Optimal Model Order Reduction For Maximal Real Part Norms

A. Megretski*

October 28, 2002

Abstract

An observation is made that a polynomial time algorithm exists for the problem of optimal model order reduction of finite order systems in the case when the approximation error measure to be minimized is defined as maximum of the real part of the model mismatch transfer function over a certain set of frequencies. Applications to H-Infinity model reduction and comparison to the classical Hankel model reduction are discussed.

I am grateful for the opportunity to present this paper to honor Professor Anders Lindquist who was a profound positive influence in my reserach career.

1 Introduction

This paper deals with problems of model order reduction for linear time-invariant (LTI) systems. Reduced order transfer functions are frequently used in modeling, design, and computer simulation of complex engineering systems. Despite significant research efforts, several fundamental questions concerning LTI model reduction remain unsolved.

A mathematical formulation of a model reduction problem can be given in terms of finding a stable transfer function $\hat{G}$ (the *reduced model*) of order less than k such that $\|G - \hat{G}\|$ (the *approximation error measure* quantifying the

*ameg@mit.edu. This work was supported by NSF, AFOSR, and DARPA

size of the *model mismatch* $\Delta = G - \hat{G}$) is minimal. Here G is a given stable LTI system (the *non-reduced model*), and $\|\cdot\|$ is a given norm (or sometimes a semi-norm) on the vector space of stable transfer functions. While G is not rational in many applications, it is usually reasonable to assume that a high order high quality finite order approximation G_0 of G is available, and therefore the optimal model reduction problem is formulated as

$$\|G_0 - \hat{G}\| \to \min, \ \ (\text{order}(\hat{G}) < k), \tag{1}$$

entirely in terms of finite order transfer functions.

Based on the wisdom of modern robust control, the most desirable norms $\|\cdot\|$ to be used in optimal model reduction are the so-called *weighted H-Infinity norms*

$$\|\Delta\| = \|W\Delta\|_\infty, \tag{2}$$

where W is a given rational transfer function. However, to the author's knowledge, no polynomial time algorithms are known for solving (1) in the case of a non-zero weighted H-Infinity norm $\|\cdot\|$ (even when $W \equiv 1$). (It is also not known whether the problem is NP-hard or not.)

The case when

$$\|\Delta\| = \|\Delta\|_h = \min_{\delta} \|\Delta + \bar{\delta}\|_\infty \tag{3}$$

(δ ranges over the set of stable transfer functions) is the so-called *Hankel norm* appears to be the only situation in which a polynomial time algorithm for solving (1) is commonly known. The theory of *Hankel model reduction* is the main concentration of rigorous results on model reduction. Since

$$\|\Delta\|_h \leq \|\Delta\|_\infty \tag{4}$$

for every stable transfer function Δ, solving the Hankel model reduction problem provides a *lower bound* in the (unweighted) H-Infinity model reduction problem. In addition, there is some evidence, both formal and experimental, that, for "reasonable" systems, the H-Infinity model matching error delivered by the Hankel optimal reduced model is not much larger than the optimal Hankel model reduction error.

The positive statements concerning Hankel model reduction and its relation to H-Infinity model reduction do not cover the case of weighted Hankel norms

$$\|\Delta\| = \|D\|_{h|W} = \min_{\delta} \|W(\Delta + \bar{\delta})\|_\infty.$$

The theory also does not extend to the case of G being defined by a finite number of frequency samples. The main point of this paper is that an alternative class of system norms $\|\cdot\|$, called *weighted maximal real part* norms, yields most of the good properties known of the Hankel model reduction, while providing the extra benefits of using sampled data and freqency weighted modeling error measures. For the weighted maximal real part model reduction, the paper provides a polynomial time optimization algorithm, and states a number of results concerning its relation to H-Infinity and Hankel model reduction. Outcomes of some numerical experiments are also presented.

2 Maximal Real Part Model Reduction

For convenience, model reduction of *discrete-time systems* will be considered. Thus, a *stable transfer function* will be defined as a continuous function $f: \mathbf{T} \to \mathbf{C}$ for which the Fourier coefficients

$$\hat{f}[n] = \int_{-\pi}^{\pi} f(e^{jt}) e^{-jnt} dt$$

are all real and satisfy the condition

$$\hat{f}[n] = 0 \quad \forall\ n > 0.$$

Here $\mathbf{C}$ is the set of all complex numbers, and

$$\mathbf{T} = \{z \in \mathbf{C} : \ |z| = 1\}$$

is the unit circle centered at $z = 0$. The set of all stable transfer functions will be denoted by $\mathbf{A}$. The set of all rational stable transfer functions of order less than k will be denoted by $\mathbf{A}_k$.

2.1 The Unsampled Setup

The unsampled version of the *maximal real part model reduction problem* is defined as the task of finding $\hat{G} \in \mathbf{A}_k$ which minimizes $\|G_0 - \hat{G}\|_{r|W}$, where $W = |H|^2$,

$$\|\Delta\|_{r|W} = \|W\mathrm{Re}(\Delta)\|_\infty,$$

$G_0, H \in \mathbf{A}$ are given rational transfer functions, and

$$\|f\|_\infty = \max_{z\in\mathbf{T}} |f(z)|$$

for every continuous function $f:\ \mathbf{T} \to \mathbf{C}$.

It is easy to see that

$$0.5\|\Delta\|_{h|W} \leq \|\Delta\|_{r|W} \leq \|W\Delta\|_\infty \quad \forall\ \Delta \in \mathbf{A},$$

where the first inequality takes place because of

$$\|D\|_{h|W} = \min_{\delta} \|W(\Delta + \bar{\delta})\|_\infty \leq \|W(\Delta + \bar{\Delta})\|_\infty = 2\|\Delta\|_{r|W}.$$

Therefore $\|\Delta\|_{r|W}$ is a norm which relaxes the corresponding weighted H-Infinity norm and is stronger than the associated weighted Hankel norm.

It will be shown later in this section that the unsampled maximal real part optimal model reduction problem can be reduced to a semidefinite program of the size which grows linearly with k and the orders of G_0 and H. This extends significantly the set of model reduction setttings for which a solution can be found efficiently.

2.2 The Sampled Setup

Applications of model reduction often deal with the situation in which the order of G_0 (hundreds of thousands) is so large that it becomes not practical to handle the exact state-space or transfer funcion representations of G_0. In such cases it may be useful to work with sampled frequency domain values of G.

The *sampled* version of the maximal real part model reduction problem is defined as the task of finding $\hat{G} \in \mathbf{A}_k$ which minimizes $\|G_0 - \hat{G}\|_{s|V}$, where $V = \{(W_i, t_i)\}_{i=1}^N$ is a given sequence of pairs of real numbers $t_i \in [0, \pi]$, $W_i > 0$,

$$\|\Delta\|_{s|V} = \max_{1\leq i\leq N} W_i|\mathrm{Re}(\Delta(e^{jt_i}))|,$$

and G_0 is a stable transfer function which is given (incompletely) by its samples $G_{0,i} = G(e^{jt_i})$.

Note that $\|G_0 - \hat{G}\|_{s|V}$ is completely determined by $\hat{G}$, $V = \{(W_i, t_i)\}_{i=1}^N$, and the samples $G_{0,i} = G(e^{jt_i})$. Practical use of the sampled setup usually relies on an assumption that $G(e^{jt})$ does not vary too much between the sample points t_i.

It is possible to propose various modifications of the sampled modeling error measure $\|\cdot\|_{s|W}$. For example, assuming that $0 = t_0 \le t_1 \le t_2 \le \cdots \le t_N \le t_{N+1}$, one can use the mixed cost

$$J(G_0, \hat{G}) = \max_{1\le i\le N} W_i \max_{t\in[t_{i-1},t_{i+1}]} |G(e^{jt_i}) - \hat{G}(e^{jt})|.$$

The main point, however, is the possibility to reduce the model reduction setup to an equivalent semidefinite program, to be shown in the next subsection.

2.3 The Convex Parameterization

By a *trigonometric polynomial* f of degree $m = \deg(f)$ we mean a function $f:\ \mathbf{T} \to \mathbf{R}$ of the form

$$g(e^{jt}) = \sum_{k=0}^{m} g_k \cos(kt),$$

where $g_k \in \mathbf{R}$ and $g_m \neq 0$.

The following simple observation is a key to the convexification of maximal real part optimal model reduction problems.

Lemma 1 *For every $f \in \mathbf{A}_m$ there exist trigonometric polynomials a, b such that*

$$\deg(a) < m,\ \deg(b) < m,\quad a(z) > 0\ \ \forall\ z \in \mathbf{T}, \tag{5}$$

and

$$\mathrm{Re}(f(z)) = \frac{b(z)}{a(z)}\ \ \forall\ z \in \mathbf{T}. \tag{6}$$

Conversely, for every pair (a, b) of trigonometric polynomials satisfying (5) there exists $f \in \mathbf{A}_m$ such that (6) holds.

Proof. If $f \in \mathbf{A}_m$ then

$$f(z) = \frac{p(z)}{q(z)}\ \ \forall\ z \in \mathbf{T},$$

where p, q are polynomials of degree less than m, with real coefficients, and $q(z) \neq 0$ for $|z| \geq 1$. Hence (6) holds for a, b defined by

$$a(z) = q(z)q(1/z), \quad b(z) = \frac{1}{2}(q(z)p(1/z) + q(1/z)p(z)) \quad (z \neq 0), \tag{7}$$

or, equivalently, by

$$a(z) = |q(z)|^2, \quad b(z) = \mathrm{Re}(p(z)q(\bar{z})) \quad (|z| = 1).$$

It is easy to see that a, b defined by (7) are trigonometric polinomials satisfying (5).

Conversely, let a, b be trigonometric polinomials satisfying (5). Let $r = \deg(a)$. Then $h(z) = z^r a(z)$ is an ordinary polynomial of degree $2r$. Since $a(z) > 0$ for all $z \in \mathbf{T}$, $h(z)$ has no zeros in $\mathbf{T} \cup \{0\}$. Since $a(z) = a(1/z)$, all zeros of $h(z)$ can be arranged in pairs $(z_i, 1/z_i)$, where $|z_i| < 1$, $i = 1, \dots, r$, i.e. $h(z) = q_0 z^r q_r(z) q_r(1/z)$ where q_0 is a constant, and

$$q_r(z) = (z - z_1)(z - z_2) \cdot \dots \cdot (z - z_r)$$

is a polynomial with no zeros in the region $|z| \geq 1$. Moreover, since h has real coefficients, the non-real zeros of h come in conjugated pairs, and hence q_r has real coefficients as well, and $q_0 \in \mathbf{R}$. Equivalently, we have

$$a(z) = q_0 |q_r(z)|^2 \quad \forall \; z \in \mathbf{T}.$$

Since $a(z) > 0$ for all $z \in \mathbf{T}$, we have $q_0 > 0$. Let

$$q(z) = q_0^{1/2} z^{m-r-1} q_r(z).$$

Then q is a polinomial of degree $m - 1$ with no zeros in the region $|z| \geq 1$, and $|q(z)|^2 = a(z)$ for all $z \in \mathbf{T}$.

It is left to show that a polynomial $p(z)$ of degree less than m with real coefficients can be found such that

$$2b(z) = p(z)q(1/z) + p(1/z)q(z).$$

Indeed, the set V of all real polynomials p of degree less than m forms an m-dimensional real vector space. The map

$$M_q : \quad p(z) \;\mapsto\; p(z)q(1/z) + p(1/z)q(z)$$

is a linear transformation from V into the m-dimensional real vector space of trigonometric polinomials of degree less than m. Moreover, $\ker M_q = \{0\}$, since

$$p(z)q(1/z) = -p(1/z)q(z)$$

would imply that p and q have same set of zeros (here we use the fact that all zeros of q are in the open unit disc $|z| < 1$), hence $p(z) = cq(z)$ and $c = 0$, i.e. $p = 0$. Therefore M_q is a bijection. ■

Using Lemma 1 it is easy to convexify the maximal real part optimal model reduction problems. In particular, the unsampled version originally has the form

$$y \to \min \text{ subject to } |H(z)|^2|(Re)(G_0(z) - \hat{G}(z))| \leq y \ \forall \ z \in \mathbf{T}, \ \hat{G} \in \mathbf{A}_m.$$

Replacing $\mathrm{Re}(\hat{G}(z))$ by $b(z)/a(z)$ where a, b are the trigonometric polynomials from Lemma 1, we obtain an equivalent formulation

$$y \to \min \text{ subject to } |H(z)|^2|a(z)Re(G_0(z)) - b(z)| \leq ya(z) \ \forall \ z \in \mathbf{T},$$

where the decision parameters a, b are constrained by (5). Since, for a given y, the constraints imposed on a, b are convex, the optimization problem is quasi-convex, and can be solved by combining a binary search over y with a convex feasibility optimization over a, b for a fixed y. For practical implementation, using an interior point cutting plane algorithm with a feasibility oracle utilizing the Kalman-Yakubovich-Popov lemma is advisable here. Another option would be to use the Kalman-Yakubovich-Popov lemma to transform the frequency domain inequalities

$$a(z) > 0, \ \pm|H(z)|^2(a(z)Re(G_0(z)) - b(z)) \leq ya(z) \ \forall \ z \in \mathbf{T},$$

(which defines an infnite set of inequalities which are linear with respect to the coefficients of a, b but infinitely parameterized by $z \in \mathbf{T}$) into a set of *three* matrix inequalities, linear with respect to the coefficients of a, b, and three auxiliary symmetric matrix variables P_0, P_+, P_-. If n denotes the sum of m and the orders of H and G_0, the matrix inequalities will have sizes $m+1$,$n+1$ and $n+1$ respectively, and the sizes of $P_0, P_\pm$ will be m-by-m and n-by-n. Therefore, the model reduction problem will be reduced to semidefinite programming.

Similarly, the sampled version of the maximal real part optimal model reduction problem can be reduced to the convex optimization problem

$$y \to \min \text{ subject to } W_i|a(e^{jt_i})\mathrm{Re}(G_{0,i}) - b(e^{jt_i})| \leq y \ \ (1 \leq i \leq N),$$

where the decision parameters a, b are constrained by (5).

Note that the complexity of this quasi-convex optimization grows slowly with N, which can be very large. The complexity grows faster with m, but this does not appear to be a significant problem, since m, the desired reduced order, is small in most applications.

3 H-Infinity Modeling Error Bounds

In this section, H-Infinity approximation quality of maximal real part optimal reduced models is examined. For a given stable transfer function $G \in \mathbf{A}$ let

$$d_m^\alpha(G) = \min_{\hat{G} \in \mathbf{A}_m} \|G - \hat{G}\|_\alpha,$$

$$\hat{G}_m^\alpha = \arg \min_{\hat{G} \in \mathbf{A}_m} \|G - \hat{G}\|_\alpha,$$

denote the minimum and the argument of minimum in the corresponding optimal model reduction problems, where $\alpha \in \{\infty, h, r\}$ indicates one of the unweighted norms: H-Infinity, Hankel, or maximal real part. In the case when the optimal reduced model is not unique, the model delivered by a particular optimization algorithm can be considered. Let

$$\tilde{G}_m^\alpha = \hat{G}_m^\alpha + \arg \min_{c \in \mathbf{R}} \|G - \hat{G}_m^\alpha - c\|_\infty.$$

In other words, let $\tilde{G}_m^\alpha$ be the result of an adjustment of $\hat{G}_m^\alpha$ by an additive constant factor (which obviously does not change the order) to further optimize the H-Infinity model reduction error. Practically, the modified $\tilde{G}_m^\alpha$ is easy to calculate. Obviously, $\hat{G}_m^\infty = \tilde{G}_m^\infty$.

Since

$$\|\Delta\|_\infty \geq \|\Delta\|_h, \quad \|\Delta\|_\infty \geq \|\Delta\|_r$$

for all $\Delta \in \mathbf{A}$, the quantities $d_m^h(G)$ and $d_m^r(G)$ are lower bounds of $d_m^\infty(G)$. One can argue that a maximal real part modified optimal reduced model $\tilde{G}_m^r$ (or, alternatively, a Hankel optimal modified reduced model $\tilde{G}_m^h$) is an acceptable surrogate of $\hat{G}_m^\infty$ when $\|G - \tilde{G}_m^r\|_\infty$ is not much larger than d_m^r (respectively, when $\|G - \tilde{G}_m^h\|_\infty$ is not much larger than d_m^h). For the Hankel optimal reduced models, a theoretical evidence that this will frequently be the case is provided by the inequality

$$\|G - \tilde{G}_m^h\|_\infty \leq \sum_{k \geq m} d_m^h(G). \tag{8}$$

Since for a "nice" smooth transfer function G the numbers $d^h_m(G)$ converge to zero quickly, (8) gives some assurance of good asymptotic behavior of H-Infinity modeling errors for Hankel optimal reduced models.

Since $2\|\Delta\|_r \geq \|\Delta\|_h$ for all $\Delta \in \mathbf{A}$, one can argue that the maximal real part norm is "closer" to the H-Infinity norm than the Hankel norm. However, the author was not able to prove a formal statement confirming the conjectured asymptotic superiority of maximal real part reduced models over the Hankel reduced models. Instead, a theorem demonstrating asymptotic behavior roughly comparable to that of the Hankel reduced models is given below.

Theorem 1 *For all $G \in \mathbf{A}$ and $m > 0$*

$$\|G - \tilde{G}^r_m\|_\infty \leq 12 \sum_{k=0}^{\infty} 2^k m d^r_{2^k m}(G). \tag{9}$$

Proof. From (8), for every $f \in \mathbf{A}_n$

$$\|f - \tilde{f}^h_0\|_\infty \leq \sum_{k=0}^{\infty} d^h_k(f) \leq n d^h_0(f) = n\|f\|_h \leq 2n\|f\|_r.$$

Note that $\tilde{f}^h_0$ is a constant transfer function. We have

$$\|\hat{G}^r_n - \hat{G}^r_{2n}\|_r \leq \|\hat{G}^r_n - G\| + \|\hat{G}^r_{2n} - G\|_r \leq 2d^r_n(G).$$

Hence

$$\|\hat{G}^r_n - \hat{G}^r_{2n} - c_n\|_\infty \leq 12nd^r_n(G)$$

for an appropriately chosen real constant c_n. Hence

$$\|\hat{G}^r_m - G - c\|_\infty \leq 12 \sum_{k=0}^{\infty} 2^k m d^r_{2^k m}(G)$$

for an appropriately chosen constant c, which in turn implies (9). ■

It may appear that (8) is a much better bound than (9), since $d_m(G)$ is not being multiplied by m in (8). However, the calculations for $d_m \approx 1/m^q$ as $m \to \infty$, where $q > 1$, result in the same rate of asymtotic convergence for the two upper bounds:

$$\sum_{k=m}^{\infty} \frac{1}{k^q} \approx \frac{c_1}{m^{q-1}},$$

$$\sum_{k=m}^{\infty} 2^k m \frac{1}{(2^k m)^q} \approx \frac{c_2}{m^{q-1}}.$$

4 Minor Improvements and Numerical Experiments

There is a number of ways in which the H-Infinity quality of the maximal real part optimal reduced models can be improved. One simple trick is to re-optimize the numerator p of $\hat{G}_m^r = p/q$ with the optimal q being fixed (this is partially used in Hankel model reduction when $\hat{G}_m^h$ is being replaced by $\tilde{G}_m^h$). A further improvement of the *lower bound* $d_m^r(G)$ can be achieved when, in a weighted H-Infinity model reduction setup $\|W(G - \hat{G}_m)\|_\infty \to \min$ $\hat{G}_m$ will be replaced by

$$\hat{G}_m^e = \frac{b(z) + (z - \bar{z})c(z)}{a(z)},$$

where a, b are constrained by (5), and c is an arbitrary trigonometric polynomial of degree less than $m - 1$. Then optimization with respect to a, b, c is convex, and the optimal a can be used to get the *denominator* q of the reduced model, after which the numerator p is to be found via a separate optimization round. Note that, since

$$|G(z) - \hat{G}_m^e(z)| \geq |\mathrm{Re}(G(z)) - b(z)/a(z)|,$$

where the equality is achieved for $c \equiv 0$, the original maximal real part model reduction is a special case of the general scheme.

With these improvements implemented, the maximal real part model reduction performs reasonably well, as demonstrated by the following examples. The software used to produce the data (requires MATLAB and CPLEX) can be obtained by sending a request to ameg@mit.edu.

4.1 Functions With Delay

Here G is the infinite dimensional transfer function

$$G = \frac{1}{(1 - .9e^{-s})(1 + .3s)}.$$

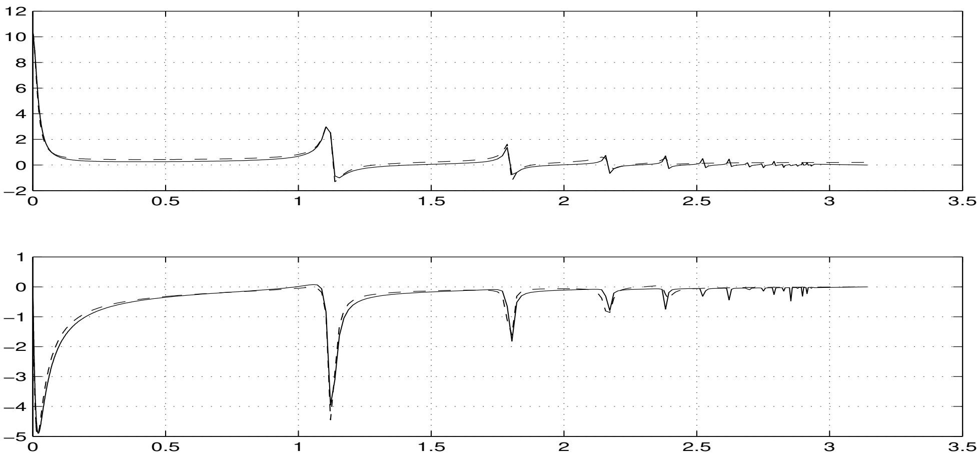

Figure 1: functions with delay

For a 10th order approximation, a lower bound $d^r_{10}(G) \geq 0.35$ was found. The resulting reduced 10th order model $\tilde{G}^r_{10}$ satisfies

$$\|G - \hat{G}^r_{10}\|_\infty < 1.4.$$

4.2 Focus Servo of a DVD Player

59 frequency samples obtained as an experimental data were provided. For a 10th order fit, the lower bound of about 1.6 was obtained. The actual H-Infinity error on the sampled data was approximately 6.2.

4.3 Fluid Dynamics Control

A jet engine outlet pressure is to be controlled by regulating a discharge valve in midstream. The complete dynamical model of the actuator dynamics is given by partial differential equations. Computational fluid dynamics simulations provided a 101 sample of the transfer function. A 5th order reduced model was sought for the system. The lower bound of achievable H-Infinity quality is 0.0069. A reduced model of quality 0.009 was found.

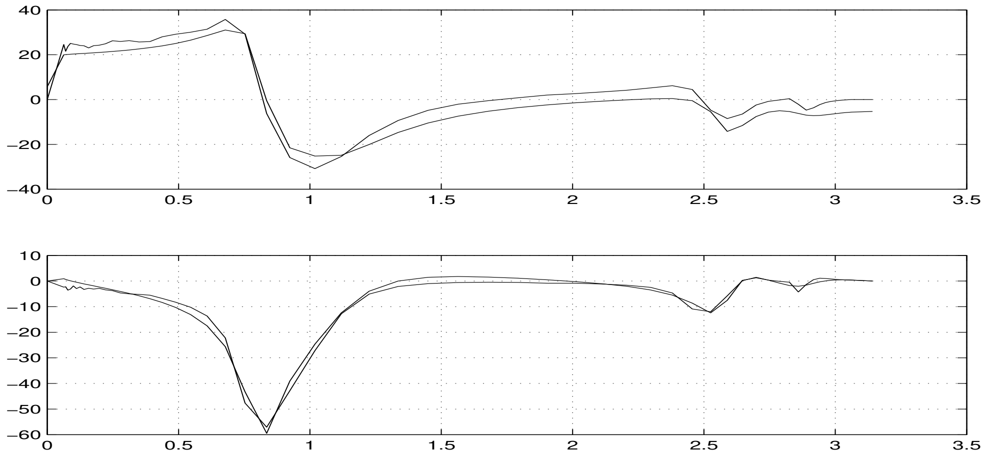

Figure 2: DVD focus servo

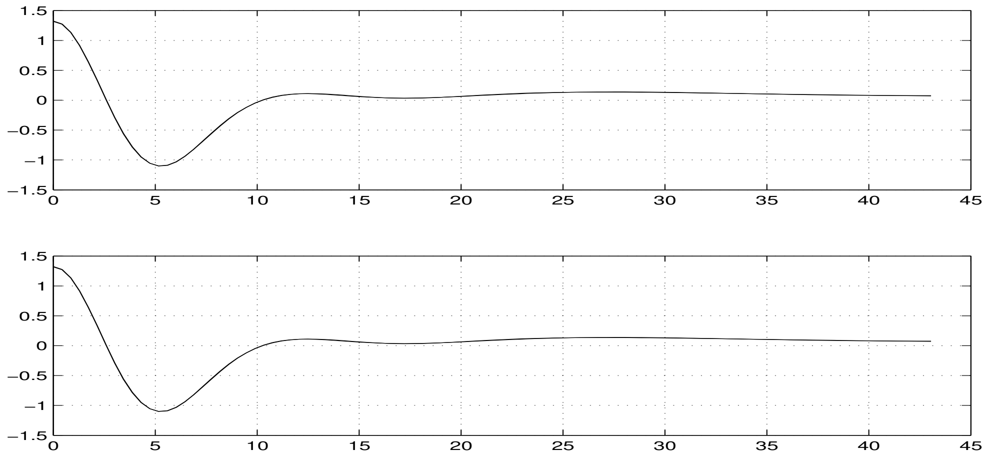

Figure 3: fluid dynamics control