2010 American Control Conference
Marriott Waterfront, Baltimore, MD, USA
June 30-July 02, 2010

Time and Output Warping of Control Systems: Comparing and Imitating Motions

Peter Kingston and Magnus Egerstedt
kingston@gatech.edu, magnus@ece.gatech.edu
Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, GA 30332, USA

Abstract— **How can one system "mimic" a motion generated by another? To address this question we introduce an optimal tracking problem which additionally optimizes over functions which deform or "warp" the time axis and the output space. Parametric and nonparametric versions of the time-warped tracking problem are introduced and reduced to standard Bolza problems. The output warping problem is treated for piecewise affine output warping functions.**

I. Introduction

How can a system be controlled to execute a motion that is "like" a desired motion? Although this may seem to be a rather ill-posed question, it is exactly what robots are asked to do in some areas of human-robot interactions, particularly in Programming by Demonstration (for example, [1]). Our particular goal is for systems (e.g. robotic marionettes) to produce motions which are recognizably and aesthetically "the same" as recorded human motions, even when the systems which are being controlled are unable to track those reference trajectories in a traditional (e.g., L_2) sense.

In [2], this problem was partially addressed via supervised learning, and very little structure was assumed; the concept of "similarity" was learned from scratch. In this paper, we instead fix a particular definition for similarity a priori: Namely, we treat mimicking as a special optimal tracking problem with additional degrees of freedom. Specifically, we allow for both the time axis and the output space to be deformed elastically, and optimize over the homeomorphisms defining these warpings.

The first portion of this paper is concerned with the deformation of the time axis, and is motivated by the recognition that the controlled system either simply may not be able to move as quickly as a human, or may have natural modes of oscillation (imagine a marionette swinging from its strings) which can be exploited by interpreting time liberally. It differs from existing work in time warping (e.g. [3], [4], [5], [6]) in that *the control problem and the "warping" problems are inextricably coupled*; and from work on time reparametrization (e.g., [7]) and path following (e.g., [8],[9],[10]) in that we are interested in approximate tracking for finite-duration moves, rather than in exact or asymptotic tracking.

Of the previous work, that which is most similar in approach to our own appears in [11], the primary concern of which is the comparison of motions for computer vision purposes. Key differences include that in [11] time warping is applied to the input to the systems rather than the output as in our case, and in [11] the control effort required to effect the motion is not penalized in the optimization problem. These differences follow naturally from the different goals of the two papers.

The second portion of this paper is concerned with "warping" the output space. For motivation, consider a marionette which cannot lift its arm above shoulder level but which is asked to mimic a human who waves above her head; although this motion cannot be tracked well in e.g. an L_2 sense, the puppet can nevertheless perform a recognizably equivalent motion lower in space and scaled. Output warping is inspired in large part by work in image and video processing (e.g. [12], [13], [14], [15], [16]), but as we will be "warping" not images but output trajectories, there are however very few technical similarities to this work (besides the use of simplicial complexes to define warping functions in [16]).

II. Time Warping

A *time warping function* is a continuously-differentiable function $w : [0,T] \to \mathbb{R}$, $T \in \mathbb{R}_+$, $w(0) = 0$ which

This work was supported by the U.S. National Science Foundation through grant number 0820004. We would also like to acknowledge Amy LaViers' help in producing the data used in the dancing puppet experiment.

978-1-4244-7427-1/10/$26.00 ©2010 AACC

is strictly increasing. Such a function is a bijection, and moreover a homeomorphism. We will denote the set of all time warping functions (those satisfying the above conditions) by Ω, and the set of derivatives of time warping functions by Ω'; i.e., Ω' is the set of continuous non-negative functions which are zero at at most a finite number of points.

Given a reference signal $r : [0,T] \to \mathbb{R}^p$, $T \in \mathbb{R}_+$, and another signal $y : \mathbb{R} \to \mathbb{R}^p$, the usual goal of time warping is to find a function $w \in \Omega$ that minimizes the functional $J : \Omega \to \mathbb{R}_+ \cup \{0\}$ defined,

$$J(w) = \int_0^T ||r(\tau) - (y \circ w)(\tau)||_Q^2 d\tau \quad \forall w \in \Omega \quad (1)$$

in some norm $|| \cdot ||_Q \triangleq \sqrt{(\cdot)^T Q(\cdot)}$.[1]

However, we can also use the time warping idea to formulate a "time-warped" output-tracking problem as well, and it is this which will be the focus of the first half of this paper. We will do this in two ways: First, we will solve the problem when any time warping function is allowed; we call this the *nonparametric* time warping problem. We follow this with a look at techniques which fix a particular parametrized form for the time warping function – we address linear functions and, more generally, polynomials – and which optimize over those parameters.

A. Tracking with Nonparametric Time Warping

Given a (possibly) nonlinear system of the form,

$$\begin{aligned} \frac{dx_t}{dt}(t) &= f(x_t(t), u_t(t), t) \\ y_t(t) &= h(x_t(t)) \end{aligned} \quad (2)$$

(with $x_t(t) \in \mathbb{R}^n$, $u_t(t) \in \mathbb{R}^m$, $y_t(t) \in \mathbb{R}^p$, and compatible dimensions for the domain and codomain of f and h)[2] we consider the optimal control problem of minimizing

$$J_t(u_t, w) = \left[\int_0^T ||y_t(w(\tau)) - r(\tau)||_Q^2 d\tau + \int_0^{w(T)} ||u_t(t)||_{R_u}^2 dt + \int_0^T R_v (w'(\tau) - 1)^2 d\tau \right]$$

where $r : [0,T] \to \mathbb{R}^p$ is a reference signal. The first term penalizes tracking error, but differs from the usual L_2 tracking problem by the introduction of the time warping function $w : [0,T] \to [0,\infty)$, and also in that we integrate tracking error over *reference time* (τ time) instead of *system time* (t time). The second term penalizes control effort as usual. The third term penalizes large deviations of $w'(\tau)$ from one, both to regularize the problem and to capture the intuition that signals which must be "warped" by a great deal are more dissimilar than those which do not need to be warped as much.

Augmenting the state with the time $t = w(\tau)$, defining $v \triangleq w'$ as the derivative of w, $x \triangleq x_t \circ w$, $u \triangleq u_t \circ w$, and $y \triangleq y_t \circ w$, this problem can be restated as the following standard (Bolza) optimal control problem: Given the system,

$$\begin{aligned} \frac{d}{d\tau} \begin{bmatrix} x \\ t \end{bmatrix} (\tau) &= \begin{bmatrix} v(\tau) f(x(\tau), u(\tau), t(\tau)) \\ v(\tau) \end{bmatrix} \\ y(\tau) &= h(x(\tau)) \end{aligned} \quad (3)$$

with known initial conditions $(x,t)(0) = (x_0, 0)$, minimize the cost functional

$$\begin{aligned} J &: L^2([0,T], \mathbb{R}^m) \times \Omega' \to \mathbb{R} \\ J(u,v) &= J_{\text{track}}(u,v) + J_{\text{timewarp}}(v) = \\ &\int_0^T \left[||y(\tau) - r(\tau)||_Q^2 + v(\tau) ||u(\tau)||_{R_u}^2 \right. \\ &\left. + R_v (v(\tau) - 1)^2 \right] d\tau \ . \end{aligned} \quad (4)$$

over the functions u and v (these functions can be viewed as control inputs to the system).

The small but important insight here is that time warping can be viewed as *modifying the dynamics of the system* in an appropriate time coordinate. The example that follows clarifies this point.

Example 2.1: Consider the underdamped simple harmonic oscillator described by transfer function $h(s) = 1/(s^2 + 2\zeta\omega_0 s + \omega_0^2)$ (with $\zeta \in (0,1) \subset \mathbb{R}$, $\omega_0 \in \mathbb{R}_+$), and compatible state-space realization (A_t, B_t, C_t). Next suppose that we would like to solve the infinite-time problem (a modified version of (4)),

$$\min_{u,w} \lim_{T \to \infty} \int_0^T \left[\frac{1}{T} ||(y_t \circ w)(\tau) - r(\tau)||_Q^2 + \frac{1}{w(T)} w'(\tau) ||(u_t \circ w)(\tau)||_{R_u}^2 \right] d\tau \quad (5)$$

where the reference signal to be tracked is the sinusoid,

$$r(t) = \cos(\omega_r t) \ \forall t \in [0,\infty) \ . \quad (6)$$

with $\omega_r \in \mathbb{R}_+$. Moreover for clarity of exposition we will limit our attention to time warping functions of the form $w(\tau) \equiv \xi\tau$ for some $\xi \in \mathbb{R}_+$ (Such *parametric time*

[1] In (1) and elsewhere, we denote $|| \cdot ||_M^2 \triangleq (\cdot)^T M(\cdot)$ and assume $M = M^T \succ 0$, for whichever matrix M is used in the subscript.

[2] Note that the subscript t is simply part of the function names x_t, y_t, etc, and is used to distinguish these functions from others to be introduced later.

warping functions are discussed in more detail in section II-B).

We note that the presence of frequencies other than ω_r in $u_t \circ w$ (and hence in $y_t \circ w$) increases both terms of (5),[3] so $u_t \circ w$ and $y_t \circ w$ must approach sinusoids with angular frequency ω_r as $t \to \infty$; without loss of generality[4] we will assume that they are in fact sinusoids. Also observing that the phase of $u_t \circ w$ has no effect on the second term of (5), it must be that $y \circ w = a\cos(\omega_r t)$ for some $a \in \mathbb{R}_+$, and

$$u_t(t) = \text{Re}\left(\frac{a}{h(i\omega_r/\xi)} e^{i\frac{\omega_r}{\xi}t}\right) \ . \tag{7}$$

It follows that the minimization problem (5) then reduces to,

$$\min_{a,\xi}\left[\frac{Q}{2}(a-1)^2 + \frac{R_u}{2}\left|\frac{a}{h(i\omega_r/\xi)}\right|^2\right] \ . \tag{8}$$

For any fixed a, this is minimized with respect to ξ when the magnitude of the transfer function $|h(i\omega_r/\xi)|$ is maximized. This occurs when $\xi = \xi^* \triangleq \omega_r/(\omega_0\sqrt{1-\zeta^2})$ – that is, when ξ is chosen so that the resonant frequency of the system with system matrix ξA_t coincides with the frequency of the reference signal.

The key point demonstrated by this example is that the steady-state effect of time warping is to scale the frequency axis (by the Fourier Dilation Theorem) so that passbands of the system coincide with concentrations of energy in the reference signal.

1) Optimality Conditions:

Theorem 2.1: The first order necessary optimality conditions for the minimization of (4) are

$$\begin{aligned} 2v(\tau)u^T(\tau)R_u + \\ v(\tau)\lambda^T(\tau)\frac{\partial f}{\partial u}(x(\tau),u(\tau),t(\tau)) &= \mathbf{0}^T \\ ||u(\tau)||^2_{R_u} + 2R_v(v(\tau-1) + \\ \lambda^T(\tau)f(x(\tau),u(\tau),t(\tau)) + \mu(\tau) &= 0 \end{aligned} \tag{9}$$

for all $\tau \in [0,T]$, where

$$\begin{aligned} -\frac{d\lambda}{d\tau}(\tau) &= 2h'(x)^T Q(h(x) - r(\tau)) + \\ &\quad v(\tau)\frac{\partial f}{\partial x}^T(x(\tau),u(\tau),t(\tau))\lambda(\tau) \\ -\frac{d\mu}{d\tau}(\tau) &= v(\tau)\frac{\partial f}{\partial t}^T(x(\tau),u(\tau),t(\tau))\lambda(\tau) \end{aligned} \tag{10}$$

with $(\lambda,\mu)(T) = 0$.

Proof : See [17]. ■

Note that when f is not time-varying, $\frac{\partial f}{\partial t} = 0$, which gives the simplification that $\mu(\tau) = 0 \ \forall \tau \in [0,T]$.

In fact, these equations (9) can be given the stronger interpretation of stating that the gradient of the functional (4) in the functional space of which (u,v) is an element must be zero; we will leverage this interpretation in the later section II-B.2 which describes an algorithm for computing the optimal (u,v).

B. Tracking with Parametric Time Warping

In some situations, we may be interested only in time warping functions with a particular parametric form. One example is linear time-warping functions, which are of special interest since they represent a uniform scaling of the time axis. Another motive for investigating time warping functions with given parametric forms is the discretization of the problem for numerical solution.

To express these ideas, we introduce a parameter vector ξ in some parameter set $\Xi \subset \mathbb{R}^q$, and a *parametrization function* $\phi_v : \Xi \to \Omega'$ which, given a parameter vector, returns the derivative of a time warping function. Then, we are in fact considering the problem,

$$\min_{\xi,u} J(u, \phi_v(\xi)) \ . \tag{11}$$

In the following subsections, we will first consider parametrization functions that return polynomial time warping functions – whose structure allow them to be treated nicely under the Bolza framework – (with linear time warping functions as a special case), and then give a more general view of the problem.

1) Polynomial Time Warping: Polynomial time warping functions are of the form,

$$w(\tau) = \phi_v(\xi)(\tau) = \sum_{i=1}^{N_v} \xi_i \tau^i \tag{12}$$

for some integer $N \geq 1$, and with discrete parameter vector $\xi = [\xi_1, \ldots, xi_N] \in \mathbb{R}^{N_v}$ (Note that the requirement

[3]These arguments can be made rigorous using Plancherel's identity for Fourier series and considering a sequence of values for T that are multiples of $\frac{2\pi}{\omega_r}$; we have omitted this lengthier development for the purposes of our informal discussion.

[4]This is explained in more detail in [17].

that $w(0) = 0$ implies that there is no constant term in the polynomial).

Theorem 2.2: The FONCs for the polynomial time warping problem are given by (9) and

$$\frac{\partial J}{\partial \xi}^T(\xi) = \mathrm{diag}(1, \tfrac{1}{2}, \tfrac{1}{3!}, ..., \tfrac{1}{N_v!})\nu(0) = 0 \qquad (13)$$

where (and dropping time arguments to x, λ, μ, u, ν),

$$\begin{aligned} -\frac{d\nu}{d\tau} &= \begin{bmatrix} 2R_v + \lambda^T f(x,u,t) + \mu \\ \nu_1 \\ \cdots \\ \nu_{N_v - 1} \end{bmatrix} \\ &= \mathbf{0} \end{aligned} \qquad (14)$$

for all $\tau \in [0, T]$, and with $\nu(T) = \mathbf{0}$.

Proof : See [17] ∎

2) The Chain Rule for Parametrization Functions: We note briefly that if the Fréchet derivatives of both the parametrization function ϕ_v and the cost J [defined in (4)] exist, then we may in fact apply the solution given in section II-A directly to the discretized problem (11) through a simple application of the chain rule. This hinges on the interpretation that the partial derivative of the Hamiltonian with respect to the control input (as a function of time) is the gradient of J with respect to the control input (projected onto the dynamical constraint); this is discussed more in [17]. Here, we will simply introduce a second parametrization function $\phi_u : \Sigma \to L^2([0,T], \mathbb{R}^m)$ for some $N_u \in \mathbb{N}$ and parameter space $\Sigma \subset \mathbb{R}^{N_u}$ which, given some finite-dimensional $\sigma \in \Sigma$, returns a control input function $u \in L^2([0,T], \mathbb{R}^m)$; this yields a problem which is now fully discretized both in u and v.

III. Output Warping

It is not just the dynamics of the mimicking system that may differ from those of the system that generated the reference motion, but also spatial constraints and scales – a problem evident even in the prototypical example of a large industrial robot arm asked to imitate a human operator. To treat this problem of spatial correspondence, we will assume that the reference signal r that we have been considering so far is in fact the composition of two functions: the "actual" reference signal $\bar{r} : [0,T] \to \mathbb{R}^p$, and an "output warping function" $s : \mathbb{R}^p \to \mathbb{R}^p$ of our choosing which transforms values of $\bar{r}$ before they are compared to those of the output signal y. In other words, $r = s \circ \bar{r}$. More precisely, an *output warping function* s : $\mathbb{R}^p \to \mathbb{R}^p$ is a continuous bijective map with continuous inverse (That is, s is a homeomorphism from $\mathbb{R}^p$ to $\mathbb{R}^p$). We will denote the set of all such functions by $\mathcal{S}$.

We will additionally assume that s has a particular parametric form. This is expressed by saying that s is returned by a parametrization function $\phi_s : \mathcal{C} \to \mathcal{S}$, where $\mathcal{C}$ is some finite-dimensional parameter space. Specifically, $\mathcal{C}$ is a space diffeomorphic to $\mathbb{R}^{N_s}$ for some $N_s \in \mathbb{N}$.

With these definitions, we can extend the original cost functional (4) to obtain the new cost functional to be minimized,

$$\begin{aligned} \bar{J} &: L^2([0,T], \mathbb{R}^m) \times \Omega' \times \mathcal{C} \to \mathbb{R} \\ \bar{J}(u,v,c) &= \bar{J}_{\text{track}}(u,v,c) + \bar{J}_{\text{timewarp}}(v) + \bar{J}_{\text{outwarp}}(c) \\ &= \int_0^T \left[||y(\tau) - (\phi_s(c) \circ \bar{r})(\tau)||_Q^2 + \right. \\ &\quad \left. ||u(\tau)||_{R_u}^2 + R_v(v(\tau) - 1)^2 \right] d\tau + \\ &\quad \bar{J}_{\text{outwarp}}(c) \end{aligned} \qquad (15)$$

where $\bar{J}_{\text{outwarp}}$ is some cost used to penalize large output warpings, regularize the problem, and in certain cases enforce constraints; its form will be determined by the choice of ϕ_s and is discussed in more detail later.

A. Piecewise Affine Output Warping

The essential idea of piecewise affine output warping will be that we divide the space $\mathcal{R}$ into some number of p-simplices, and use an affine warping function within each of these, chosen in such a way that the resulting piecewise function is continuous. In order to enforce that s remain a bijection, this will require both that the individual affine warping functions be full rank, and that the images of their domains remain disjoint.

To begin, let S (the "input simplices") be a finite pure simplicial p-complex covering $\mathcal{R}$, and R ("the output simplices") be another finite pure simplicial p-complex, which is isomorphic to S.[5] Basically, we will optimize over the positions of vertices in R, and use the induced simplicial map (which interpolates vertex positions barycentrically) as our output warping function.

In more detail: We denote the p-simplices contained in S by $S^1, \ldots, S^{|S|}$, and those contained in R by $R^1, \ldots, R^{|S|}$; that is, the p-simplices in S and R are indexed. We also denote the vertices (0-simplices) of S

[5] *Pure* means that the only simplices of dimension less than p are the faces of higher-dimensional simplices. *p-complex* means that the highest dimensional simplices are p-simplices. *Isomorphic* means that there is a bijection between elements of R and S that preserves topology.

and R by $V_1^R, V_2^R, \ldots$ and $V_1^S, V_2^S, \ldots$, respectively, and the vertices of a given simplex R^i by $R_1^i, \ldots, R_p^i$. Then, we define the output warping function by,

$$s(\bar{r}) = \sum_{i=1}^{p} R_i^{\pi_S(\bar{r})} \beta_i(\bar{r}, S^{\pi_S(\bar{r})}) \tag{16}$$

where $\pi_S(r) : \mathcal{R} \to \mathbb{N}$ is the function that, given a point $\bar{r} \in \mathcal{R}$, returns the index of the p-simplex in S containing $\bar{r}$, and $\beta : \mathcal{R} \times S \to \mathbb{R}^{p+1}$ is the function that, given a point $\bar{r} \in \mathcal{R}$ and a simplex $S^i \in S$, returns the barycentric coordinates of r in S^i if $r \in S^i$ and $\mathbf{0}$ otherwise. The map s is called the *simplicial map* induced by the vertex map from S to R.

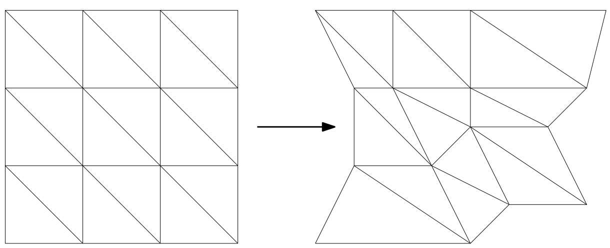

Fig. 1. Given the set S of input simplices, the output warping function s is determined by the positions of the vertices of the corresponding output simplices R. This example uses a Coxeter-Kuhn-Freudenthal tessellation of a regular grid of cubes in $\mathbb{R}^2$.

Defining for each simplicial complex $\mathcal{K} \in \{S, R\}$ a graph $\mathcal{G}^{\mathcal{K}}$ whose vertices are the 0-simplices in $\mathcal{K}$, and in which an edge exists between two vertices iff they are both contained within the same 1-simplex, then a cost which tends to maintain the bijectivity of s is given by,

$$\bar{J}_{\text{outwarp}}(c) = \frac{1}{2} \sum_{(V_i^R,\, V_j^R) \,\in\, \text{edges}(\mathcal{G}^R)} \left(||V_i^R - V_j^R||_K - ||V_i^S - V_j^S||_K \right)^2 . \tag{17}$$

The idea here is that $\mathcal{G}^R$ is a rigid graph, and that by maintaining edge distances we ensure that simplices can neither "collapse" nor "collide." If $\mathcal{G}^R$ is visualized as a network of springs, then (17) gives their overall potential energy.

The partial gradient of $\bar{J}_{\text{outwarp}}$ with respect to each V_i^R is then,

$$\sum_{V_j^R \,\in\, \mathcal{N}_{\mathcal{G}^R}(V_i^R)} \frac{||V_i^R - V_j^R||_K - ||V_i^S - V_j^S||_K}{||V_j^R - V_i^R||_K} K (V_j^R - V_i^R) \tag{18}$$

where $\mathcal{N}_{\mathcal{G}^R}(V_i^R)$ is the neighborhood of V_i^R in $\mathcal{G}^R$.

Admittedly, this cost does leave something to be desired, since simplices can collapse with finite energy. Nevertheless, we believe it is useful for its simplicity. One may wish to also apply (59) in [17] for each simplex in cases where (17) is not sufficient.

Now, define $c = \begin{bmatrix} (V_1^R)^T & \cdots & (|V_{\text{verts}(R)|}^R)^T \end{bmatrix}^T$. Letting $i_1(r), \ldots i_{p+1}(r)$ be the indexes into $\text{verts}(S)$ corresponding to the vertices of the simplex in S containing r, letting $\pi_S(r)$ be the simplex in S containing r, and defining the $p|\text{verts}(S)| \times p$ matrix

$$Z(r) = \begin{bmatrix} I\alpha_1 \\ \vdots \\ I\alpha_{|\text{verts}(S)|} \end{bmatrix} \tag{19}$$

where $\alpha_{i_1(r)} = \beta_1(r, \pi_S(r)), \ldots, \alpha_{i_{p+1}(r)} = \beta_{p+1}(r, \pi_S(r))$ and $\alpha_i = 0 \ \forall i \notin \{i_1(r), \ldots, i_{p+1}(r)\}$, then the the partial gradient of (15) without the last term $\bar{J}_{\text{outwarp}}$ is given by,

$$\nabla_c(\bar{J} - \bar{J}_{\text{outwarp}})(R) = \\ -2 \int_0^T Z(r(\tau)) Q \left[y(\tau) - (\phi_s(c) \circ r)(\tau) \right] d\tau \ . \tag{20}$$

Hence the partial gradient of $\bar{J}$ with respect to c is simply the sum of (18) and (20).

We apply piecewise affine output warping together with linear time warping in a short example below.

Example 3.1: Suppose we would like the state of an autonomous Van der Pol oscillator to track that of a damped pendulum driven by a fixed-frequency sinusoid, allowing for linear time warping and piecewise affine output warping. That is, the system is given by,

$$\begin{aligned} \frac{d}{dt} \begin{bmatrix} x_{t,1} \\ x_{t,2} \end{bmatrix} (t) &= \begin{bmatrix} x_{t,2}(t) \\ \zeta_{\text{vp}}(1 - x_{t,1}^2(t)) \end{bmatrix} \\ y_t(t) &= x_t(t) \end{aligned}$$

(with in our case $\zeta_{\text{vp}} = 0.9$) and the reference signal r is the solution to (dropping time arguments to r_1, r_2),

$$\frac{d}{d\tau} \begin{bmatrix} r_1 \\ r_2 \end{bmatrix} = \begin{bmatrix} r_2 \\ \sin(\tau) - \omega_0^2 \sin(r_1) - \zeta_{\text{pend}} r_2 \end{bmatrix}$$

with in our case $\omega_0 = 1$, $\zeta_{\text{pend}} = 0.5$, $x(0) = r(0) = [0.1, 0.1]^T$, and $\tau \in [0, T] = [0, 10]$. We will use essentially the same costs introduced earlier, but with some specially-chosen constants (described in more detail in [17]). Then, performing gradient descent using the gradient given by (18) and (20), we obtain the results shown in figures 2 and 3.

IV. Conclusions

In order to allow one system to mimic a reference signal, we have introduced several versions of a modified output

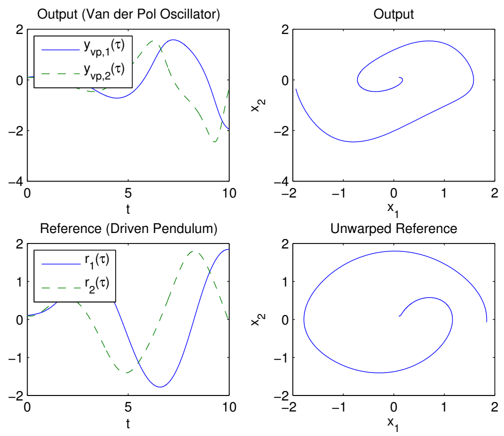

Fig. 2. Van der Pol oscillator vs. driven pendulum, before warping. We wish to scale the time axis of the output (top left) and the output space of the reference (bottom right) to align the two signals.

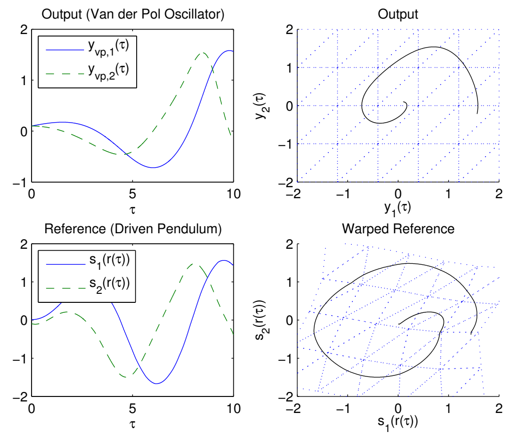

Fig. 3. Van der Pol oscillator vs. driven pendulum, after warping. Time warping matches the first part of the Van der Pol oscillator's transient to that of the pendulum (left top, bottom), and output warping rotates and deforms the reference output space to better match the output (right top, bottom).

tracking problem that also includes time and output warping functions as decision variables, and given FONCs for all of these. The basic motivation has been that this captures a measure of qualitative similarity which the usual error metrics used in tracking problems (like the generalized L^2 metric) do not.

The chief limitations of this approach are related to computational tractability, and are common to many problems in numerical optimal control: The gradients, being the solutions to ordinary differential equations, are fairly expensive to compute; and the nonconvexity of the problem means that only local optima are guaranteed. Nevertheless, it is possible to compute local optima which do give results that achieve our ultimate goal of qualitative similarity.

References

[1] *IEEE Transaction on Systems, Man, and Cybernetics, Special Issue on Robot Learning by Observation, Demonstration, and Imitation, Part B*, vol. 37, Apr 2007.

[2] Peter Kingston and Magnus Egerstedt, "Comparing apples and oranges through partial orders: An empirical approach," in *American Control Conference*, 2009.

[3] H. Sakoe and S. Chiba, "A dynamic programming approach to continuous speech recognition," in *Proc. Int'l. Cong. Acoust.*, 1971.

[4] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," in *IEEE Transactions on Acoustics, Speech and Signal Processing*, Feb 1978.

[5] A.E. Rosenberg C. Myers, L.R. Rabiner, "Performance tradeoffs in dynamic time warping algorithms for isolated word recognition," in *IEEE Transactions on Acoustics, Speech and Signal Processing*, Dec 1980, vol. ASSP-6, pp. 623–635.

[6] A. Waibel and Yegnanarayana, "Comparative study of nonlinear time warping techniques in isolated word speech recognition," in *IEEE Transactions on Acoustics, Speech and Signal Processing*, Feb 1978.

[7] George Pappas, "Avoiding saturation by trajectory reparameterization," in *Proceedings of the 35th IEEE Decision and Control*, Dec 1996, vol. 1, pp. 76–81.

[8] J. Hauser and R. Hindman, "Aggressive flight maneuvers," in *Decision and Control, 1997., Proceedings of the 36th IEEE Conference on*, 1997, vol. 5, pp. 4186–4191.

[9] R. Skjetne, T. I. Fossen, and P. V. Kokotovic, "Adaptive maneuvering, with experiments, for a model ship in a marine control laboratory," in *Automatica*, 2005, vol. 41, pp. 289–298.

[10] A.P. Aguiar and J.P. Hespanha, "Trajectory-tracking and path-following of underactuated autonomous vehicles with parametric modeling uncertainty," in *Automatic Control, IEEE Transactions on*, vol. 52.

[11] Michalis Raptis, Matteo Bustreo, and Stefano Soatto, "Time warping under dynamic constraints," in *Eleventh IEEE International Conference on Computer Vision, Workshop on Dynamical Vision*, 2007.

[12] Daniel Keysers and Walter Unger, "Elastic image matching is NP-complete," *Pattern Recognition Letters*, vol. 24, pp. 445–453, 2003.

[13] Jan Kybic, *Elastic Image Registration using Parametric Deformation Models*, Phd in biomedical image processing, Ecole Polytechnique Fédérale de Lausanne, July 2001.

[14] Hansheng Lei and V. Govindaraju, "Direct image matching by dynamic warping," in *Computer Vision and Pattern Recognition Workshop, 2004. CVPRW '04. Conference on*, 2004, p. 76.

[15] Zdenek Prochazka, Takayuki Ito, and Toshio Okamoto, "Image correspondences by warping functions and its applications," in *Systems and Computers in Japan*, 2002, vol. 33, pp. 22–30.

[16] Ergun Akleman, "Making caricatures with morphing," in *SIGGRAPH: ACM Special Interest Group on Computer Graphics and Interactive Techniques*, 1997, p. 145.

[17] Peter Kingston and Magnus Egerstedt, "Time and output warping of control systems: Comparing and imitating motions," Tech. Rep., Georgia Institute of Technology, `http://www.prism.gatech.edu/~pkingston3/papers/time_out_warp.pdf`.