that remains to be done is to provide a concrete description of $W_0 \cap W_M$ in terms of the basis functions to yield an explicit description of the homogeneous constraints within W_M.

Theorem 6. *Let $\{w_i\}_{i=1}^M$ be a set of operators such that $W_M = span\left(\{w_i\}_{i=1}^M\right)$. Let $\{\tilde{w}_i\}_{i=1}^P \in W_M$ be an independent set. Then, $W_0 \cap W_M = span\left(\left\{L_{\{\tilde{w}_i\}_{i=1}^P}(w_i)\right\}_{i=1}^M\right)$.*

Proof. Using Theorem 4 we have that $span\left(\left\{L_{\{\tilde{w}_i\}_{i=1}^P}(w_i)\right\}_{i=1}^M\right) \subset W_0 \cap span(\{w_i\}_{i=1}^M)$. Next, let $w \in W_0 \cap W_M$. As $w \in W_M$ it admits the representation $w = \sum_{i=1}^M b_i w_i$. Since $L_{\{\tilde{w}_i\}_{i=1}^P}(w)$ is a linear operator, $L_{\{\tilde{w}_i\}_{i=1}^P}(w) = \sum_{i=1}^M b_i L_{\{\tilde{w}_i\}_{i=1}^P}(w_i)$. However, for every $w \in W_0$, $L_{\{\tilde{w}_i\}_{i=1}^P}(w)$ is the identity map, *i.e.*, $L_{\{\tilde{w}_i\}_{i=1}^P}(w) = w$. Hence, $w = L_{\{\tilde{w}_i\}_{i=1}^P}(w) = \sum_{i=1}^M b_i L_{\{\tilde{w}_i\}_{i=1}^P}(w_i) \in span\left(\left\{L_{\{\tilde{w}_i\}_{i=1}^P}(w_i)\right\}_{i=1}^M\right)$. □

We have thus derived a constructive procedure for representing the space $W_0 \cap W_M$ in terms of a set of basis functions, and hence a complete representation of any set of operators $\{w_i\}_{i=1}^M \in W$ in terms of a direct sum representation of particular and homogeneous constraints. In this representation each particular constraint $\tilde{w}_i$ provides a solution for a single deformation model parameter a_i.

3.1 Numerical Example: Piecewise Linear Deformations

Let $[A, B]$ be the support of the template $g(t)$, and let $r(t) = \begin{cases} t & 0 < t < 1 \\ 0 & t \le 0 \\ 1 & t \ge 1 \end{cases}$. Let $\{c_i\}_{i=0}^K$ be some division of $[A, B]$ such that $A = c_0 < c_1, \ldots, c_i < c_{i+1}, \ldots, c_K = B$. Let $E_k(t) = (c_k - c_{k-1}) r(\frac{t - c_{k-1}}{c_k - c_{k-1}})$. The piecewise linear transformation is then defined by $\phi(t) = \sum_{k=1}^K a_k E_k(t)$ where we assume for simplicity that $\phi(A) = 0$, while a_k is the slope of the function on the kth interval. The corresponding division of the range of $\phi(t)$ (which is the domain of

$\phi^{-1}(t)$) is given by $\tilde{c}_i = \sum_{k=1}^{i} a_k(c_k - c_{k-1})$. The slope of the inverse transformation on the kth interval is then given by $\frac{1}{a_k}$. Hence, $\phi^{-1}(t) = \sum_{k=1}^{K}(c_k - c_{k-1})r(\frac{t-\sum_{l=1}^{k-1} a_l(c_l-c_{l-1})}{a_k(c_k-c_{k-1})})$, and therefore it belongs to the same space of functions defined by the forward transformation. Thus the representation of the deformation function, or its inverse, are equivalent. As already indicated in Section 2.1, $\phi^{-1}(t)$ models a map from the coordinate system of $g(x)$ which is fixed, as $g(x)$ is the reference template. Hence, in the following example the deformation is defined in terms of the inverse transformation $\phi^{-1}(t)$.

Example 1: To illustrate the results and conclusions of the derivation in the preceding sections we consider the following example. Figure 1 provides four different examples (one in each row) of deformations of the same template and the results of the estimation procedure. The leftmost figure in each row depicts the template and the deformed observation. The error between the observed signal and its reconstruction obtained by applying the estimated deformation $\hat{\phi}(x)$ to the template, is shown to its right. The figures on the right depict the errors in estimating the deformation: $\phi(x) - \hat{\phi}(x)$, and its derivative: $\phi'(x) - \hat{\phi}'(x)$, respectively. In order to approximate the above derivation, performed in the continuum, the observations are densely sampled (10000 points on $[0, 1]$) and their amplitude are quantized into 2^{32} levels. The deformation is piecewise linear over a random number K of intervals. The initial choice of the nonlinear operators is $w_i(y) = \begin{cases} 1 & \frac{(i-1)}{K} < y \leq \frac{i}{K} \\ 0 & \text{otherwise} \end{cases}$. Obviously, integrals are replaced by summations of the sampled and quantized signals and hence we observe small errors in evaluating them which result in small errors (of the same order as the sampling step) in the estimated deformations.

Example 2: Next, we consider an example where the parameter vector of the inverse deformation is given by $\mathbf{a} = [1.71, 1.54, 1.07, 0.79]$, and the interval division is given by $[0, 0.25, 0.5, 0.75, 1]$. The template $g(t)$ is depicted in Fig. 2. In this example we chose the nonlinear operators such that $w_i(y) = y^i,\ i = 1, \ldots, 4$. The four particular solutions $\tilde{w}_i(y)$ corresponding to the four parameters in $\mathbf{a}$ for the initial choice $w_i(y) = y^i$ are depicted in Figure 3. The four particular solutions $\tilde{w}_i(y)$ corresponding to the four parameters in $\mathbf{a}$ for the initial choice $w_i(y) = sin(2\pi iy)$ are depicted in Figure 4. In each case, and regardless of the initial choice of the nonlinear operators, we have for every i that $a_i = \int \tilde{w}_i(h(y))dy$. Example homogeneous constraints are depicted in Fig. 5.

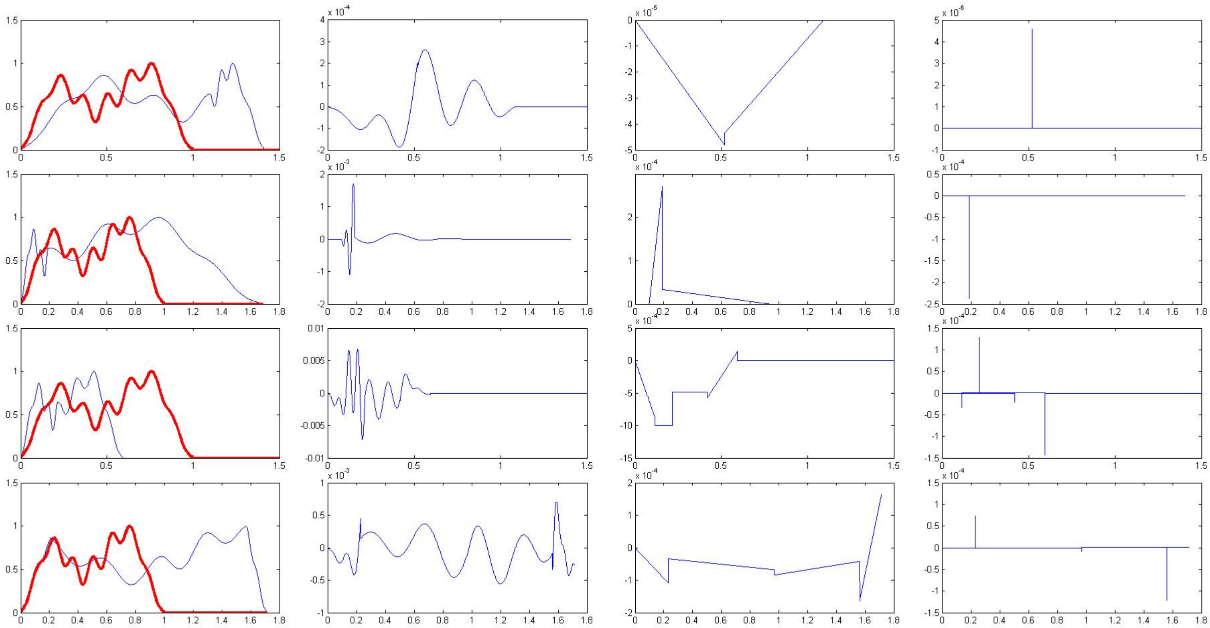

Figure 1: From left to right, in each row: The template (red) and observation (blue); The error in estimating the deformed observation: $h(x) - g(\hat{\phi}(x))$; The error in estimating the deformation: $\phi(x) - \hat{\phi}(x)$; The error in estimating the deformation derivative: $\phi'(x) - \hat{\phi}'(x)$.

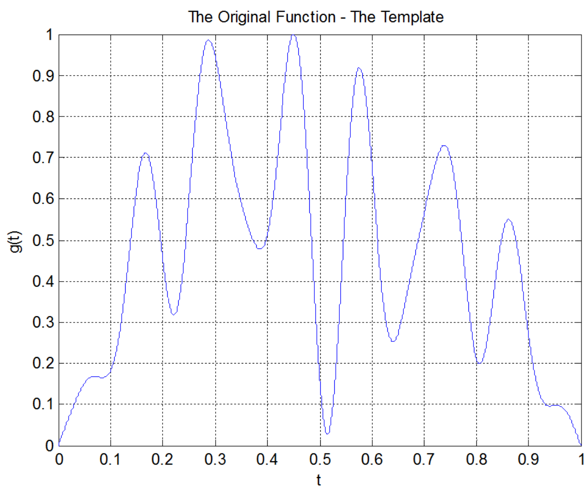

Figure 2: The template.

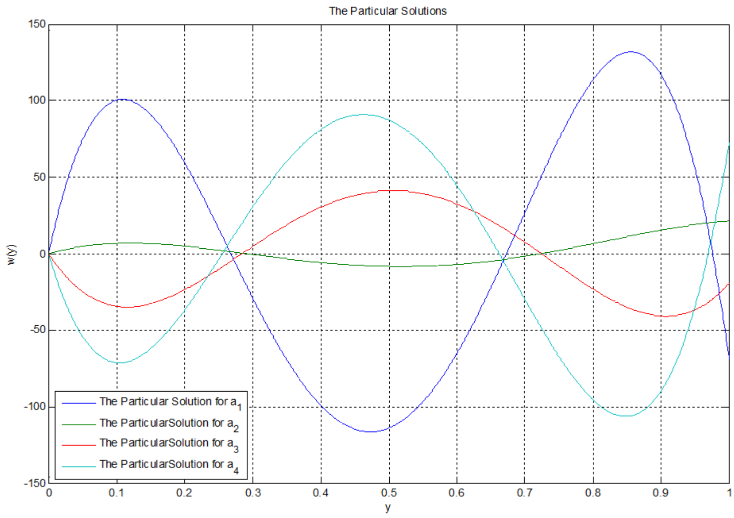

Figure 3: The four particular solutions for the initial choice $w_i(y) = y^i$. Blue: $\tilde{w}_1(y)$, the particular solution for a_1; Green: $\tilde{w}_2(y)$, the particular solution for a_2; Red: $\tilde{w}_3(y)$, the particular solution for a_3; Cyan: $\tilde{w}_4(y)$, the particular solution for a_4.

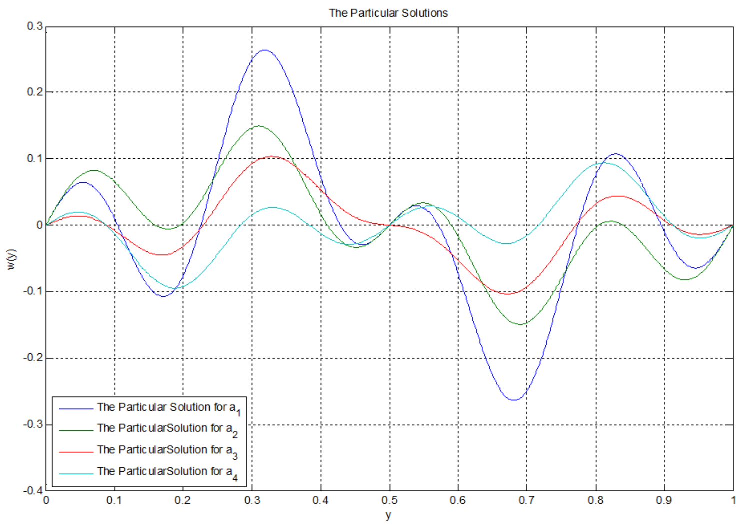

Figure 4: The four particular solutions for the initial choice $w_i(y) = sin(2\pi iy)$. Blue: $\tilde{w}_1(y)$, the particular solution for a_1; Green: $\tilde{w}_2(y)$, the particular solution for a_2; Red: $\tilde{w}_3(y)$, the particular solution for a_3; Cyan: $\tilde{w}_4(y)$, the particular solution for a_4.

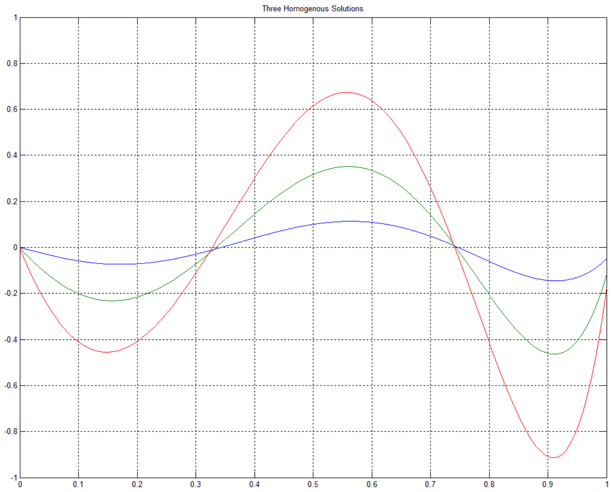

Figure 5: Example homogeneous solutions.

4 Observations Subject to Additive Noise: Unbiased Estimation

In the presence of noise the observed data is given by

$$h(x) = g(\phi(x)) + \eta(x) \ . \tag{14}$$

Assuming that the noise has a zero mean, and that its higher order statistics are known, we first address questions related to issue of the *optimal choice of the set* $\{w_k\}$ *for each template function* g. We begin by adapting the solution derived in the previous section for the deterministic case, to a least squares solution for the model parameters. In the presence of noise the basic equation (1) becomes

$$\begin{aligned} \int_{-\infty}^{\infty} w_k(h(x))dx &= \int_{-\infty}^{\infty} w_k[g(\phi(x)) + \eta(x)]dx \\ &= \int_{-\infty}^{\infty} w_k[g(z) + \eta(\phi^{-1}(z))](\phi^{-1})'(z)dz \\ &= \int_{-\infty}^{\infty} w_k(g(z))(\phi^{-1})'(z)dz + \epsilon_k^g \end{aligned} \tag{15}$$

where we define the random variable

$$\begin{aligned}\epsilon_k^g &= \int_{-\infty}^{\infty} \Big(w_k[g(z) + \eta(\phi^{-1}(z))] - w_k(g(z)) \Big) (\phi^{-1})'(z) dz \\ &= \int_{-\infty}^{\infty} \Big(w_k[g(\phi(x)) + \eta(x)] - w_k(g(\phi(x))) \Big) dx \end{aligned} \tag{16}$$

Substituting (3) into (15), we obtain the linear system of equations

$$\int_{-\infty}^{\infty} w_k(h(x)) dx = \sum_{i=1}^{P} a_i \int_{-\infty}^{\infty} e_i(x) w_k(g(x)) dx + \epsilon_k^g \quad k = 1, \ldots, N \tag{17}$$

The system (17) represents a linear regression problem where the noise sequence $\{\epsilon_k^g\}$ is non-stationary since its statistical moments depend on the choice of w_k for each k. The regressors are functions of w_k and the template g, and hence are *deterministic*. Provided that the sequence of composition functions $\{w_k\}$ is chosen such that the resulting regressors matrix is full rank, the system (17) is solved by a linear least squares method such that the l_2 norm of the noise vector is minimized.

The dependence of the noise sequence $\{\epsilon_k^g\}$ on the choice of w_k suggests that different choices of the composition sequence $\{w_k\}$ may provide different solutions. We shall be first interested in systems such that for each k, the linear constraint imposed by w_k is unbiased (and thus the "effective noise" that corresponds to each w_k is zero mean).

4.1 Construction of Unbiased Linear Constraints

Consider the case where we choose $w_k(x) = \sum_l \alpha_l^k x^l$, and the additive noise is an i.i.d. process with zero mean and variance σ^2. We next evaluate the mean term, $E\epsilon_k^g$, of the "effective noise". To simplify the notation we will take advantage of the linear structure of $w_k(x)$, and analyze first only the case where $w_k(x) = x^k$. Thus, in this case

$$\begin{aligned}\epsilon_k^g &= \int_{-\infty}^{\infty} \{[g(z) + \eta(\phi^{-1}(z))]^k - g^k(z)\} (\phi^{-1})'(z) dz \\ &= \int_{-\infty}^{\infty} \sum_{j=1}^{k} \binom{k}{j} g^{k-j}(z) \eta^j(\phi^{-1}(z)) (\phi^{-1})'(z) dz \end{aligned} \tag{18}$$

Since the additive noise is i.i.d., $E[\eta^j(\phi^{-1}(z))] = E[\eta^j(z)]$ for every j. Thus we have

$$\begin{aligned} E\epsilon_k^g &= \sum_{i=1}^{P} a_i \int_{-\infty}^{\infty} e_i(z) \left(\sum_{j=1}^{k} \binom{k}{j} g^{k-j}(z) E[\eta^j(\phi^{-1}(z))] \right) dz \\ &= \sum_{i=1}^{P} a_i \int_{-\infty}^{\infty} e_i(z) \left(\sum_{j=1}^{k} \binom{k}{j} g^{k-j}(z) E[\eta^j(z)] \right) dz \end{aligned} \tag{19}$$

Hence, for the case where $w_k(x) = x^k$, we can rewrite (17) in the form:

$$\begin{aligned} \int_{-\infty}^{\infty} h^k(x)dx &= \sum_{i=1}^{P} a_i \Bigg\{ \int_{-\infty}^{\infty} e_i(z) g^k(z) dz + \int_{-\infty}^{\infty} e_i(z) \left(\sum_{j=1}^{k} \binom{k}{j} g^{k-j}(z) E[\eta^j(z)] \right) dz \Bigg\} + \bar{\epsilon}_k^g \\ &= \sum_{i=1}^{P} a_i \Bigg\{ \int_{-\infty}^{\infty} e_i(z) \left(\sum_{j=0}^{k} \binom{k}{j} g^{k-j}(z) E[\eta^j(z)] \right) dz \Bigg\} + \bar{\epsilon}_k^g \end{aligned} \tag{20}$$

where $\bar{\epsilon}_k^g$ is a zero mean random variable. Comparing the expression in (20) to (3) and (4) of the deterministic case, it is clear that $\int_{-\infty}^{\infty} e_i(z)g^k(z)dz$ is the (k, i) entry of the matrix $\mathbf{G}(g, \mathbf{e}, \mathbf{w})$ for the choice $w_k(x) = x^k$, while the term $\int_{-\infty}^{\infty} e_i(z) \left(\sum_{j=1}^{k} \binom{k}{j} g^{k-j}(z)E[\eta^j(z)]\right) dz$ is a deterministic correction term for the (k, i) entry of $\mathbf{G}(g, \mathbf{e}, \mathbf{w})$, due to the noise contribution, such that the noise term $\bar{\epsilon}_k^g$ has a zero mean. Thus, let $\mathbf{P}(\epsilon, \mathbf{e}, \mathbf{w})$ be the matrix whose (k, i) entry is $\int_{-\infty}^{\infty} e_i(z) \left(\sum_{j=1}^{k} \binom{k}{j} g^{k-j}(z)E[\eta^j(z)]\right) dz$, and let $\bar{\boldsymbol{\epsilon}} = [\bar{\epsilon}_1^g, \ldots, \bar{\epsilon}_N^g]^T$. Then, provided that $\mathbf{G}_\epsilon(g, \mathbf{e}, \mathbf{w}) = \mathbf{G}(g, \mathbf{e}, \mathbf{w}) + \mathbf{P}(\epsilon, \mathbf{e}, \mathbf{w})$ is a full rank matrix, there exists an unbiased least-squares solution for $\mathbf{a}$, such that the norm of $\bar{\boldsymbol{\epsilon}}$ is minimized. Due to the linearity of the constraints the above conclusion holds for the case where $w_k(x) = \sum_{l=0}^{L} \alpha_l^k x^l$, as well. In this case, the (k, i) entry of $\mathbf{P}(\epsilon, \mathbf{e}, \mathbf{w})$ is given by $\int_{-\infty}^{\infty} e_i(z) \sum_{l=0}^{L} \alpha_l^k \left(\sum_{j=1}^{l} \binom{k}{j} g^{l-j}(z)E[\eta^j(z)]\right) dz$.

Extending the concept, developed in previous sections, of obtaining particular solutions for each a_i (constraints on e_i only) to the case where the observations are noisy, we are interested in obtaining *unbiased particular solutions* and thus an unbiased LWE. We have the following:

Lemma 8. *Let $h(x) = g(\phi(x)) + \eta(x)$, where $\eta(x)$ is an i.i.d. process with zero mean and variance σ^2. For a set of functions $\{w_k\}_{k=1}^N \in W$ such that $w_k(x) = \sum_{l=0}^{L} \alpha_l^k x^l$, where $\mathbf{G}_\epsilon(g, \mathbf{e}, \mathbf{w})$ is full rank, there exists a corresponding set of functions $\{\tilde{w}_i\}_{i=1}^P \in W$ where $\tilde{w}_i \in span\left(\{w_i\}_{i=1}^N\right)$ such that for every i, $\int \tilde{w}_i(h(x)) = a_i + \epsilon_i(g, \tilde{\mathbf{w}}, \mathbf{e})$, and $\epsilon_i(g, \tilde{\mathbf{w}}, \mathbf{e})$ has zero mean.*

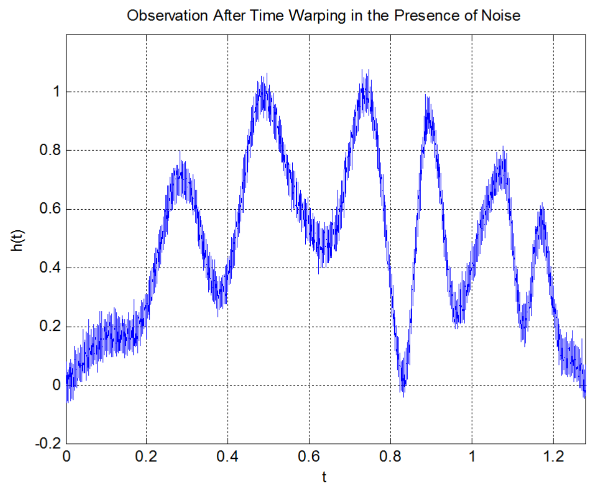

Figure 6: Noisy observation

Proof. Set $\mathbf{A} = (\mathbf{G}_\epsilon^T \mathbf{G}_\epsilon)^{-1}\mathbf{G}_\epsilon^T$. Then similarly to Lemma 2, $\mathbf{I} = \mathbf{A}\mathbf{G}_\epsilon(g, \mathbf{e}, \mathbf{w}) = \mathbf{G}_\epsilon(g, \mathbf{e}, \mathbf{A}\mathbf{w}) = \mathbf{G}_\epsilon(g, \mathbf{e}, \tilde{\mathbf{w}})$ where $\tilde{\mathbf{w}} = \mathbf{A}\mathbf{w}$. Let $\tilde{\boldsymbol{\epsilon}} = [\epsilon_1(g, \tilde{\mathbf{w}}, \mathbf{e}), \ldots, \epsilon_P(g, \tilde{\mathbf{w}}, \mathbf{e})]^T$. Rewriting (20) in a matrix form and multiplying it from the left by $\mathbf{A}$, the proof follows where $\tilde{\boldsymbol{\epsilon}} = \mathbf{A}\bar{\boldsymbol{\epsilon}}$ and hence is zero mean. $\square$

Note that the system (20) represents a linear regression problem where the observation noise is *non-stationary*, but with a zero mean. The regressors are functions of $\{w_k\}$, the template g, and the known statistics of the noise. Hence the regressors are *deterministic*.

4.1.1 Numerical Example

Example 1: Consider the same setting as in Example 2 of Section 3.1, yet in the present case the observation is subject to an additive zero-mean white Gaussian noise with standard deviation $\sigma = 0.03$. An example realization is depicted in Fig. 6. In this case we have chosen again the nonlinear operators to be $w_i(y) = y^i$, $i = 1, \ldots 4$. Following the derivation in Section 4.1, we have for the setting of this example that

$$\mathbf{G}(g, \mathbf{e}, \mathbf{w}) = \begin{bmatrix} 0.082 & 0.179 & 0.123 & 0.072 \\ 0.038 & 0.138 & 0.077 & 0.029 \\ 0.020 & 0.113 & 0.053 & 0.014 \\ 0.0122 & 0.096 & 0.039 & 0.007 \end{bmatrix}$$

while the deterministic correction matrix due to the presence of noise is given by

$$\mathbf{P}(\epsilon, \mathbf{e}, \mathbf{w}) = 10^{-3} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0.225 & 0.225 & 0.225 & 0.225 \\ 0.222 & 0.484 & 0.334 & 0.194 \\ 0.207 & 0.750 & 0.417 & 0.160 \end{bmatrix}$$

where for example, the first row of $\mathbf{P}(\epsilon, \mathbf{e}, \mathbf{w})$ results from substituting $k = 1$ which yields for $i = 1, \ldots, 4$, that $\int_{-\infty}^{\infty} e_i(z) g^0(z) E[\eta(z)] dz = 0$ since the noise is zero mean.

Example 2: The performance of the estimation algorithm is tested using a sequence of 100,000 Monte-Carlo experiments. In each experiment, each of the deformation parameters (the interval slopes) is drawn from a uniform distribution on $[1, 3]$. Then, in order to guarantee that the probability of shrinking the interval or expanding it are equal, the slope parameter is substituted by its reciprocal with probability of 0.5. The bias of the estimated parameter vector for the case where the estimates were obtained by a least squares solution of the system (17) without introducing the bias correction is given by $[2.450, 0.045, -0.558, -1.958]$. However, employing the unbiased linear constraints by incorporating the deterministic correction matrix into the solution and solving the system (20), using the least squares method, we find that the bias of the estimated parameter vector is now $10^{-3}[0.253, 0.126, -0.290, -0.087]$ and the standard deviation $[0.3114, 0.0173, 0.0775, 0.2550]$ which verifies experimentally that indeed the estimator (20) is unbiased.

4.2 Minimum Variance Unbiased Estimators

Consider the setting defined in Lemma 8, *i.e.*, we have that $h(x) = g(\phi(x)) + \eta(x)$, where $\eta(x)$ is an i.i.d. process with zero mean and variance σ^2, and we have a set of functions $\{w_k\}_{k=1}^N \in W$ such that $w_k(x) = \sum_{l=0}^{L} \alpha_l^k x^l$, where $\mathbf{G}_\epsilon(g, \mathbf{e}, \mathbf{w})$ is full rank. As concluded in Lemma 8, there exists a corresponding set of functions $\{\tilde{w}_k\}_{k=1}^P \in W$ where $\tilde{w}_k \in$ *span* $\left(\{w_k\}_{k=1}^N\right)$ such that for every k, $\int \tilde{w}_k(h(x)) = a_k + \epsilon_k(g, \tilde{\mathbf{w}}, \mathbf{e})$, and $\epsilon_k(g, \tilde{\mathbf{w}}, \mathbf{e})$ has zero mean. Hence, we define

$$\hat{a}_k = \int \tilde{w}_k(h(x)) \tag{21}$$

to be our estimator of a_k. Using Lemma 8 we also have that $\tilde{w}_k(x)$ admits the following representation $\tilde{w}_k(x) = \sum_{l=0}^{L} p_l^k x^l$. Our goal is therefore to find, jointly for all k, the coefficients of the minimum variance estimator of a_k among all the unbiased estimators resulting from applying to the observed data operators of the form $w_k(x) = \sum\limits_{l=0}^{L} \alpha_l^k x^l$.

Due to the polynomial structure of $\tilde{w}_k$ we have following the same considerations as in the previous section that

$$\begin{aligned}
\int\limits_{-\infty}^{\infty} \tilde{w}_k(h(x))dx &= \int\limits_{-\infty}^{\infty} \sum_{l=0}^{L} p_l^k \,(h(x))^l dx = \int\limits_{-\infty}^{\infty} \sum_{l=0}^{L} p_l^k \,(g(\phi(x)) + n(x))^l dx \\
&= \sum_{i=1}^{P} a_i \int\limits_{-\infty}^{\infty} e_i(x) \sum_{l=0}^{L} p_l^k \left(\sum_{j=0}^{l} \binom{l}{j} g(x)^{l-j} n(\phi^{-1}(x))^j \right) dx \\
&= \mathbf{V}_k^T \mathbf{a} \qquad (22)
\end{aligned}$$

where $\mathbf{V}_k$ is a P-dimensional column vector whose ith entry is given by $[\mathbf{V}_k]_i = \int\limits_{-\infty}^{\infty} e_i(x) \sum\limits_{l=0}^{L} p_l^k \left(\sum\limits_{j=0}^{l} \binom{l}{j} g(x)^{l-j} n(\phi^{-1}(x))^j \right) dx.$

Using the notation defined in (6), $\int\limits_{-\infty}^{\infty} \tilde{w}_k(h(x))dx$ is the kth element of $\mathbf{h}(h, \mathbf{w})$ for the above choice of $\{\tilde{w}_k(h(x))\}_{k=1}^{P}$. Hence, rewriting (22) in a matrix from we have $\mathbf{h}(h, \mathbf{w}) = \mathbf{V}\mathbf{a}$, where $\mathbf{V}_k^T$ is the k-th row of $\mathbf{V}$.

Since for every k, $\hat{a}_k = \int \tilde{w}_k(h(x)) = \mathbf{h}_k(h, \mathbf{w})$ we have

$$\begin{aligned}
\mathrm{cov}(\mathbf{a}) &= E[(\mathbf{h}(h, \mathbf{w}) - E(\mathbf{h}(h, \mathbf{w}))][(\mathbf{h}(h, \mathbf{w}) - E(\mathbf{h}(h, \mathbf{w}))]^T \\
&= \mathbf{a}^T E[(\mathbf{V} - E(\mathbf{V}))][(\mathbf{V} - E(\mathbf{V})]^T \mathbf{a} \qquad (23)
\end{aligned}$$

Our goal in this framework is to find $\{\tilde{w}_k(h(x))\}_{k=1}^{P}$ such that $\mathrm{tr}\{\mathrm{cov}(\mathbf{a})\}$ is minimized. Therefore we have to evaluate $E[(\mathbf{V}_k - E(\mathbf{V}_k))][(\mathbf{V}_k - E(\mathbf{V}_k)]^T$.

By the construction in Lemma 8 $\tilde{w}_k$ is a particular solution for a_k such that $E[\epsilon_k(g, \tilde{\mathbf{w}}, \mathbf{e})] = 0$. (This is since $\int \tilde{w}_k(h(x)) = a_k + \epsilon_k(g, \tilde{\mathbf{w}}, \mathbf{e})$). Hence, for $i = k$, $E[\mathbf{V}_k]_k = 1$, while for $i \neq k$,$E[\mathbf{V}_k]_i = 0$. Hence,

$$[\mathbf{V}_k - E(\mathbf{V}_k)]_i = \begin{cases} \int\limits_{-\infty}^{\infty} e_i(x) \sum\limits_{l=0}^{L} p_l^k \left(\sum\limits_{j=0}^{l} \binom{l}{j} g(x)^{l-j} n(\phi^{-1}(x))^j \right) dx & i \neq k \\ \int\limits_{-\infty}^{\infty} e_k(x) \sum\limits_{l=0}^{L} p_l^k \left(\sum\limits_{j=0}^{l} \binom{l}{j} g(x)^{l-j} n(\phi^{-1}(x))^j \right) dx - 1 & i = k \end{cases} \qquad (24)$$

Then it can be easily verified that

Lemma 9.

$$E\{[\mathbf{V}_k - E(\mathbf{V}_k)]_n \ [\mathbf{V}_k - E(\mathbf{V}_k)]_m\} = \tag{25}$$
$$\begin{cases} \int\limits_{-\infty}^{\infty} e_n(x)e_m(x) \sum\limits_{l=0}^{L} p_l^k \sum\limits_{q=0}^{l} \binom{l}{q} \sum\limits_{r=0}^{L} p_r^k \sum\limits_{t=0}^{r} \binom{r}{t} g(x)^{l+r-q-t}\sigma^{q+t}dx & n \neq k \ or \ m \neq k \\ \int\limits_{-\infty}^{\infty} e_n(x)e_m(x) \sum\limits_{l=0}^{L} p_l^k \sum\limits_{q=0}^{l} \binom{l}{q} \sum\limits_{r=0}^{L} p_r^k \sum\limits_{t=0}^{r} \binom{r}{t} g(x)^{l+r-q-t}\sigma^{q+t}dx - 1 & n = m = k \end{cases}$$

Thus, we have expressed tr{cov($\mathbf{a}$)} in terms of the unknown coefficients $\{p_l^k\}$. By using numerical minimization, such as the Newton-Raphson method, in order to minimize tr{cov($\mathbf{a}$)} with respect to the unknown parameters, a minimum variance LSE for $\mathbf{a}$ is found.

4.2.1 Numerical Example

Example 1: To illustrate the performance of the minimum variance LSE, we depict in Figures 7-10 the results of applying the proposed minimum variance unbiased estimator for signals observed with different deformations and different noise levels. Each of the figures depict (from right to left) the template (red), the deformed template $h(x) = g(\phi(x))$ (cyan); noisy observation (blue); Original and estimated deformed noiseless template: $h(x)$ and $\hat{h}(x)$; The deformation and its estimate: $\phi(x)$ and $\hat{\phi}(x)$; The deformation derivative and its estimate: $\phi'(x)$ and $\hat{\phi}'(x)$. Note that in each of the five examples the deformation parameters are different as they are sampled from a random distribution, as detailed in Example 2 of Section 4.1.1. In all the examples we have chosen the nonlinear operators to be $w_i(y) = y^i$, $i = 1, \ldots 200$, and we look for the minimum variance unbiased estimator of $\mathbf{a}$.

Example 2: Next, we consider again the same setting as discussed in Example 1 of Section 4.1.1. We have chosen the nonlinear operators to be $w_i(y) = y^i$, $i = 1, \ldots 200$, and we look for the minimum variance unbiased estimator of $\mathbf{a}$. The four particular solutions $\tilde{w}_i(y)$ corresponding to the four parameters are depicted in Figure 11.

Example 3: Finally, we repeat with the minimum variance unbiased LWE, the same Monte-Carlo experiment described in Example 2 of Section 4.1.1. The standard deviation of the estimation error using the minimum variance unbiased estimator of $\mathbf{a}$ is given by $[0.0460, 0.0121, 0.0287, 0.0372]$, which is smaller than that obtained by any of the previous methods.

Example 4: As indicated in the Introduction, the DTW provides the best piece-

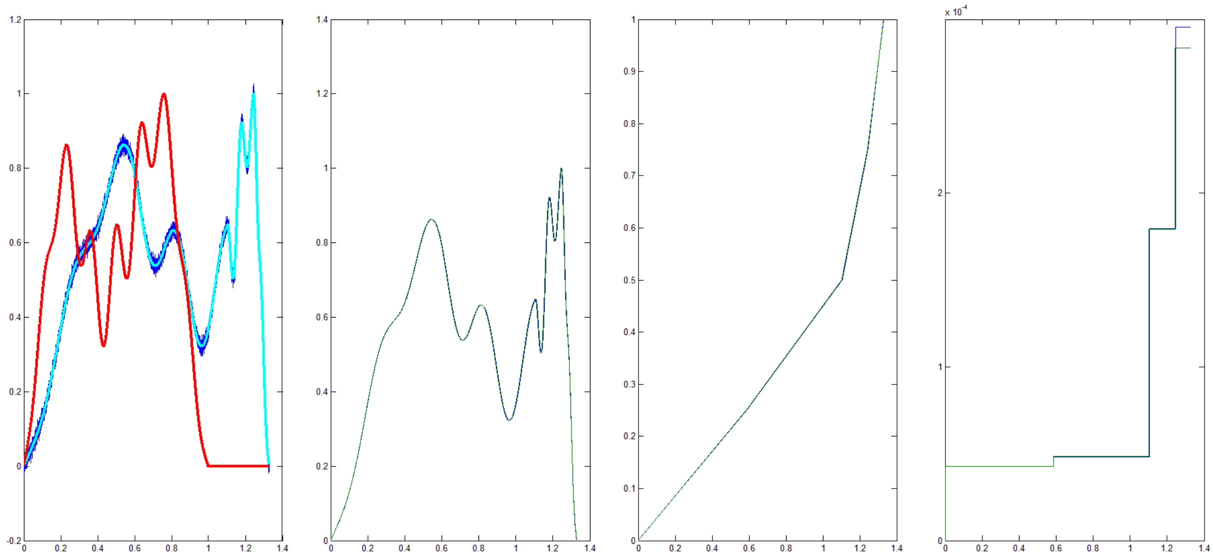

Figure 7: From left to right: The template (red), the deformed template $h(x) = g(\phi(x))$ (cyan); noisy observation (blue) where the noise standard deviation is 0.01; Original and estimated deformed noiseless template: $h(x)$ and $\hat{h}(x)$; The deformation and its estimate: $\phi(x)$ and $\hat{\phi}(x)$; The deformation derivative and its estimate: $\phi'(x)$ and $\hat{\phi}'(x)$.

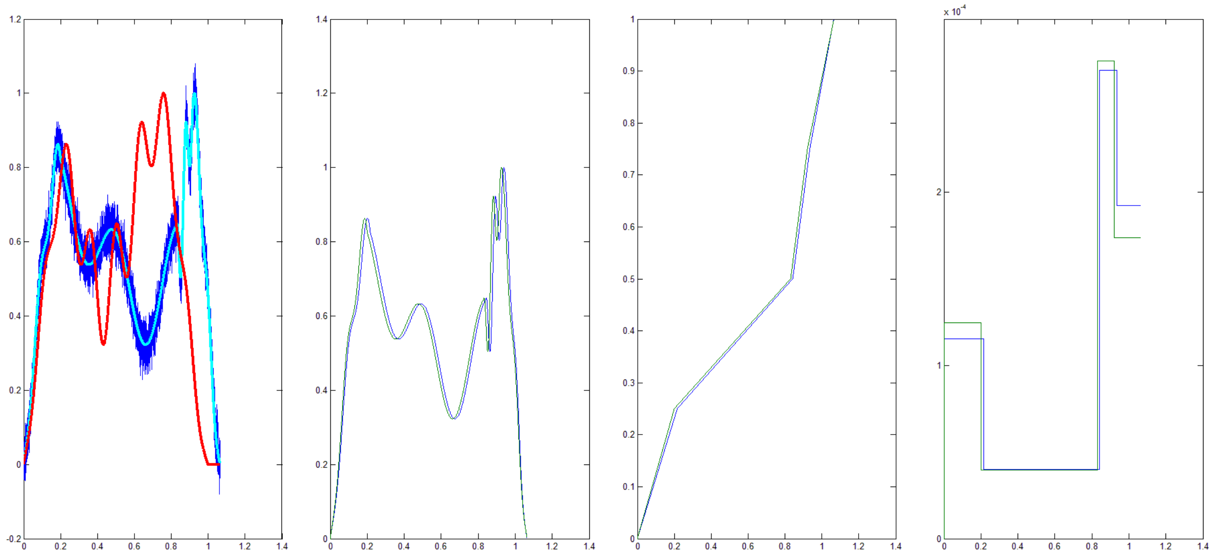

Figure 8: From left to right: The template (red), the deformed template $h(x) = g(\phi(x))$ (cyan); noisy observation (blue) where the noise standard deviation is 0.03; Original and estimated deformed noiseless template: $h(x)$ and $\hat{h}(x)$; The deformation and its estimate: $\phi(x)$ and $\hat{\phi}(x)$; The deformation derivative and its estimate: $\phi'(x)$ and $\hat{\phi}'(x)$.

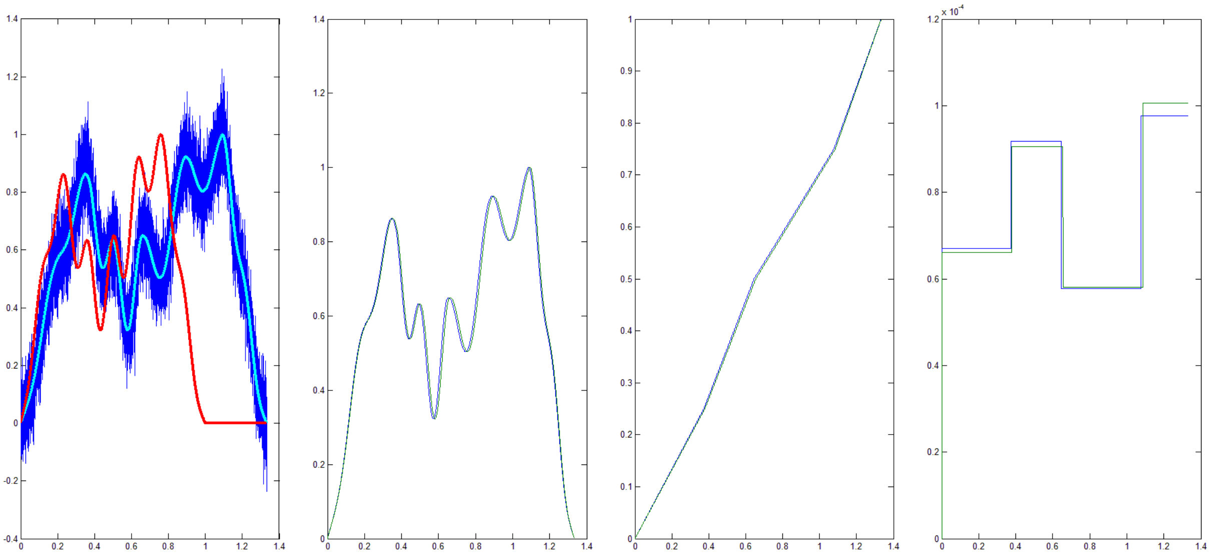

Figure 9: From left to right: The template (red), the deformed template $h(x) = g(\phi(x))$ (cyan); noisy observation (blue) where the noise standard deviation is 0.07; Original and estimated deformed noiseless template: $h(x)$ and $\hat{h}(x)$; The deformation and its estimate: $\phi(x)$ and $\hat{\phi}(x)$; The deformation derivative and its estimate: $\phi'(x)$ and $\hat{\phi}'(x)$.

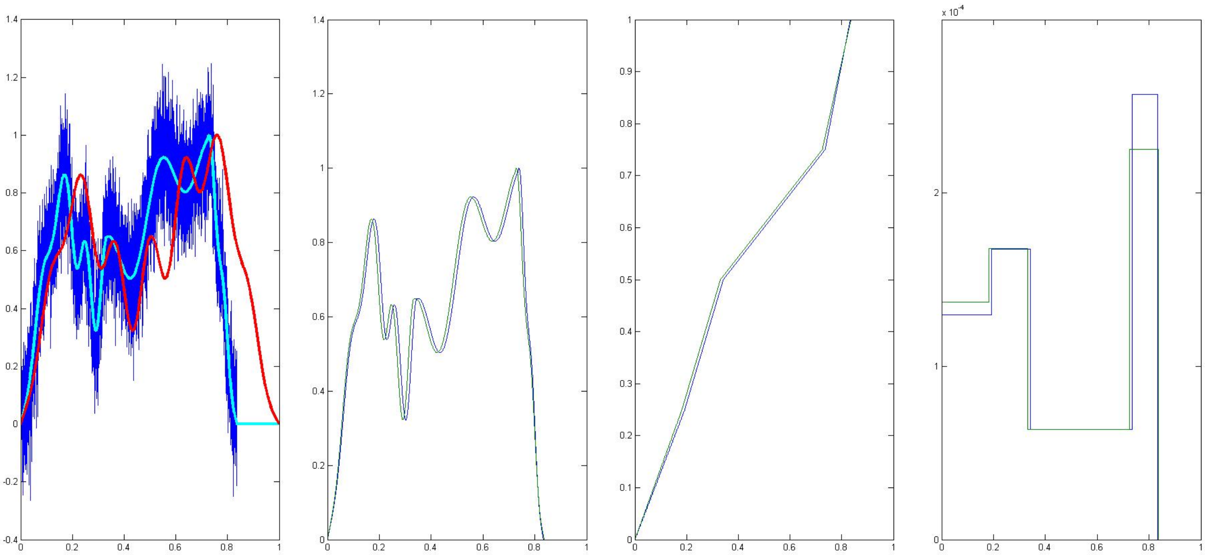

Figure 10: From left to right: The template (red), the deformed template $h(x) = g(\phi(x))$ (cyan); noisy observation (blue) where the noise standard deviation is 0.1; Original and estimated deformed noiseless template: $h(x)$ and $\hat{h}(x)$; The deformation and its estimate: $\phi(x)$ and $\hat{\phi}(x)$; The deformation derivative and its estimate: $\phi'(x)$ and $\hat{\phi}'(x)$.

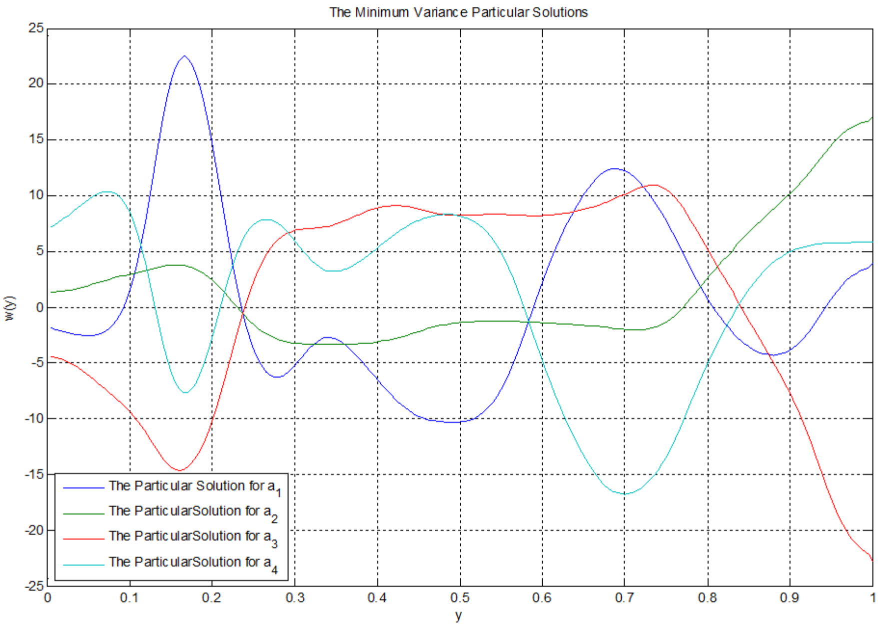

Figure 11: The four particular solutions that provide the minimum variance unbiased estimator. Blue: $\tilde{w}_1(y)$, the particular solution for a_1; Green: $\tilde{w}_2(y)$, the particular solution for a_2; Red: $\tilde{w}_3(y)$, the particular solution for a_3; Cyan: $\tilde{w}_4(y)$, the particular solution for a_4.

wise linear approximation of the deformation function on a discrete grid, with respect to the ℓ_2 norm. This solution is in fact an efficient implementation of a grid search for the deformation function, when both the template and the observed functions are provided on a discrete grid. Hence, in terms of minimizing the ℓ_2 norm for piece-wise linear deformations, this solution is the optimal. However this optimality in performance is achieved at the cost of huge memory and computational requirements as the algorithm holds a matrix whose dimensions are $M_T \times M_O$ where M_T and M_O are the lengths of the template and the observed signals, respectively. Thus in cases where the length of the signal is long, the memory and computational requirements of the DTW make it impractical. On the other hand, the memory and computational requirements of the linear method proposed in this paper are minimal: Only P functions $\{\tilde{w}_k\}_{k=1}^{P}$ and the observed signal need to be stored, while the solution is computed based on the summation (integrals computations) of P vectors ($P = 4$ in the examples). In the following example the lengths of the template and observed signals were 1000 samples. The average computation time of the deformation using the proposed LWE was 10000 times faster than that required by the DTW. In Figure 12 we compare the statistical performance of the proposed linear method with that of the DTW, using a sequence of 100 Monte-Carlo experiments at each noise level. Since the DTW minimizes the ℓ_2 distance between the template and observed realization, while the proposed method is parametric and hence minimizes the ℓ_2 distance in the parameter space of the deformation model, we chose to compare the performance of the two methods using a metric that measures the maximal distance between the true and estimated deformations, *i.e.*, $Q = ||\phi(x) - \hat{\phi}(x)||_\infty$, averaged over the set of Monte-Carlo experiments. In the experiment we compare the performance of the proposed LWE for different choices of the number of non-linear functionals employed, to that of the DTW. The number of non-linear functionals employed varies from $N = 5$, which is very close to the minimal possible number as the model order is $P = 4$, to $N = 10$ and $N = 50$. As expected, the performance of the proposed LWE improves with increasing the number of employed nonlinear functionals. The experimental results indicate that when the observation noise is low, and the number of employed nonlinear functionals is larger than P which is the minimal number that guarantees the existence of a solution in the deterministic case, the proposed linear method outperforms the DTW in all aspects: accuracy, memory requirements, and computational requirements. However, in cases where the observation noise is high, the DTW achieves better performance than the proposed method. This is the result of restricting the proposed method to use only small numbers of non linear functionals on the data. When the noise level is high the computed functionals are noisy themselves, and thus a larger number of

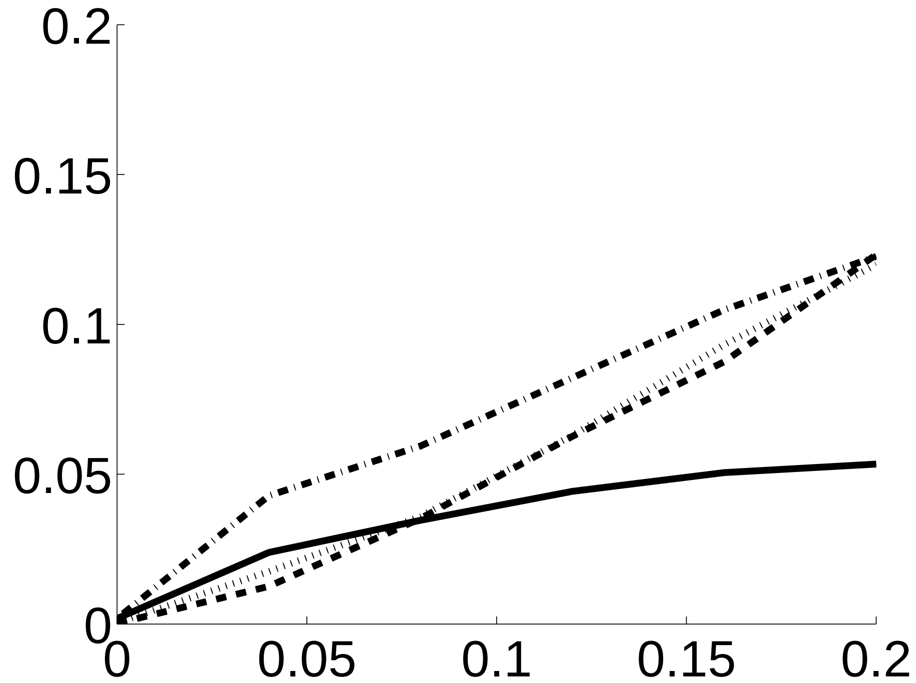

Figure 12: The averaged maximal distance between the true and estimated deformations, *i.e.*, $Q = ||\phi(x) - \hat{\phi}(x)||_\infty$, for the DTW (solid line) and the proposed minimum variance unbiased LWE as a function of the number, N, of employed non-linear functionals: $N = 5$ dashed-dotted line; $N = 10$ dotted line; $N = 50$ dashed line.

functionals is required in order to reduce the error of the linear solution. Since in the experiment these numbers are held low, we observe the phenomenon of a larger estimation error as the noise level increases. Obviously, if due to some reason the number of employed non-linear functionals cannot be increased, the LWE based solution can always serve as an initializer of the DTW based optimization, thus reducing significantly the computational requirements of the DTW as the search performed by the DTW algorithm can now be confined to the small region of deformations provided by the LWE estimate.

4.3 Analysis of the High SNR Case

In this section we analyze the proposed method assuming $w_k(y) = y^k$, when it is assumed that the signal to noise ratio is high. Since $h(x) = g(\phi(x)) + \eta(x)$ we have under the high SNR assumption that the contribution of high noise powers can be neglected, *i.e.*,

$$h^k(x) \approx g^k(\phi(x)) + k\ \eta(x) g^{k-1}(\phi(x)) \tag{26}$$

Hence, the error term in (16) is approximated under the high SNR assumption by

$$\bar{\epsilon}_k^g \;=\; k\int_{-\infty}^{\infty} \eta(x) g^{k-1}(\phi(x))dx \quad k=1,\ldots,N \tag{27}$$

Clearly, $E[\bar{\epsilon}_k^g]=0,\ k=1,\ldots,N$. We next evaluate the error covariances of the system, under the high SNR assumption. Let $\tilde{\boldsymbol{\epsilon}}^g=[\tilde{\epsilon}_1^g,\ldots,\tilde{\epsilon}_N^g]^T$, and $\mathbf{\Gamma}=E[\tilde{\boldsymbol{\epsilon}}^g(\tilde{\boldsymbol{\epsilon}}^g)^T]$. Thus, the (k,l) element of $\mathbf{\Gamma}$ is given by

$$\begin{aligned}
\mathbf{\Gamma}_{k,l} &= klE[\int_{-\infty}^{\infty} \eta(x)g^{k-1}(\phi(x))dx \int_{-\infty}^{\infty} \eta(y)g^{l-1}(\phi(y))dy] \\
&= kl\sigma^2 \int_{-\infty}^{\infty} g^{k+l-2}(\phi(x))dx \\
&= kl\sigma^2 \int_{-\infty}^{\infty} (\phi^{-1}(z))' g(z)^{k+l-2}dz \\
&= kl\sigma^2 \sum_{i=1}^{P} a_i \int_{-\infty}^{\infty} e_i(z)g(z)^{k+l-2}dz
\end{aligned} \tag{28}$$

Rewriting (28) in matrix form we have

$$\mathbf{\Gamma}=\sigma^2\sum_i a_i\mathbf{u}_i \tag{29}$$

where

$$\mathbf{u}_i=\begin{pmatrix} \int_{-\infty}^{\infty} e_i(z)dz & \ldots & N\int_{-\infty}^{\infty} e_i(z)g(z)^{N-1}dz \\ \vdots & \ddots & \vdots \\ N\int_{-\infty}^{\infty} e_i(z)dz & \ldots & N^2\int_{-\infty}^{\infty} e_i(z)g(z)^{2N-2}dz \end{pmatrix} \tag{30}$$

Assuming further that the observation noise $n(x)$ is Gaussian, we have under the high SNR assumption that $\bar{\boldsymbol{\epsilon}}^g$ is a zero mean Gaussian random vector with covariance matrix $\mathbf{\Gamma}$ given in (29). Hence the log-likelihood function $\log p(\mathbf{h};\mathbf{a})$ of the observation vector $\mathbf{h}$ is easily obtained. The MLE of the deformation model parameters can then be found by maximizing $\log p(\mathbf{h};\mathbf{a})$ with respect to the model parameters $\mathbf{a}$.

5 Conclusions

We introduced a novel methodology for geometric deformation estimation of a known object, where the deformation belongs to a known family of deformations. As a result of the action of the set of all possible deformations in the family, the set of different realizations of each object is generally a manifold in the space of observations. We showed that in cases where the family of possible deformations the object may undergo, admits a finite dimensional representation, there is a *nonlinear* mapping from the space of observations to a low dimensional linear space. We have rigorously analyzed the structure of the nonlinear operators achieving this mapping, and showed their decomposition into particular and homogeneous solutions. As a result of the derived mapping, the manifold corresponding to each object is mapped to a linear subspace with the same dimension as that of the manifold. In this setting, the problem of estimating the parametric model of the warping function is solved by a *linear* system of equations in the low dimensional Euclidian space.

References

[1] U. Grenander, *General Pattern Theory*, Oxford University Press, 1993.

[2] R. Hagege and J. M. Francos, "Parametric Estimation of Affine Transformations: An Exact Linear Solution," *Jou. of Math. Imaging and Vision.*, pp. 1-16, 2010.

[3] L. Younes, "Tutorial: Deformation, Warping and Object Comparison," Proc. ECCV'2000.

[4] A. Trouve and L. Younes, "On a Class of Diffeomorphic Matching Problems in One Dimension," *SIAM Jou. Cont. Opt.*, pp. 1112 - 1135, 2000.

[5] L. Younes,"Computable Elastic Distances Between Shapes," *SIAM Jou. Appl. Math.* pp. 565-586, 1998.

[6] M. Piccioni, S. Scarlatti and A. Trouve,"A Variational Problem Arising from Speech Recognition," *SIAM Jou. Appl. Math.* pp. 753-771, 1998.

[7] Y. Amit, *2-D Object Detection and Recognition*, MIT Press, 2002.

[8] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Trans. Acoust., Speech and Signal Process.*, pp. 43-49, 1978.

[9] C. Myers and L. Rabiner, "A Level Building Dynamic Time Warping Algorithm for Connected Word Recognition," *IEEE Trans. Acoust., Speech and Signal Process.*, pp. 284-297, 1981.

[10] L. Rabiner and B. Juang, *Fundamentals of Speech Recognition*, Prentice Hall, 1993.

[11] T. Gasser and K. Wang, "Alignment of Curves by Dynamic Time Warping," *Annal. Stat.*, pp. 1251-1276, 1997.

[12] D. Berndtand J. Clifford, "Using Dynamic Time Warping to Find Patterns in Time Series," *Proc. AAAI-94 Workshop on Knowledge Discovery in Databases* , pp. 229248, 1994.

[13] E. Keogh and C. A. Ratanamahatana, "Exact Indexing of Dynamic Time Warping," *Knowledge and Information Systems*, pp. 358-386, 2005.

[14] J. Kybic, P. Thevenaz, A. Nirkko, and M. Unser, "Unwarping of Unidirectionally Distorted EPI Images," *IEEE Trans. Med. Imaging*, pp. 80-93, 2000.

[15] C. Ekanadham, D. Tranchina, and E. P. Simoncelli, "Recovery of Sparse Translation-Invariant Signals With Continuous Basis Pursuit," *IEEE Trans. Signal Process.*, pp. 4735-4744, 2011

[16] David G. Luenberger, *Linear & Nonlinear Programming*, 2^{nd} ed., Addison-Wesley, 1984.

[17] L. Lucchese, "A Frequency Domain Technique Based on Energy Radial Projections for Robust Estimation of Global 2D Affine Transformations," *Computer Vision and Image Understanding,* vol. 81, pp. 72-116, Feb. 2001.

[18] Y. Keller, A. Averbuch and M. Israeli, "Pseudopolar-based estimation of large translations, rotations, and scalings in images," *IEEE Trans. Image Process.* vol. 14, pp. 12-22, 2005.

[19] H. Foroosh, J. Zerubia, M. Berthod, "Extension of Phase Correlation to Subpixel Registration," *IEEE Trans. Image Process.* vol. 11, pp. 188 - 200, 2002.

[20] J. M. Francos and R. Hagege, *Parametric Estimation of Multi-Dimensional Homeomorphic Transformations*, United States Patent 7,406,197.

[21] H. L. Royden, *Real Analysis*, Prentice-Hall, 1988.