INTERNATIONAL JOURNAL OF ROBUST AND NONLINEAR CONTROL
Int. J. Robust Nonlinear Control 2000; **10**:439–456

Approximate nonlinear output regulation based on the universal approximation theorem

Jin Wang[1], Jie Huang[1,*] and Stephen S.T. Yau[2]

[1] *Department of Mechanical and Automation Engineering, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong SAR, People's Republic of China*
[2] *Department of Mathematics, Statistics and Computer Science, University of Illinois, Chicago, IL 60607-7045, U.S.A.*

SUMMARY

The regulator equations arising from the nonlinear output regulation problem are a set of mixed partial and algebraic equations. Due to the nonlinear nature, it is difficult to obtain the exact solution of the regulator equations. This paper presents an approximation method for solving the regulator equations based on a class of feedforward neural networks. It is shown that a three-layer neural network can solve the regulator equations up to a prescribed arbitrarily small error, and this small error can be translated into a guaranteed steady-state tracking error for the closed-loop system. The method has led to an effective approach to approximately solving the nonlinear output regulation problem. Copyright © 2000 John Wiley & Sons, Ltd.

KEY WORDS: regulator equations, nonlinear servomechanism problem; neural networks; universal approximation theorem

1. INTRODUCTION

A central control problem is to design a feedback control law for a plant such that the output of the plant can asymptotically track a class of reference inputs and reject a class of disturbances while maintaining the closed-loop stability. Both the reference inputs and disturbances are generated by an autonomous time-invariant differential equation called exosystem. This problem is called output regulation problem or, alternatively, servomechanism problem or asymptotic tracking and disturbance rejection. For the class of linear systems, the solvability of the output regulation problem was thoroughly studied in the 1970s by many researchers including Davison, Francis, and Wonham, and Desoer, to name just a few [1–4]. For the class of nonlinear systems, research had long been limited to the special case where the reference inputs and disturbances are constant [4–6] until 1990 when Isidori and Byrnes published their work on the output regulation problem for nonlinear systems with time-varying reference inputs and disturbances [7]. A celebrated result in Reference [7] is that the solvability of the nonlinear output regulation problem is tied to the solvability of a set of partial differential and algebraic equations.

* Correspondence to: Jie Huang, Department of Mechanical and Automation Engineering, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong SAR, People's Republic of China.

Contract/grant sponsor: Hong Kong Research Grants Council; contract/grant number: CUHK4400/99E

This set of partial differential and algebraic equations is now known as the regulator equations or Isidori and Byrnes equations. Due to the nonlinear nature, it is usually impossible to obtain the exact solution for the regulator equations. Thus, it is necessary to develop approximation approaches to solving these equations. In fact, an approximation method based on the Taylor series expansion was proposed in Reference [11]. However, the Taylor series approximation method depends on the availability of the Taylor series expansions of several multivariable nonlinear functions defining the nonlinear plant and exosystem. In addition, Taylor-theorem-based approximation is valid only in a sufficiently small open neighbourhood of the origin. To overcome these difficulties, we have turned to a neural-network-based approximation method to solve the regulator equations. The foundation of our approach is the well-known universal approximation theorem [8]. This theorem enables the solution of the regulator equations to be approximately represented by three-layer feedforward neural networks. As a result, the problem of finding the solution of the regulator equations can be cast into a parameter optimization problem that can be handled by various gradient-based methods. This new method not only leads to a non-local approximation scheme, but also offer some well-known advantages associated with neural networks such as computational efficiency, and hardware realizability.

In the next section, after a brief review of the output regulation problem, we introduce the so-called approximate output regulation problem which aims to design a feedback control law that stabilizes the closed-loop system, and results in a guaranteed steady-state tracking error. In Section 3, we show that using three-layer neural networks, it is possible to solve the regulator equations up to a prescribed arbitrarily small error. In Section 4, we further show that the control law based on the neural networks approximation solution of the regulator equations can solve the approximate output regulation problem as stated in Section 2. In Section 5, we illustrate the application of our approach to the difficult asymptotic tracking problem of the inverted pendulum on a cart system. In comparison with the Taylor-series-based controller, our controller shows a significant performance enhancement in tracking sinusoidal signals with larger amplitudes. Finally, we close this paper with some remarks.

We will adopt the following notations in the sequel. Let $X \subset R^n$ and $f\colon X \to R$ be a real-valued function. An $n-tuple$ $\alpha = (\alpha_1, \ldots, \alpha_n)$ of non-negative integers is called a multiindex. We denote $|\alpha| = \alpha_1 + \cdots + \alpha_n$ the order of the multiindex α and

$$D^\alpha f(x) = \frac{\partial^{\alpha_1 + \cdots + \alpha_n} f}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}}(x)$$

for the corresponding partial derivative of a sufficiently smooth function f of $x \in X$. Let $C^k(X)$ be the space of all functions f which, together with all their partial derivatives D^α of order $|\alpha| \leqslant k$, are continuous on X. Let $f_i\colon X \to R$, $i = 1, \ldots, n$, be real valued functions. Then $f = [f_1, \ldots, f_n]^{\mathrm{T}} \in C^k(X)$ if $f_i \in C^k(X)$, $i = 1, \ldots, n$. For a n by m matrix $A = [a_{ij}]$, define $\|A\| = \sqrt{(\sum_{i=1}^n \sum_{j=1}^m a_{ij}^2)}$.

2. PROBLEM DESCRIPTION

We begin by introducing the output regulation problem. Consider a nonlinear plant described by

$$\begin{aligned} \dot{x}(t) &= f(x(t), u(t), v(t)), \quad x(0) = x_0, \ t \geqslant 0 \\ y(t) &= h(x(t), u(t), v(t)) \end{aligned} \tag{1}$$

where $x(t)$ is the n-dimensional plant state, $u(t)$ is the m-dimensional plant input, $y(t)$ is the p-dimensional plant output representing tracking error, and $v(t)$ is a q-dimensional exogenous signal representing both reference inputs and disturbances. It is assumed that $v(t)$ is generated by a q-dimensional exosystem

$$\dot{v}(t) = a(v(t)), \quad v(0) = v_0, \quad t \geqslant 0 \tag{2}$$

For simplicity, we assume $f \in C^k(R^n \times R^m \times R^q)$, $h \in C^k(R^n \times R^m \times R^q)$, and $a \in C^k(R^q)$ with $k \geqslant 2$, and $f(0, 0, 0) = 0$, $h(0, 0, 0) = 0$, and $a(0) = 0$.

We will concentrate on the state feedback control law of the form

$$u = \psi(x, v) \tag{3}$$

where $\psi \in C^k(R^n \times R^q)$ and $\psi(0, 0) = 0$. This control law results in a closed-loop system as follows:

$$\begin{aligned} \dot{x}(t) &= f_c(x, v) \triangleq f(x, \psi(x, v), v) \\ y &= h_c(x, v) \triangleq h(x, \psi(x, v), v) \end{aligned} \tag{4}$$

In terms of the closed-loop system, we can describe the output regulation problem as follows [7]:

Output regulation problem: Designing a feedback control law of the form $u = \psi(x, v)$ such that the closed-loop system satisfies

R1: The eigenvalues of $(\partial f_c / \partial x)(0, 0)$ have negative real part.

R2: For sufficiently small x_0 and v_0, the solution of the closed-loop system exists for all $t \geqslant 0$, and

$$\lim_{t \to \infty} \| h_c x(t), v(t) \| = 0 \tag{5}$$

Remark 1

It is shown in Reference [7] that the first requirement ensures the (local) asymptotic stability of the closed-loop system, which in turn ensures the (local) bounded-input bounded state property for the closed-loop system (4).

The solvability of the above problem is given in Reference [7] and can be rephrased as follows:

Theorem 1

Assume

A1. The pair $\{(\partial f/\partial x)(0, 0, 0), (\partial f/\partial u)(0, 0, 0)\}$ is stabilizable, and

A2. The equilibrium of exosystem (2) is stable, and there is an open neighbourhood of the point $v = 0$ in which every point is Poisson stable.

Then the output regulation problem is solvable if and only if

A3. There exist two sufficiently smooth functions $\mathbf{x}(v)$ and $\mathbf{u}(v)$ defined in an open neighbourhood V of the origin of R^q such that $\mathbf{x}(0) = 0$, $\mathbf{u}(0) = 0$, and, for all $v \in V$,

$$\frac{\partial \mathbf{x}(v)}{\partial v} a(v) = f(\mathbf{x}(v), \mathbf{u}(v), v) \tag{6}$$

$$0 = h(\mathbf{x}(v), \mathbf{u}(v), v) \tag{7}$$

Remark 2

A simple state feedback control law that solves the output regulation problem is given by Isidori and Byrnes [7]

$$u(t) = \mathbf{u}(v(t)) + K[x(t) - \mathbf{x}(v(t))] \tag{8}$$

where K is a feedback gain such that all the eigenvalues of the matrix

$$\frac{\partial f}{\partial x}(0, 0, 0) + \frac{\partial f}{\partial u}(0, 0, 0)K \tag{9}$$

have negative real parts. Such a feedback gain always exists under assumption A1.

Equations (6) and (7) are known as the regulator equations. Due to the nonlinear nature of the plant in question, it is usually impossible to obtain the exact solution of the regulator equations. For example, for the well-known ball and beam system as studied in Reference [9], and for the inverted pendulum on a cart system as studied in Reference [10], there is no known method that is able to give the exact solution to the associated regulator equations. Thus, it is necessary to develop approximation method to solve the regulator equations, and study the performance of the control law synthesized on the basis of the approximate solution of the regulator equations. For this purpose, we introduce the following,

Approximate output regulation problem: Given $\varepsilon > 0$, designing a control law of the form $u = \psi(x, v)$ such that the closed-loop system (4) has the property that, for all sufficiently small initial conditions x_0 and v_0, the closed-loop system has a bounded solution for all $t \geqslant 0$, and

$$\limsup_{t \to \infty} \| h_c(x(t), v(t)) \| \leqslant \varepsilon \tag{10}$$

Remark 3

Approximate output regulation problem has been studied in Reference [11] where Taylor series is used to approximately solve the regulator equations. Nevertheless, this approach suffers from some disadvantages as mentioned in the Introduction. The objective of this paper is to develop an alternative approach to solving the approximate output regulation problem based on the universal approximation theorem of the multilayer neural networks [8]. This approach is not only computationally more efficient but also leads to an approximation that is valid in any compact set on which the mapping is defined.

3. NEURAL NETWORKS APPROXIMATION OF THE REGULATOR EQUATIONS

The universal approximation theorem has been developed by several people. One version of this theorem taken from Reference [8] can be rephrased as follows:

Theorem 2

Let $\phi \in C^k(R)$ be non-constant and bounded real-valued function. Let Γ be any compact subset of R^n and $f \in C^k(\Gamma)$ be a real-valued function on Γ. Then for any $\varepsilon > 0$, there exists an integer N, and real numbers w_i^O, w_{ij}^I, and b_i, $i = 1, \ldots, N$, and $j = 1, \ldots, n$ such that

$$\hat{f}(x_1, \ldots, x_n) = \sum_{i=1}^{N} w_i^O \phi\left(\sum_{j=1}^{n} w_{ij}^I x_j + b_i\right) \tag{11}$$

satisfies

$$\max_{|\alpha| \leqslant k} \sup_{x \in \Gamma} |D^{\alpha}(\hat{f}(x) - f(x))| < \varepsilon \tag{12}$$

The mapping defined by (11) is called a three-layer feedforward neural network where x_i, $i = 1, \ldots, n$, is the input, f the output, and the integer N the number of hidden neurons. The real numbers w_i^O, w_{ij}^I, and b_i, $i = 1, \ldots, N$, and $j = 1, \ldots, n$ are called the weights of the neural network. The function ϕ is called activation function. The common activation functions include sigmoid function or hyper tangent function.

Now, assume $\mathbf{x}(v)$ and $\mathbf{u}(v)$ are the solution of the regulator equations (6) and (7) defined in an open neighbourhood V of the origin of R^q, and $\mathbf{x}, \mathbf{u} \in C^k(V)$ with $k \geqslant 2$. Let Γ be any compact subset of V, by the universal approximation theorem, given any $\gamma > 0$, there exist integers $N_{x1}, \ldots, N_{xn}$, $N_{u1}, \ldots, N_{um}$, scalars $w_{1j}^{XO}, \ldots, w_{nj}^{XO}$, $b_{1j}^X, \ldots, b_{nj}^X$, $w_{1j}^{UO}, \ldots, w_{mj}^{UO}$, $b_{1j}^U, \ldots, b_{mj}^U$, and q-dimensional row vectors $W_{1j}^{XI}, \ldots, W_{nj}^{XI}$, $W_{1j}^{UI}, \ldots, W_{mj}^{UI}$, $j = 1, \ldots, N$ such that

$$\hat{\mathbf{x}}(v) = \begin{bmatrix} \sum_{j=1}^{N_{x1}} w_{1j}^{XO} \phi(W_{1j}^{XI} v + b_{1j}^X) \\ \vdots \\ \sum_{j=1}^{N_{xn}} w_{nj}^{XO} \phi(W_{nj}^{XI} v + b_{nj}^X) \end{bmatrix}, \quad \hat{\mathbf{u}}(v) = \begin{bmatrix} \sum_{j=1}^{N_{u1}} w_{1j}^{UO} \phi(W_{1j}^{UI} v + b_{1j}^U) \\ \vdots \\ \sum_{j=1}^{N_{um}} w_{mj}^{UO} \phi(W_{mj}^{UI} v + b_{mj}^U) \end{bmatrix} \tag{13}$$

satisfy

$$\begin{aligned} &\max_{v \in \Gamma} \| \mathbf{x}(v) - \hat{\mathbf{x}}(v) \| < \gamma \\ &\max_{v \in \Gamma} \| \mathbf{u}(v) - \hat{\mathbf{u}}(v) \| < \gamma, \\ &\max_{v \in \Gamma} \left\| \frac{\partial(\mathbf{x}(v) - \hat{\mathbf{x}}(v))}{\partial v} \right\| < \gamma \end{aligned} \tag{14}$$

Next we will show that, by having γ sufficiently small, the two functions $\hat{\mathbf{x}}(v)$ and $\hat{\mathbf{u}}(v)$ as described in (13) can solve the regulator equations up to an arbitrarily small error ε_r.

Lemma 1

Under assumption A3, let $Q = \{(\mathbf{x}(v), \mathbf{u}(v), v) \mid v \in V\}$, G an open, connected subset of Q relatively compact in Q, and V_G the projection of G onto V. Then, given any $\varepsilon_r > 0$, there exist two functions $\hat{\mathbf{x}}(v) \in C^k(V_G)$ and $\hat{\mathbf{u}}(v) \in C^k(V_G)$ of the form (13) satisfying, for all $v \in V_G$,

$$\left\| \frac{\partial \hat{\mathbf{x}}(v)}{\partial v} a(v) - f(\hat{\mathbf{x}}(v), \hat{\mathbf{u}}(v), v) \right\| < \varepsilon_r \tag{15}$$

$$\| h(\hat{\mathbf{x}}(v), \hat{\mathbf{u}}(v), v) \| < \varepsilon_r \tag{16}$$

Proof. Since $\bar{G}$ and $\bar{V}_G$ are compact, and f, h, and a are C^k, $k \geqslant 2$, there exists a finite number Λ such that

$$\left\{ \max_{(x,u,v)\in\bar{G}} \left\| \frac{\partial f(x,u,v)}{\partial x} \right\|, \sup_{(x,u,v)\in\bar{G}} \left\| \frac{\partial f(x,u,v)}{\partial u} \right\|, \sup_{(x,u,v)\in\bar{G}} \left\| \frac{\partial h(x,u,v)}{\partial x} \right\| \right.$$
$$\left. \sup_{(x,u,v)\in\bar{G}} \left\| \frac{\partial h(x,u,v)}{\partial u} \right\|, \sup_{v\in\bar{V}_G} \| a(v) \| \right\} < \Lambda$$

Let $z = (x, u, v)$, and define the distance between z and $\bar{G}$ by

$$d(z, \bar{G}) = \min_{(x(v),\mathbf{u}(v),v)\in\bar{G}} (\| x - \mathbf{x}(v) \|^2 + \| u - \mathbf{u}(v) \|^2)^{1/2}$$

Let

$$Q_\delta = \{ z \in R^n \times R^m \times R^q \,|\, d(z, \bar{G}) < \delta \}$$

Then there exists sufficiently small $\delta > 0$ such that

$$\max \left\{ \sup_{z\in Q_\delta} \left\| \frac{\partial f(x,u,v)}{\partial x} \right\|, \sup_{z\in Q_\delta} \left\| \frac{\partial f(x,u,v)}{\partial u} \right\| \right.$$
$$\left. \sup_{z\in Q_\delta} \left\| \frac{\partial h(x,u,v)}{\partial x} \right\|, \sup_{z\in Q_\delta} \left\| \frac{\partial h(x,u,v)}{\partial u} \right\|, \sup_{v\in\bar{V}_G} \| a(v) \| \right\} < \Lambda$$

Fix δ, and let

$$\gamma = \min \left\{ \frac{\varepsilon_r}{3\Lambda}, \frac{\delta}{\sqrt{2}} \right\}$$

Then, by the universal approximation theorem, there exist two functions $\hat{\mathbf{x}}(v)$ and $\hat{\mathbf{u}}(v)$ described in (11) such that (14) holds with $\Gamma = \bar{V}_G$.

We now first show that $\{(\hat{\mathbf{x}}(v), \hat{\mathbf{u}}(v), v) \,|\, v \in V_G\} \subset Q_\delta$. To this end, let

$$z(v) = \begin{bmatrix} \mathbf{x}(v) \\ \mathbf{u}(v) \\ v \end{bmatrix}, \quad \hat{z}(v) = \begin{bmatrix} \hat{\mathbf{x}}(\mathbf{v}) \\ \hat{\mathbf{u}}(\mathbf{v}) \\ v \end{bmatrix}$$

Then, $\forall v \in V_G$,

$$(d(\hat{z}(v), \bar{G}))^2 \leqslant \| \hat{\mathbf{x}}(v) - \mathbf{x}(v) \|^2 + \| \hat{\mathbf{u}}(v) - \mathbf{u}(v) \|^2$$
$$< 2\gamma^2 \leqslant \delta^2$$

Thus, $\{(\hat{\mathbf{x}}(v), \hat{\mathbf{u}}(v), v) \mid v \in V_G\} \subset Q_\delta$.

Since Q_δ is a convex open set in $R^n \times R^m \times R^q$, and $\forall v \in V_G$, $(\hat{\mathbf{x}}(v), \hat{\mathbf{u}}(v), v), (\mathbf{x}(v), \mathbf{u}(v), v) \in Q_\delta$, applying Theorem 9.19 of Reference [12] gives, $\forall v \in V_G$,

$$\| f(\hat{\mathbf{x}}(v), \hat{\mathbf{u}}(v), v) - f(\mathbf{x}(v), \mathbf{u}(v), v) \| \leqslant \Lambda \| \mathbf{x}(v) - \hat{\mathbf{x}}(v) \| + \Lambda \| \mathbf{u}(v) - \hat{\mathbf{u}}(v) \|$$
$$< 2\Lambda\gamma$$

Thus, $\forall v \in V_G$,

$$\left\| \frac{\partial \hat{\mathbf{x}}(v)}{\partial v} a(v) - f(\hat{\mathbf{x}}(v), \hat{\mathbf{u}}(v), v) \right\| = \left\| \frac{\partial \mathbf{x}(v)}{\partial v} a(v) - \frac{\partial \hat{\mathbf{x}}(v)}{\partial v} a(v) + f(\hat{\mathbf{x}}(v), \hat{\mathbf{u}}(v), v) - f(\mathbf{x}(v), \mathbf{u}(v), v) \right\|$$

$$\leqslant \left\| \frac{\partial \mathbf{x}(v)}{\partial v} a(v) - \frac{\partial \hat{\mathbf{x}}(v)}{\partial v} a(v) \right\| + \| f(\hat{\mathbf{x}}(v), \hat{\mathbf{u}}(v), v) - f(\mathbf{x}(v), \mathbf{u}(v), v) \|$$

$$\leqslant \Lambda \left\| \frac{\partial (\mathbf{x}(v) - \hat{\mathbf{x}}(v))}{\partial v} \right\| + 2\Lambda\gamma < 3\Lambda\gamma \leqslant \varepsilon_{\mathrm{r}}$$

Similarly, we have, $\forall v \in V_G$,

$$\| h(\hat{\mathbf{x}}(v), \hat{\mathbf{u}}(v), v) \| = \| h(\hat{\mathbf{x}}(v), \hat{\mathbf{u}}(v), v) - h(\mathbf{x}(v), \mathbf{u}(v), v) \| < \varepsilon_{\mathrm{r}}$$

Remark 4

Without loss of generality, we can always assume that $\hat{\mathbf{x}}(0) = 0$ and $\hat{\mathbf{u}}(0) = 0$. Otherwise, let $\hat{\mathbf{x}}_1(v) = \hat{\mathbf{x}}(v) - \hat{\mathbf{x}}(0)$ and $\hat{\mathbf{u}}_1(v) = \hat{\mathbf{u}}(v) - \hat{\mathbf{u}}(0)$. Then

$$\max_{\bar{V}_G} \| \mathbf{x}(v) - \hat{\mathbf{x}}_1(v) \| < 2\gamma$$

$$\max_{\bar{V}_G} \| \mathbf{u}(v) - \hat{\mathbf{u}}_1(v) \| < 2\gamma,$$

$$\max_{\bar{V}_G} \left\| \frac{\partial (\mathbf{x}(v) - \hat{\mathbf{x}}_1(v))}{\partial v} \right\| < \gamma$$

$\hat{\mathbf{x}}_1(0) = 0$ and $\hat{\mathbf{u}}_1(0) = 0$. Thus, if γ is sufficiently small, we still have, $\forall v \in V_G$,

$$\left\| \frac{\partial \hat{\mathbf{x}}_1(v)}{\partial v} a(v) - f(\hat{\mathbf{x}}_1(v), \hat{\mathbf{u}}_1(v), v) \right\| < \varepsilon_{\mathrm{r}}$$

$$\| h(\hat{\mathbf{x}}_1(v), \hat{\mathbf{x}}_1(v), v) \| < \varepsilon_{\mathrm{r}}$$

Remark 5

It is interesting to make a comparison between the neural network approximation scheme proposed here and Taylor series approximation scheme developed in Reference [11]. Both schemes can lead to approximation solution of the regulator equations. However, the approximation solution of the regulator equations based on the Taylor theorem is valid only in a sufficiently small neighborhood of the origin of R^q even though $V = R^q$. In contrast, the approximation solution of the regulator equations based on the universal approximation theorem is valid in V_{G} which can be large so long as $\bar{G}$ is compact relative to Q.

4. SOLVABILITY OF THE PROBLEM

In this section, we will further show that good approximation of the solution of the regulator equations as measured by the error bound ε_r can be translated into a small steady-state tracking error of the closed-loop system, thus leading to the solution of the approximate output regulation problem. Let us first consider the property of the closed-loop system under a state feedback control law of the form

$$\psi(x, v) = \hat{\mathbf{u}}(v) + K(x - \hat{\mathbf{x}}(v)), \quad x \in R^n, \quad v \in V_G \tag{17}$$

where K is such that (9) is stable, and $\hat{\mathbf{x}}(v)$ and $\hat{\mathbf{u}}(v)$ are such that (14) holds for some given ε_r for $v \in V_G$. Clearly the closed-loop system (4) resulting from this control law satisfies R1 and

$$f_c(\hat{\mathbf{x}}(v), v) = f(\hat{\mathbf{x}}(v), \hat{\mathbf{u}}(v), v), \quad v \in V_G$$

$$h_c(\hat{\mathbf{x}}(v), v) = h(\hat{\mathbf{x}}(v), \hat{\mathbf{u}}(v), v), \quad v \in V_G$$

Consequently, we have

$$\left\| \frac{\partial \hat{\mathbf{x}}(v)}{\partial v} a(v) - f_c(\hat{\mathbf{x}}(v), v) \right\| < \varepsilon_r, \quad v \in V_G \tag{18}$$

$$\| h_c(\hat{\mathbf{x}}(v), v) \| < \varepsilon_r, \quad v \in V_G \tag{19}$$

Lemma 2

Suppose the closed-loop system satisfies R1. Let $\beta(v)$ be a C^2 n-dimensional vector valued function locally defined in a neighbourhood of the origin of R^q satisfying $\beta(0) = 0$. Let $v(t)$ be the solution of (2) satisfying $v(0) = v_0$, and let $q(t) = \beta(v(t))$. Then there exist constants $\lambda, \xi_1, \xi_2, \xi_3 > 0$, and $\alpha \geqslant 1$ such that if

$$\begin{aligned} \| x(0) - q(0) \| &< \xi_1 \\ \| v(t) \| &< \xi_2, \quad t \geqslant 0, \\ \| f_c(q(t), v(t)) - \dot{q}(t) \| &< \xi_3, \quad t \geqslant 0 \end{aligned} \tag{20}$$

then the closed-loop system has a unique solution $x(t)$ that is defined for all $t \geqslant 0$, and satisfies

$$\| x(t) - q(t) \| < \alpha e^{-\lambda t} \| x(0) - q(0) \| + \int_0^t \alpha e^{-\lambda(t-\tau)} \| f_c(q(\tau), v(\tau)) - \dot{q}(\tau) \| \, d\tau \tag{21}$$

Proof. The proof is similar to that of Lemma 2.1 of Reference [11], and is thus omitted.

Theorem 3.

Suppose the closed-loop system satisfies R1, (18), and (19) for some sufficiently small ε_r, and exosystem (2) satisfies Assumption A2. Then,

(i) For all sufficiently small x_0 and v_0, the closed-loop system has a unique bounded solution $x(t)$ defined for all $t \geqslant 0$, and
(ii) There exist $M > 0$ such that

$$\overline{\lim_{t \to \infty}} \| h_c(x(t), v(t)) \| < M \varepsilon_r \tag{22}$$

Proof. First note that $\hat{\mathbf{x}}(v)$ has all the properties of $\beta(v)$ described in Lemma 2. We can let $q(t) = \hat{\mathbf{x}}(v(t))$. Since the closed-loop system satisfies R1, by Lemma 2, there exist ξ_1, ξ_2, and ξ_3 such that (20) implies (21). Fix ξ_1, ξ_2, and ξ_3. We will show that (20) can be made satisfied by having x_0, v_0, and ε_r sufficiently small. In fact,

$$\| x(0) - q(0) \| \leqslant \| x(0) \| + \| \hat{\mathbf{x}}(v(0)) \| = \| x_0 \| + \| \hat{\mathbf{x}}(v_0) \|$$

Thus, the first inequality of (20) can always be satisfied with x_0 and v_0 sufficiently small. Also, since the exosystem satisfies assumption A2, the second inequality of (20) can also be satisfied with v_0 sufficiently small. Finally, note that the closed-loop system satisfies (18), and for sufficiently small v_0, $v(t) \in V_G$, $\forall t \geqslant 0$. Thus,

$$\begin{aligned} \| f_c(q(t), v(t)) - \dot{q}(t) \| &= \left\| f_c(q(t), v(t)) - \frac{\partial \hat{\mathbf{x}}(v(t))}{\partial v} a(v(t)) \right\| \\ &\leqslant \max_{v \in \bar{V}_G} \left\| f_c(\hat{\mathbf{x}}(v), v) - \frac{\partial \hat{\mathbf{x}}(v)}{\partial v} a(v) \right\| < \varepsilon_r \end{aligned}$$

Therefore the third inequality of (20) is satisfied with $\xi_3 = \varepsilon_r$. Similarly, noting that the closed-loop system satisfies (19) gives

$$\| h_c(q(t), v(t)) \| \leqslant \max_{v \in \bar{V}_G} \| h_c(\hat{\mathbf{x}}(v), v) \| < \varepsilon_r \tag{23}$$

Since the closed-loop system satisfies (18), by Lemma 2, there exist constants $\lambda > 0$, and $\alpha \geqslant 1$, such that (21) holds. Let $M_1 = \alpha/\lambda$ and taking upper limits on both sides of (21) gives

$$\overline{\lim_{t \to \infty}} \| x(t) - q(t) \| \leqslant M_1 \xi_3 = M_1 \varepsilon_r \tag{24}$$

Next, note that there exists a finite number b_1 such that

$$\sup_{\| v \| < \xi_2} \| \hat{\mathbf{x}}(v) \| < b_1$$

Thus, $\| q(t) \| = \| \hat{\mathbf{x}}(v(t)) \| < b_1$ for $\forall t \geqslant 0$. Also note that (21) implies

$$\sup_{0 \leqslant t \leqslant \infty} \| x(t) - q(t) \| \leqslant \alpha \xi_1 + M_1 \xi_3$$

Therefore, there exists a finite number b_2 such that

$$\| x(t) \| \leqslant \| q(t) \| + \| x(t) - q(t) \| < b_2, \quad \forall t \geqslant 0$$

Let

$$M_2 = \sup_{\| x \| \leqslant b_2, \, \| v \| \leqslant \xi_2} \left\| \frac{\partial h_c(x, v)}{\partial x} \right\|$$

Then M_2 is finite, and is such that

$$\| h_c(x(t), v(t)) - h_c(q(t), v(t)) \| \leqslant M_2 \| x(t) - q(t) \|, \quad t \geqslant 0 \tag{25}$$

Thus, using (23) and (25) gives

$$\begin{aligned}\| h_c(x(t), v(t)) \| &\leqslant \| h_c(x(t), v(t)) - h_c(q(t), v(t)) \| + \| h_c(q(t), v(t)) \| \\ &\leqslant M_2 \| x(t) - q(t) \| + \varepsilon_r, \quad t \geqslant 0\end{aligned} \tag{26}$$

Taking upper limits gives

$$\begin{aligned}\limsup_{t \to \infty} \| h_c(x(t), v(t)) \| &\leqslant M_2 M_1 \varepsilon_r + \varepsilon_r \\ &= (M_1 M_2 + 1) \varepsilon_r\end{aligned} \tag{27}$$

Letting $M = M_1 M_2 + 1$ concludes the proof. □

Remark 6

Under assumptions A1, and A3, there always exist control laws of the form (17) such that the closed-loop system satisfies R1, (18), and (19). Thus, Theorem 3 can also be rephrased as the following:

Under assumptions A1–A3, there exist a control law of the form (17) with $\hat{\mathbf{x}}(v)$ and $\hat{\mathbf{u}}(v)$ being in the form (13) that solves the approximate output regulation problem.

5. ALGORITHM AND EXAMPLE

Having shown that neural networks can be used to approximately solve the output regulation problem with arbitrarily small error, we turn to the problem of developing an algorithm to find the desired neural networks. To explicitly indicate the reliance of the neural network approximation of the solution of the regulator equations on the weights of the neural networks, we will adopt the notation $\hat{\mathbf{x}}(W, v)$ and $\hat{\mathbf{u}}(W, v)$ in the sequel where W is a parameter vector consisting of all of the weights of the neural networks as given in (13). The dimension of W is denoted by s_N which is determined by the vector $N = [N_{x1}, \ldots, N_{xn}, N_{u1}, \ldots, N_{um}]$. By the universal approximate theorem and Lemma 1, for some given $\varepsilon_r > 0$, there exist an vector N, and a parameter vector $\hat{W} \in R^{s_N}$ such that

$$\sup_{v \in \Gamma} \left\| \frac{\partial \hat{\mathbf{x}}(\hat{W}, v)}{\partial v} a(v) - f(\hat{\mathbf{x}}(\hat{W}, v), \hat{\mathbf{u}}(\hat{W}, v), v) \right\| \leqslant \varepsilon_r \tag{28}$$

$$\sup_{v \in \Gamma} \| h(\hat{\mathbf{x}}(\hat{W}, v), \hat{\mathbf{u}}(\hat{W}, v), v) \| \leqslant \varepsilon_r \tag{29}$$

where $\Gamma \in V_G$ is some compact subset of R^q.

Next let

$$J(W, v) = \left\| \frac{\partial \hat{\mathbf{x}}(W, v)}{\partial v} a(v) - f(\hat{\mathbf{x}}(W, v), \hat{\mathbf{u}}(W, v), v) \right\|^2 + \| h(\hat{\mathbf{x}}(W, v), \hat{\mathbf{u}}(W, v), v) \|^2 \tag{30}$$

Clearly, if for some W,

$$\sup_{v\in\Gamma} J(W, v) \leqslant \varepsilon_{\mathrm{r}} \tag{31}$$

both (28) and (29) will be satisfied.

Since $J(W, v)$ depends on both v and W, there is no effective numerical method to solve (31). To circumvent this difficulty, we discretize (31) by letting

$$Q(W) = \sum_{v\in\Gamma_{\mathrm{d}}} J(W, v) \tag{32}$$

where Γ_{d} is a subset of Γ consisting of finitely many elements of Γ. If for some N and $\hat{W} \in R^{s_N}$,

$$Q(\hat{W}) < \varepsilon_{\mathrm{r}}^2 \tag{33}$$

Then we have

$$\max_{v\in\Gamma_{\mathrm{d}}} \left\| \frac{\partial \hat{\mathbf{x}}(\hat{W}, v)}{\partial v} a(v) - f(\hat{\mathbf{x}}(\hat{W}, v), \hat{\mathbf{u}}(\hat{W}, v), v) \right\| < \varepsilon_{\mathrm{r}} \tag{34}$$

$$\max_{v\in\Gamma_{\mathrm{d}}} \| h(\hat{\mathbf{x}}(\hat{W}, v), \hat{\mathbf{u}}(\hat{W}, v), v) \| < \varepsilon_{\mathrm{r}} \tag{35}$$

Existence of such N and $\hat{W}$ is also guaranteed by the universal approximation theorem.

When Γ_{d} is sufficiently *dense* in Γ, we have reason to believe that the solution of (34) and (35) is a good approximation of that of (28) and (29).

Remark 7

Since, for each fixed N, $Q(W)$ relies only on the parameter W, the optimal weight that minimizes $Q(W)$ can be searched by any minimization technique. For example, applying the steepest descent method leads to the following update law of the weight vector:

$$W_{j+1} = W_j - \eta_j \frac{\partial Q(W_j)}{\partial W}, \quad j = 0, 1, \ldots, \tag{36}$$

with η_j the step size. Thus, the problem of looking for the approximation solution of the regulator equation is converted into a parameter optimization problem.

Remark 8

Though gradient-based methods do not necessarily lead to a weight that minimize $Q(W)$, there is no need, in practice, to really search the optimal weight. What are needed are some vector N and weight W that make $Q(W)$ sufficiently small. Of course, the particular vector N that guarantees the satisfaction of (33) is not known *a priori*. Therefore, iteration on N is often inevitable.

Inverted pendulum on a cart example: The inverted-pendulum on a cart illustrated in Figure 1 is a well-known unstable nonlinear system that can be found in many universities' control labs. The motion of the system can be described by

$$(M + m)\ddot{x} + ml(\ddot{\theta}\cos\theta - \dot{\theta}^2 \sin\theta) + b\dot{x} = u$$

$$m(\ddot{x}\cos\theta + l\ddot{\theta} - g\sin\theta) = 0$$

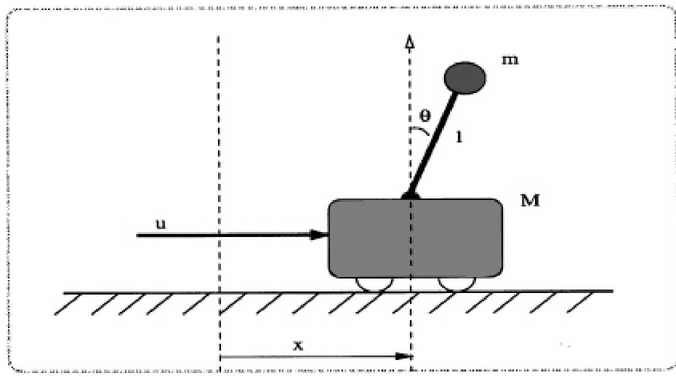

Figure 1: Inverted Pendulum with cart.

where M is the mass of the cart, m is the mass of the block on the pendulum, l the length of the pendulum, g the acceleration due to gravity, b the coefficient of viscous friction for motion of the cart, θ is the angle the pendulum makes with vertical, $\mathbf{x}$ the position of the cart, and u is the applied force [13]. With the choice of the state variables $x_1 = x$, $x_2 = \dot{x}$, $x_3 = \theta$, $x_4 = \dot{\theta}$, the state space equations of the system are

$$\begin{aligned}
\dot{x}_1 &= x_2 \\
\dot{x}_2 &= \frac{1}{M + m(\sin x_3)^2}(u + mlx_4^2 \sin x_3 - bx_2 - mg\cos x_3 \sin x_3) \\
\dot{x}_3 &= x_4 \\
\dot{x}_4 &= \frac{1}{l(M + m(\sin x_3)^2)}((M+m)g\sin x_3 - u\cos x_3 + bx_2\cos x_3 - mlx_4^2 \sin x_3 \cos x_3) \\
y(t) &= x_1(t)
\end{aligned} \tag{37}$$

where the parameters are given by $b = 12.98$ kg/s, $M = 1.378$ kg, $l = 0.325$ m, $g = 9.8$ m/s^2, and $m = 0.051$ kg.

We will consider the problem of designing a state feedback control law for this system such that the position of the cart can asymptotically track a sinusoidal function $y_d(t) = A\sin\omega t$. It is known that this is a non-minimum-phase system [10]. Thus, it is impossible to solve this problem by using conventional inversion-based control methods such as input–output linearization. Reference [10] has used the Taylor-series-based approach to solving this problem approximately. In the sequel, we will apply the method developed in this paper to achieve asymptotic tracking approximately. To this end, let us first put the inverted-pendulum system into the following standard form:

$$\begin{aligned}
\dot{x} &= f(x) + g(x)u \\
y &= h(x)
\end{aligned} \tag{38}$$

where

$$f(x) = \begin{bmatrix} x_2 \\ \dfrac{1}{M + m(\sin x_3)^2}(mlx_4^2 \sin x_3 - bx_2 - mg \cos x_3 \sin x_3) \\ x_4 \\ \dfrac{1}{l(M + m(\sin x_3)^2)}((M+m)g \sin x_3 + bx_2 \cos x_3 - mlx_4^2 \sin x_3 \cos x_3) \end{bmatrix}$$

$$g(x) = \begin{bmatrix} 0 \\ \dfrac{1}{M + m(\sin x_3)^2} \\ 0 \\ \dfrac{-\cos x_3}{l(M + m(\sin x_3)^2)} \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$$

and $h(x) = x_1$.

Also, introduce the following exosystem:

$$\dot{v} = Sv, \quad t \geqslant 0, \quad v(0) = v_0 \tag{39}$$

with

$$v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}, \quad S = \begin{bmatrix} 0 & \omega \\ -\omega & 0 \end{bmatrix}, \quad v_0 = \begin{bmatrix} 0 \\ A \end{bmatrix}$$

Then clearly, $y_d(t) = v_1$.

The Jacobian linearization of the inverted pendulum on a cart system can be calculated as follows:

$$\frac{\partial f(0)}{\partial x} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & -\dfrac{b}{M} & -\dfrac{mg}{M} & 0 \\ 0 & 0 & 0 & 1 \\ 0 & \dfrac{b}{lM} & \dfrac{(M+m)}{lM}g & 0 \end{bmatrix}, \quad g(0) = \begin{bmatrix} 0 \\ \dfrac{1}{M} \\ 0 \\ -\dfrac{1}{lM} \end{bmatrix}$$

which can be easily verified to be controllable. Thus, a state feedback gain K that makes the matrix $\partial f(0)/\partial x + g(0)K$ stable exists. For example, using the **ITAE**-based prototype design gives

$$K = [0.7311 \quad 13.9671 \quad 20.3326 \quad 2.2018]$$

which places the eigenvalues of $\partial f(0)/\partial x + g(0)K$ at the following locations:

$$s = \{2(-0.4240 \pm 1.2630\text{i}), \quad 2(-0.626 \pm 0.4141\text{i})\}$$

To obtain the control law of the form (17), we need to solve the regulator equations associated with the above tracking problem which takes the following form:

$$\frac{\partial \mathbf{x}(v)}{\partial v} Sv = f(\mathbf{x}(v)) + g(\mathbf{x}(v))\,\mathbf{u}(v)$$
$$h(\mathbf{x}(v)) = v_1 \tag{40}$$

Due to the nonlinearity of the functions f and g, exact solution of (40) is unavailable. Reference [10] has used the Taylor series to approximately solve (40) up to third order. Here, we will try to find a three-layer neural network approximation for the solution of (40) with $\omega = 1.5$ in the compact set $\Gamma = \{v \in R^2 \mid \|v\| \leqslant 2\}$ up to a maximal error $\varepsilon_r = 10^{-3}$. To simplify matter, let us first note that Equation (40) can be partially solved as follows [10]

$$\mathbf{x}(v) = \begin{bmatrix} v_1 \\ \omega v_2 \\ \mathbf{x}_3(v) \\ \mathbf{x}_4(v) \end{bmatrix}$$
$$\mathbf{u}(v) = -(M + m\sin^2 \mathbf{x}_3(v))\,\omega^2 v_1 - ml\mathbf{x}_4^2 \sin \mathbf{x}_3(v) + mg\cos \mathbf{x}_3(v) \sin \mathbf{x}_3(v)$$

where $\mathbf{x}_3(v)$ and $\mathbf{x}_4(v)$ satisfy

$$\frac{\partial \mathbf{x}_3(v)}{\partial v} Sv = \mathbf{x}_4(v)$$
$$\frac{\partial \mathbf{x}_4(v)}{\partial v} Sv = \frac{\omega^2}{l} v_1 \cos \mathbf{x}_3(v) + \frac{g}{l} \sin \mathbf{x}_3(v) \tag{41}$$

Thus it suffices to define $J(W, v)$ as follows

$$J(W, v) = \left\| \frac{\partial \mathbf{x}_3(W, v)}{\partial v} Sv - \mathbf{x}_4(W, v) \right\|^2 + \left\| \frac{\partial \mathbf{x}_4(W, v)}{\partial v} Sv - \frac{\omega^2}{l} v_1 \cos \mathbf{x}_3(W, v) + \frac{g}{l} \sin \mathbf{x}_3(W, v) \right\|^2$$

where $\mathbf{x}_3(W, v)$ and $\mathbf{x}_4(W, v)$ are the neural network representations of $\mathbf{x}_3(v)$ and $\mathbf{x}_4(v)$, respectively.

Having defined $J(W, v)$, we can further define $Q(W)$ according to (32) with

$$\Gamma_d = \left\{ (\rho \sin\theta, \rho\cos\theta) \mid \rho = 0.2, 0.4, \ldots, 2, \theta = \frac{\pi}{15}, \frac{2\pi}{15} \ldots, 2\pi \right\}$$

After some iterations, it is found that, with the activation function given by $\phi(y) = (1 - \exp(-y))/(1 + \exp(-y))$, and $N = [20, 20]$, there exists a weight $\hat{W}$ such that $Q(\hat{W}) < 6.5 \times 10^{-4}$. Of course, the vector N is determined on the trial and error basis. We have actually tested three cases corresponding to $N = [10, 10]$, $[20, 20]$, and $[40, 40]$. The vector $N = [20, 20]$ is chosen since it leads to faster convergence than the other two cases.

Figure 2 depicts the error surface $J(\hat{W}, v)$ for $\|v\| \leqslant 2$. It can be seen that the approximation error as measured by $J(\hat{W}, v)$ is quite uniform throughout the whole region $\|v\| \leqslant 2$, and the maximal value of $J(\hat{W}, v)$ over $\|v\| \leqslant 2$ is well below 10^{-3}. Thus, the discretization scheme as manifested by (33) yields a successful solution for (28) and (29) with $\varepsilon_r = 10^{-3}$.

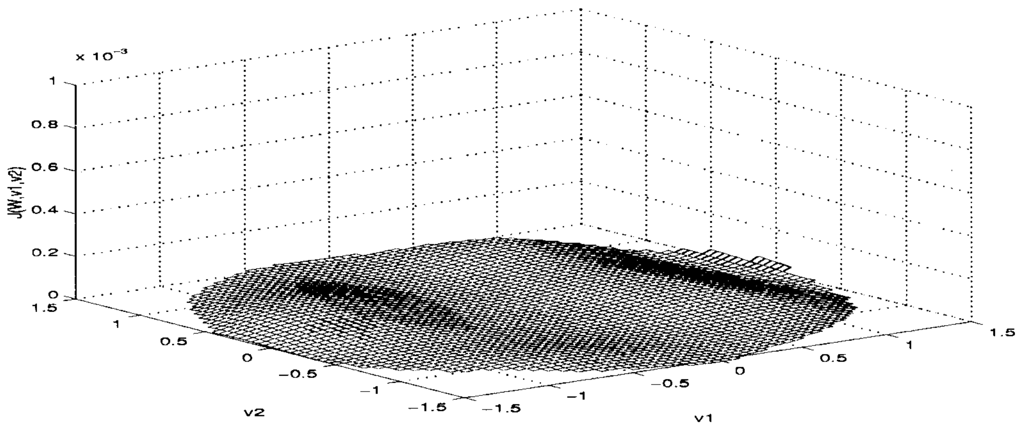

Figure 2. The error surface $J(W^*, v)$ for $\|v\| \leqslant 2$.

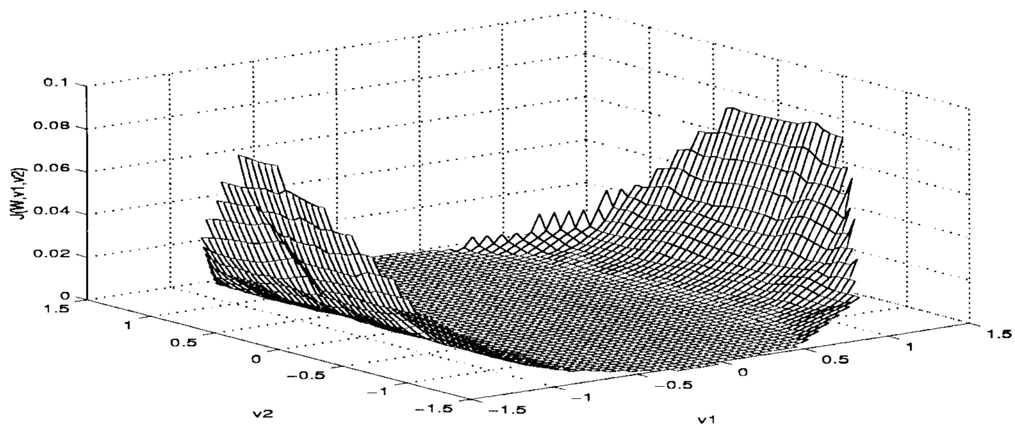

Figure 3. The error surface $J^1(v)$ for $\|v\| \leqslant 2$.

For comparison, we also show the error surfaces resulting from linear and third-order approximation of the solution of Equation (40) in Figures 3 and 4, respectively, where

$$J^i(v) = \left\| \frac{\partial \mathbf{x}_3^i(v)}{\partial v} Sv - \mathbf{x}_4^i(v) \right\|^2 + \left\| \frac{\partial \mathbf{x}_4^i(v)}{\partial v} Sv - \frac{\omega^2}{l} v_1 \cos \mathbf{x}_3^i(v) + \frac{g}{l} \sin \mathbf{x}_3^i(v) \right\|^2, \quad i = 1, 3$$

where $(\mathbf{x}_3^1(v), \mathbf{x}_4^1(v))$ and $(\mathbf{x}_3^3(v), \mathbf{x}_4^3(v))$ are the linear and third-order approximate solution of the Equation (41). It is not surprising to see that though the error resulting from the Taylor series based approach is small when v is small, it increases drastically as v does.

The performance of our control law as given by (17) has been evaluated by computer simulation with various values of the frequency ω and amplitude A. Table I lists the maximal steady-state tracking errors of the closed-loop system with $\omega = 1.5$ and $A = 0.5$, 1.0, 1.5, 2.0. Figures 5 and 6 show the time history of the tracking performance of the closed-loop system for

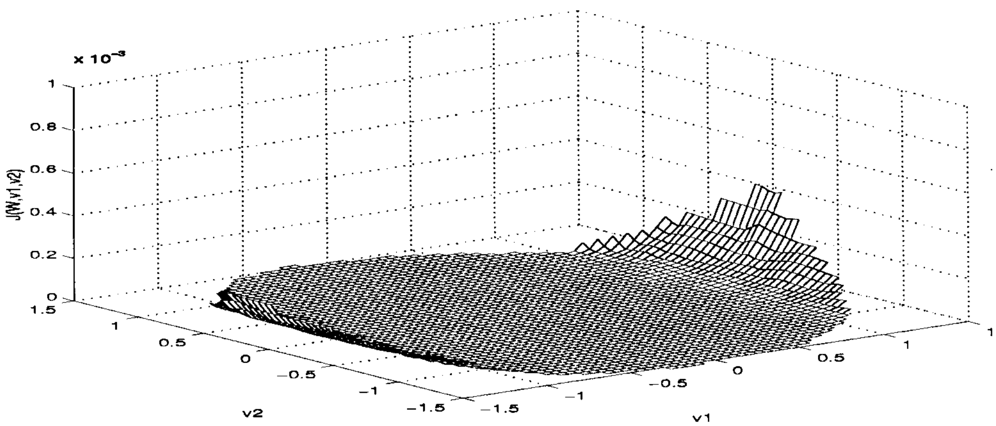

Figure 4. The error surface $J^3(v)$ for $\|v\| \leqslant 2$.

Table I. Maximal percentage steady-state tracking error with $\omega = 1.5$

Amplitude A	Neural (%)	Linear (%)	Third order (%)
0.5	0.28	1.45	0.10
1.0	1.70	6.15	0.23
1.5	0.95	14.3	0.83
2.0	1.11	25.9	2.42

the cases $A = 1.5$ and 2. For comparison, we have also presented the tracking performance of two other control laws of the following form:

$$u(t) = \mathbf{u}^i(v) + K(x - \mathbf{x}^i(v)), \quad i = 1, 3$$

where $(\mathbf{u}^1(v), \mathbf{x}^1(v))$ and $(\mathbf{x}^3(v), \mathbf{x}^3(v))$ are the linear and third-order approximate solution of the regulator equations. It is seen that the steady-state tracking errors resulting from both the linear and third-order control laws increase significantly as the amplitude of the sinusoidal signal does while the neural-network-based control law maintains a more uniform performance over different amplitudes of the the sinusoidal signal. It is clear that the neural network controller performs better than the Taylor-series-based controller when A is relatively large. These simulation results are consistent with the error surfaces shown in Figures 2–4.

It may be worth noting that it is possible to further improve the tracking performance of the Taylor-series-based control law by obtaining a higher-order approximation solution of the regulator equations. But the penalty on the computational complexity may be prohibitive. For example, a polynomial of degree k in two variables has as many as $k(k+3)/2$ coefficients. For a system with a moderate size such as the inverted pendulum on a cart system where $n = 4$, $m = 1$, and $q = 2$, the regulator equations may involve up to 5 unknown functions each of which depends on two variables v_1 and v_2. In the worse case, finding a fifth-order solution may involve solving

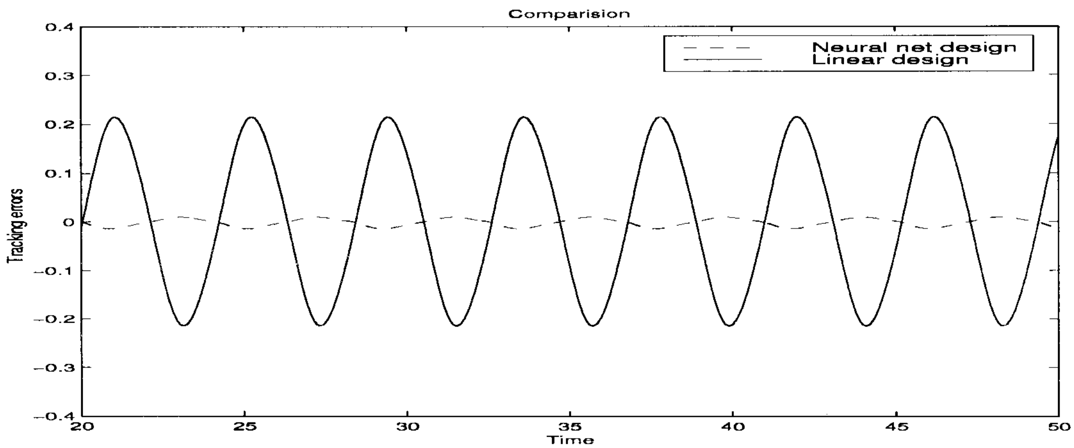

Figure 5. Results comparison with $A = 1.5$, $\omega = 1.5$.

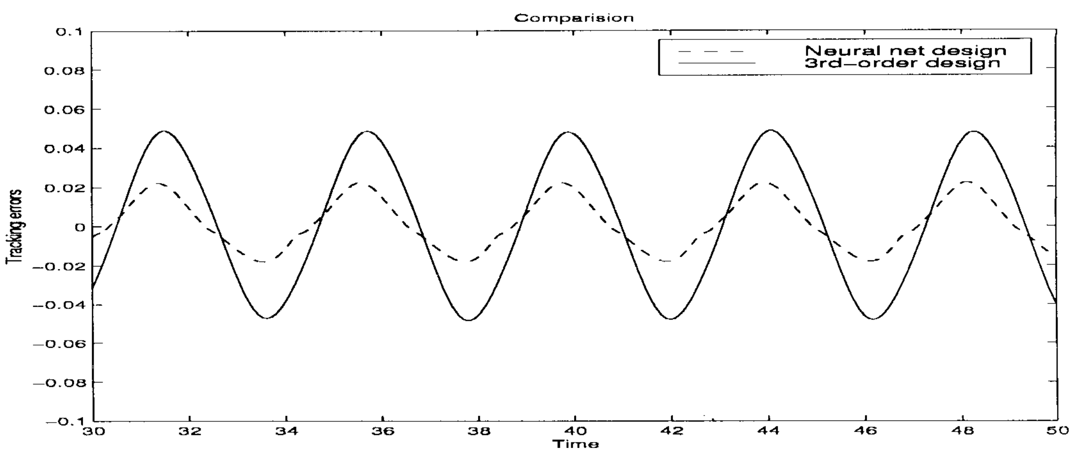

Figure 6. Results comparison with $A = 2.0$, $\omega = 1.5$.

$100(= 5 \times 20)$ linear equations not mentioning how to obtain these equations. On the other hand, it is also possible to further improve the tracking performance of the neural-network-based control law by increasing the number of hidden neurons or refining the discrete set Γ_d. Doing this will also substantially increase the computation time. But the algorithm to find a suitable weight remains almost the same since the calculation of the gradient of $Q(W)$ is almost invariant with respect to N or the way Γ_d is defined.

6. CONCLUSIONS

This paper has developed an approximation method to solve the problem of output regulation of nonlinear systems based on the three-layer neural networks. Effectiveness of our approach has been illustrated through a tracking problem associated with the inverted pendulum on a cart system. In comparison with the Taylor-series-based approach, this method offers some advantages described as follows:

- Good approximation of the solution of the regulator equations can be obtained in any compact set on which the exact solution is defined.
- As the approximation problem is converted into a parameter optimization problem, the wealth of the tools resulting from optimization can be freely exploited.
- Advantages such as numerical efficiency and hardware realization of the neural networks can be fully taken.

Finally, we note that nonlinear partial differential equations also arise in other nonlinear control problems such as the Frobenius equation from the feedback linearization, and the Hamilton–Jacobian–Issacs equation from the nonlinear H_∞ control problem. It is interesting to extend the approach proposed here to handle these two equations.

ACKNOWLEDGEMENT

The research was supported by the Hong Kong Research Grants Council under Grant CUHK4400/99E

REFERENCES

1. Davison EJ. The robust control of a servomechanism problem for linear time-invariant multivariable systems, *IEEE Transactions on Automatic Control* 1990; **35**(2):131–140.
2. Desoer CA, Wang YT. Linear time-invariant robust servomechanism problem: a self-contained exposition. *Control and Dynamic Systems* 1980; **16**:81–129.
3. Francis BA. The linear multivariable regulator problem. *SIAM Journal on Control and Optimization* 1977; **15**:486–505.
4. Francis BA, Murray Wonham W. The internal model principle of control theory. *Automatica* 1976; **12**:457–465.
5. Hepburn JSA, Wonham WM. Error feedback and internal models on differential manifolds. *IEEE Transactions on Automatic Control* 1984; **29**(5):397–403.
6. Huang J, Rugh WJ. On a nonlinear multivariable servomechanism problem. *Automatica* 1990; **26**(6):963–972.
7. Isidori A, Byrnes CI. Output regulation of nonlinear systems. *IEEE Transactions on Automatic Control* 1990; **35**(2):131–140.
8. Hornik K. Approximation capabilities of multilayer feedforward networks. *Neural Networks* 1991; **4**:251–257.
9. Hauser J, Sastry S, Kokotović P. Nonlinear control via approximate input–output linearization: The ball and beam example. *IEEE Transactions on Automatic Control* 1992; **37**(3):392–398.
10. Huang J. On a control problem of the inverted-pendulum on a cart system. *Proceedings of The* 4*th International Conference on Information Systems*, *Analysis and Synthesis*, Orlando, Florida, 1998; 51–56.
11. Huang J, Rugh WJ. An approximation method for the nonlinear servomechanism problem. *IEEE Transaction on Automatic Control* 1992; **37**(9):1395–1398.
12. Rudin W. Principles of mathematical analysis. (3rd edn.). *McGraw-Hill book Company*: New York, 1976.
13. Gurumoorthy R, Sanders SR. Controlling nonminimum phase nonlinear systems- the inverted pendulum on a cart example. *Proceedings of American Control Conference*, San Francisco, California, USA, 1993; 680–685.
14. Hepburn JSA, Wonham WM. Structural stable nonlinear regulation with step inputs. *Mathematical Systems Theory* 1984; **17**:319–333.
15. Huang J, Rugh WJ. Stabilization on zero-error manifolds and the nonlinear servomechanism problem. *IEEE Transactions on Automatic Control* 1992; **37**(7):1009–1013.
16. Krener AJ. The construction of optimal linear and nonlinear regulators. In *Systems*, *Models*, *and Feedback*, Isidori A, Tarn TJ (eds). Birkhauser: Basel, 1992; 301–322.

