On the Persistency of Excitation in Radial Basis Function Network Identification of Nonlinear Systems

Dimitry Gorinevsky, *Member, IEEE*

Abstract— **We consider radial basis function (RBF) network approximation of multivariate nonlinear mapping as a linear parametric regression problem. Linear recursive identification algorithms applied to this problem are known to converge, provided the regressor vector sequence has the persistency of excitation (PE) property. The main contribution of this paper is formulation and proof of PE conditions on the input variables. In the RBF network identification, the regressor vector is a nonlinear function of these input variables. According to the formulated condition, the inputs provide PE, if they belong to domains around the network node centers. For a two-input network with Gaussian RBF that have typical width and are centered on a regular mesh, these domains cover about 25% of the input domain volume. We further generalize the proposed solution of the standard RBF network identification problem and study affine RBF network identification that is important for affine nonlinear system control. For the affine RBF network, we formulate and prove a PE condition on both the system state parameters and control inputs.**

I. Introduction

APPROXIMATION of nonlinear mappings has recently attracted much attention of the control community in the context of nonlinear system identification. The approximation issues have been addressed by applying artificial neural network (ANN) approaches, as well as by fitting linearly parameterized nonlinearities to input/output data. This paper considers radial basis function (RBF) approximation methods that belong to the intersection of the two mentioned classes of approaches [4], [5], [27], [35].

It has recently been acknowledged that in many problems spatial filtering and approximation accuracy properties of RBF networks are advantageous as compared to other methods, including multilayered perceptron networks [5], [6], [16], [19], [22], [26], [35]. Even more important for many applications, the RBF networks provide approximation linear in the network weights. This feature makes powerful tools of the linear system theory applicable to the RBF network identification of nonlinear systems.

The RBF network approximation can be represented in the form of a linear parametric regression, as a product of a parameter matrix and a regressor vector. The parameter matrix comprises the network weights, and the regressor vector components are nonlinear radial functions of inputs. Linear identification algorithms are known to converge to the correct estimates of the regression parameter matrix, if the regressor vector sequence satisfies the persistency of excitation (PE) condition; e.g., see [11]. This condition is formulated mathematically in Section III. The physical interpretation of the PE condition is that the regressor vector sequence spans the full dimension of the input parameter space.

In the RBF network identification, unlike the linear identification, the regressor vector depends on the system inputs in a nonlinear way, and it is difficult to formulate the PE conditions on the network inputs. The PE conditions are very important for RBF networks application, as they provide the theoretical background for the analysis of the identification algorithm convergence. The only formulation of the PE condition for RBF network identification known to the author, however, is given in [36]. This condition is very restrictive, as it requires the inputs to coincide with the network node centers.

In this paper, we consider two RBF network architectures. The first is a standard RBF network, and the second one is an affine RBF network useful for approximating nonlinear controlled systems that are affine (linear) in the control components. We present and prove conditions on the input sequence to provide PE for both standard and affine RBF networks. These PE conditions are more practical than the condition of [11]. The results presented in this paper can be used as a basis for the convergence analysis of RBF network-based algorithms for identification and adaptive control of nonlinear systems.

The paper outline is as follows.

Section II gives a general formulation of recursive algorithms for RBF network identification. This section also presents the basics of the RBF network approximation.

Section III presents our results on the persistency of excitation for the standard RBF network identification.

In Section IV, we consider RBF network approximation of controlled dynamical systems. We distinguish between two types of inputs: state parameters and control inputs. The approximation is affine in the control inputs and nonlinear in the state parameters. We formulate the PE conditions that simultaneously include both types of inputs.

II. Radial Basis Function Approximation

Let us consider the problem of approximating a smooth nonlinear mapping $\Re^{N_p} \mapsto \Re^{N_y}$

$$Y = f(p); \quad Y \in \Re^{N_y}; \quad p \in \mathcal{P} \subset \Re^{N_p} \tag{1}$$

Manuscript received July 17, 1993; revised April 17, 1994. This work was in part supported by NSERC International Research Fellowship.

The author was with the Department of Mechanical Engineering, University of Toronto, Toronto, Ontario, Canada, and is currently with Measurex Devron Inc., North Vancouver, BC, Canada V7J 3S4.

IEEE Log Number 9409294.

where p is an input parameter vector, $\mathcal{P}$ is a compact domain, and Y is an output vector. We assume that a scattered (irregularly placed) set of N_t input–output pairs is available and call this set the training data set

$$\{Y^{(j)} = f(p^{(j)}),\ p^{(j)}\}, \qquad (j = 1, \cdots, N_t). \tag{2}$$

The problem is to find an approximation $\hat{Y} = \hat{f}(p)$ of the mapping (1) value for any argument $p \in \mathcal{P}$.

A. Exact RBF Interpolation

The most convenient way of representing an unknown nonlinear function is to present it as an expansion. The expansion is linear in parameters that are assumed to be unknown. Let us consider an approximation of the mapping (1) that has the form

$$\hat{Y} = \hat{f}(p) = \sum_{j=1}^{N_a} Z^{(j)} w_j(p) \tag{3}$$

where N_a is the order of the expansion, $w_j(p)$ are scalar expansion functions, and $Z^{(j)} \in \Re^{N_y}$ are the expansion parameters. The truncated Fourier series expansion, polynomial expansion, and B-spline expansion all have the form (3). In the ANN literature, approximations of the form (3) are known as functional link networks [4], [29], and we further call vectors $Z^{(j)}$ the network weight vectors.

Given the expansion functions $w_j(p)$ (3), a standard way to solve the scattered data approximation problem is to choose the parameter vectors $Z^{(j)}$ by fitting expansion (3) to the data (2) with a least error. In the special case $N_a = N_t$, when the number of the expansion weight vectors (3) coincides with the number of the training set data pairs (2), one can generally fit the training data exactly.

In this paper we consider an expansion of the form (3) with functions $w_j(\cdot)$ that depend on the radii $r_j = \|Q^{(j)} - p\|$, where $Q^{(j)} \in \Re^{N_p}$ are given vectors. Such expansion is known under the name of RBF approximation. RBF approximation has been used in computer graphics and experimental data processing applications (e.g., geophysical data) for two decades and has been demonstrated to provide for a high quality of approximation. One can find further details and references in [7]–[9], [23], [33], and [34]. The cited papers employ the method recently referred to as an exact RBF interpolation. This method uses the radial functions centered at each of the data points (2). In this method, the radial function centers are $Q^{(j)} = p^{(j)}$, and the expansion functions in (3) have the form

$$w_j(p) = h(p - p^{(j)}) \tag{4}$$

where $h(\cdot)$ is a radial function, i.e., $h(p - p^{(j)})$ depends on the radius $\|p - p^{(j)}\|$. Some most commonly used radial basis functions are

$$\begin{aligned} h(p) &= \exp(-\|p\|^2/d^2); \\ h(p) &= (1 + \|p\|^2/d^2)^{1/2}; \\ h(p) &= (1 + \|p\|^2/d^2)^{-1/2} \end{aligned} \tag{5}$$

where $\|\cdot\|$ denotes the Euclidean vector norm. The first radial function in (5) is Gaussian, and the last two are called Hardy multiquadrics and reverse multiquadrics, respectively [23]. Usually, the radial function width parameter d in (5) is chosen to be about an average distance between the neighboring node centers [2], [7], [8], [34].

Let us introduce the data matrix $\boldsymbol{Y}$ and the parameter matrix Θ built of vectors (2) and (3)

$$\begin{aligned} Y &= [Y^{(1)}, \cdots, Y^{(N_t)}] \in \Re^{N_y, N_t} \\ \Theta &= [Z^{(1)}, \cdots, Z^{(N_a)}] \in \Re^{N_y, N_a}. \end{aligned} \tag{6}$$

In the exact RBF interpolation we have $N_a = N_t$. By substituting (3) and (4) into (2), and using (6), we can write the condition of the exact fit of training data $\hat{Y}(p^{(j)}) = Y^{(j)}$ in the matrix form

$$Y = \Theta H, \quad H = \{h(p^{(i)} - p^{(j)})\}_{i,j=1}^{N_t} \in \Re^{N_t, N_t}. \tag{7}$$

The symmetrical matrix H in (7) is called interpolation matrix. This matrix has been proved to be invertible for the commonly used radial functions, if the vectors $p^{(j)}$ are distinct [25]. With (3), (6), and (7) we obtain the interpolation of the mapping (1) of the form

$$\begin{aligned} \hat{Y} &= \hat{f}(p) = \boldsymbol{Y} H^{-1} \overline{h}(p) \\ \overline{h}(p) &= \mathrm{col}(\{h(p - p^{(j)})\}_{j=1}^{N_t}) \in \Re^{N_t}. \end{aligned} \tag{8}$$

It has recently been acknowledged that RBF interpolation minimizes a certain regularization performance index that describes the interpolated surface roughness [7], [32], [33]. Different forms of radial functions (5) correspond to minimization of different regularization indexes.

Note that approximation (8) is linear in the data vectors $Y^{(j)}$ (6). Thus, the computational complexity of the method remains moderate even for a large dimension N_y of the vector Y.

The exact RBF interpolation (8) is global in the sense that it has the same form for any $p \in \mathcal{P}$. One needs to complete the most computationally expensive part of (8)—inversion of matrix H (7)—only once for any number of points, where approximation (8) of function (1) is to be computed. Yet, this advantage cannot be used if the training set (2) grows with time.

B. RBF Network Approximation

Recently, some authors have treated the RBF approximation in the connectionist network setting [1], [19], [24], [26], [30]–[32]. They consider the radial function centers $Q^{(j)}$ (we will call $Q^{(j)}$ the network node centers) that do not coincide with the training set points, so that the expansion functions in (3) have the form

$$w_j(p) = h(p - Q^{(j)}), \qquad (j = 1, \cdots, N_a). \tag{9}$$

Suppose that the node centers $Q^{(j)}$, $(j = 1, \cdots N_a)$ are given and fixed vectors. Let us fit the training set data (2) using network (3). Employing the same notation (6) as in (7), we can represent the fitting problem in the regression form

$$Y = \Theta\Phi + \mathcal{E}, \quad \Phi = \{h(p^{(k)} - Q^{(j)})\}_{j,k=1}^{N_a, N_t} \in \Re^{N_a, N_t} \tag{10}$$

where $\mathcal{E} = [e^{(1)}, \cdots, e^{(N_t)}] \in \Re^{N_y, N_t}$ is a residual error matrix. Since we cannot be sure that Φ is a well conditioned

or even a full rank matrix, we will look for a regularized least square solution to (10) that minimizes

$$\|\mathcal{E}\|_F^2 + \alpha\|\Theta\|_F^2 \rightarrow \min. \qquad 0 < \alpha \ll 1 \tag{11}$$

where $\|\cdot\|_F$ denotes a matrix norm equal to the square root of the sum of the squared matrix entries (the Frobenius norm). In (12), α is a scalar regularization parameter, introduced to obtain solution of possibly ill-conditioned problem following the regularization technique of [37]. The parameter α is small and does not influence the solution if the problem is well conditioned. Solving (10) and (11) for Θ gives

$$\Theta = \boldsymbol{Y}\Phi^T(\alpha I_{N_a} + \Phi\Phi^T)^{-1} \tag{12}$$

where I_{N_a} is the $N_a \times N_a$ identity matrix.

If $N_t \gg N_a$ and the training set inputs $p^{(j)}$ are uniformly distributed in the input domain, matrix Φ should have a rank N_a because basis functions (9) are linearly independent. The condition that Φ has the full rank is called the PE condition. The discussion of the PE condition is the main topic of this paper, and for the network (3), (9) we formulate the PE condition in Section III.

As the exact RBF interpolation is known to yield very accurate results, one can expect that an RBF network with fixed centers can provide good approximation accuracy. More discussion of the properties of the RBF network with nodes placed on a uniform grid can be found in [35]. The idea discussed in [35] and theoretically studied in more detail in [34] is that an RBF interpolation on a uniform grid performs a spatial filtering of the approximated function. Thus, the RBF approximation error is small, if the function has small high-frequency contents.

C. Recursive Identification of the RBF Model

Equation (12) describes the computation of the network parameter matrix Θ with the method that is called batch learning in the ANN literature. This method assumes that the whole training data set (2) is available at once. In many practical situations, however, the training data pairs arrive one by one, and a recursive weight updating procedure is desirable. The recursive weight update enables the user not to keep track of all upcoming data, but rather modify the parameter matrix Θ as the new data arrive. This feature is especially important for RBF network-based nonlinear adaptive control applications.

We can apply a well-known recursive least squares (RLS) equivalent of (12) to update an already available estimate of the matrix Θ (6). Let us introduce a regressor vector

$$\Phi(p) = \mathrm{col}(\{h(p - Q^{(j)})\}_{j=1}^{N_a}) \tag{13}$$

and denote $\Phi^{(k)} = \Phi(p^{(k)})$. The vectors $\Phi^{(k)}$ are the columns of the regression matrix Φ (10). Let $\hat{\Theta}^{(k)}$ be an estimate of Θ available at step k. The RLS identificator has the following standard form [11]

$$\begin{aligned}
\hat{\Theta}^{(k+1)} &= \hat{\Theta}^{(k)} + a^{(k)}e^{(k)}\Phi^{(k)^T}P^{(k)}/g^{(k)}.\\
e^{(k)} &= Y^{(k)} - \hat{\Theta}^{(k)}\Phi^{(k)}.\\
P^{(k+1)} &= P^{(k)} - a^{(k)}P^{(k)}\Phi^{(k)}\Phi^{(k)^T}P^{(k)}/g^{(k)}.\\
g^{(k)} &= (1 + a^{(k)}\Phi^{(k)^T}P^{(k)}\Phi^{(k)})^{-1}
\end{aligned} \tag{14}$$

where, as usual, $e^{(k)}$ is the kth step prediction error, $P^{(k)}$ is the inverse covariance matrix, and $g^{(k)}$ is an auxiliary variable. For $a^{(k)} = 1$ and the initial condition $P^{(0)} = \alpha^{-1}I_{N_a}$, (14) is equivalent to (12). The deadzone parameter $a^{(k)}$ can be introduced in (14) in the usual way to provide robustness with respect to the imprecise modeling and measurement noise; e.g., see [11] for more detail. The dimension of the covariance matrix $P^{(k)}$ is $N_a \times N_a$. This dimension does not depend on the dimension N_y of the output vector Y. Since there are no matrix inversions in the RLS update (14), it is feasible for a relatively large network size N_a.

Papers [4] and [5] consider an application of the orthogonal least square modification of the RLS algorithm to RBF network approximation. Papers [5] and [22] consider modifications of the RLS identification of an RBF network for the cases of dynamical node creation, update, and clustering of the node centers. In this paper, we consider the numbers of the nodes and the node centers as fixed parameters.

Another recursive estimation technique commonly used in signal processing and adaptive control is projection estimation. The projection estimation is a special case of the least mean square algorithm, which is known as the Widrow–Hoff updating rule in the signal processing literature and as a delta rule in the ANN literature. To derive the projection update, instead of minimizing a mean error index (11), let us minimize a one-step error increment index similar to (11)

$$\begin{aligned}
\|e^{(k)}\|^2 + \alpha\|\hat{\Theta}^{(k+1)} - \hat{\Theta}^{(k)}\|_F^2 &\rightarrow \min,\\
e^{(k)} &= Y^{(k)} - \hat{\Theta}^{(k)}\Phi^{(k)}
\end{aligned} \tag{15}$$

where $e^{(k)}$ is the kth step approximation error. The solution of (15) for $\hat{\Theta}^{(k)}$ has the form similar to (12)

$$\hat{\Theta}^{(k+1)} = \hat{\Theta}^{(k)} + e^{(k)}\Phi^{(k)^T}/(\alpha + \|\Phi^{(k)}\|^2) \tag{16}$$

where $\|\Phi^{(k)}\|^2 = \Phi^{(k)^T}\Phi^{(k)}$.

Projection estimation of RBF network weights in adaptive control of a nonlinear system is considered, for instance, in [27].

III. Persistency of Excitation

This section, as well as Section IV, contains the major contribution of this paper.

The well-known convergence condition for the estimation algorithms (16) and (14) is the PE condition, i.e., the requirement that for some $\delta > 0, K_e \geq N_a$ and any $n > 0$

$$\underline{\sigma}\left(\sum_{k=n}^{n+K_e}\Phi^{(k)}\Phi^{(k)^T}\right) > \delta \tag{17}$$

where $\underline{\sigma}(A)$ denotes a minimal singular value of the matrix A [11].

In accordance with (13), the PE condition depends on the training inputs $p^{(j)}$. Authors of [36] have established that the inputs consecutively coinciding with the node centers $Q^{(j)}$ provide PE. The PE condition of [36] immediately follows from the invertibility of the exact RBF interpolation matrix H (7). Condition [36] is very restrictive, since for a random input

sequence $p^{(j)}$, which clearly should provide PE, the network inputs never coincide with the node centers.

We keep assuming that the node centers $Q^{(j)}$ of the RBF network are fixed and prove that the sequence $\Phi(p^{(j)})$ provides PE if the inputs $p^{(j)}$ belong to certain neighborhoods of the node centers $Q^{(j)}$. We show numerically that, in fact, these neighborhoods may occupy a significant part of the input domain volume, and thus, the derived PE condition is not too restrictive.

Definition 1: Let us consider a finite sequence of K_e RBF network inputs $p^{(j)}$ and call it exciting, if for some $\delta > 0$ the following inequality holds

$$\underline{\sigma}(\Psi_{K_e}) = \underline{\sigma}\left(\sum_{k=1}^{K_e} \Phi^{(k)}\Phi^{(k)^T}\right) > \delta. \tag{18}$$

Clearly, an infinite sequence $\{p^{(j)}\}_{j=1}^{\infty}$ has a PE property, if for some K_e all finite sequence pieces of the length K_e contain exciting subsequences of length N_a with the same δ.

To establish the PE condition, let us use the cardinal interpolation representation of (10) and (18). Let us consider the interpolation matrix H of the form (7) and a new regressor vector $X(p)$ of the form

$$X(p) = \begin{bmatrix} \chi_1(p) \\ \vdots \\ \chi_{N_a}(p) \end{bmatrix} = H^{-1}\Phi(p) \in \Re^{N_a}$$
$$H = \{h(Q^{(i)} - Q^{(j)})\}_{i,j=1}^{N_a} \in \Re^{N_a, N_a}. \tag{19}$$

The interpolation matrix H is proved to be invertible for the commonly used radial functions [25]. The functions $\chi_j(p)$ (19) are called cardinal functions [34] and have the property

$$\chi_j(Q^{(k)}) = \delta_{jk} \tag{20}$$

where δ_{jk} is the Kronecker symbol. The regressor vector $X(p)$ (19) is the linear transformation of the vector $\Phi(p)$ (13), and by substituting (19) into (18), we obtain the excitation condition (18) in the form

$$\underline{\sigma}(\Psi_{N_a}) = \underline{\sigma}\left(\sum_{k=1}^{N_a} X^{(k)}X^{(k)^T}\right) > \varepsilon_1 \tag{21}$$

where $X^{(k)} = X(p^{(k)})$. For $\varepsilon_1 = \delta/\underline{\sigma}^2(H) > 0$, inequality (18) follows from (21). Inequality (21) easily proves the PE condition of [36]. In this condition, $p^{(j)} = Q^{(j)}$ and, according to (20), $\Psi_{N_a} = I_{N_a}$, an identity matrix. Thus, (21) holds with $\delta = 1$.

Since the cardinal functions (19) are continuous and satisfy (20), the node centers $Q^{(j)}$ are inner points of the diagonal dominance sets $\mathcal{A}_j^{\varepsilon}$ $(0 < \varepsilon)$

$$\mathcal{A}_j^{\varepsilon} = \left\{ p \in \mathcal{A}_j^{\varepsilon} : |\chi_j(p)| - \sum_{i=1, i\neq j}^{N_a} |\chi_i(p)| > \varepsilon \right\}. \tag{22}$$

Now we are in position to formulate the following theorem

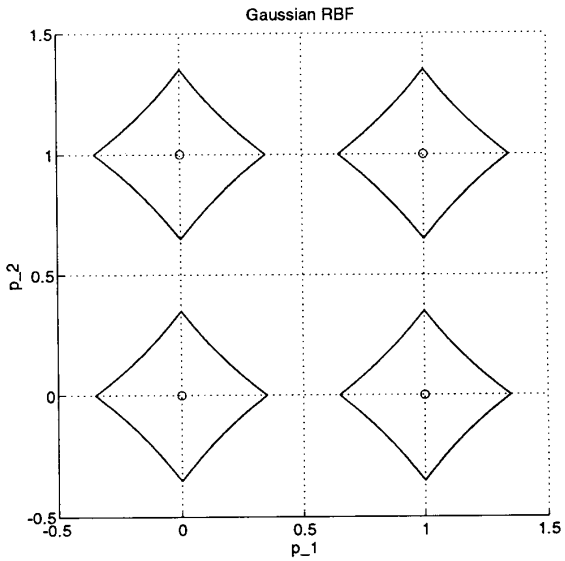


Fig. 1. View of the diagonal dominance domains $\mathcal{A}_j^0$ (22) for Gaussian RBF approximation, $d = 1$. RBF centers are at the points with whole coordinates and within the closed curves the cardinal functions (19) for the node exceed the absolute value sum of all other cardinal functions.

Theorem 1: Let us consider an infinite sequence of the RBF network inputs $\{p^{(j)}\}_{j=1}^{\infty}$. Assume that $K_e \geq N_a$ and $\varepsilon > 0$ exist, such that for any k and node number $j, 1 < j \leq N_a$ we can find an input vector $p^{(i)}, (k < i \leq k + K_e)$, such that $p^{(i)} \in \mathcal{A}_j^{\varepsilon}$.

Then the sequence $\{p^{(j)}\}_{j=1}^{\infty}$ provides PE in the sense of (17).

The proof of Theorem 1 is given in the Appendix.

Practical usefulness of Theorem 1 depends on how large the diagonal dominance sets (22) of the cardinal functions are. The shape and sizes of the sets $\mathcal{A}_j^{\varepsilon}$ depend on the radial basis functions used and on the node center positions. To get an idea on how the domains (22) look like, let us consider an important special case of the network node centers $Q^{(j)}$ placed on a uniform mesh with a unit base.

For the presentation clarity, let us consider a two-dimensional input parameter space, $N_p = 2$. We compute the borders of the domain (22) for $\varepsilon \to 0$ numerically. The cardinal functions $\chi_j(p)$ decay fast and are small far from the respective node center. Thus, remote nodes contribute little to the diagonal dominance condition (22). Therefore, in the middle of the node grid 11×11, that we use for computations in the example, shape of the domains (22) is almost the same as for the cardinal interpolation on an infinite grid of nodes.

The numerically found borders of the diagonal dominance domains $\mathcal{A}_j^0$ are plotted in Figs. 1–3. In all three figures, the RBF network nodes are at the points with whole coordinates—the centers of the domains $\mathcal{A}_j^0$. Fig. 1 shows the shapes of the domains $\mathcal{A}_j^0$ for Gaussian RBF network with $d = 0.7$. The domains have symmetrical diamond shapes. Figs. 2 and 3 show the domain borders for the Gaussian and reverse multiquadrics radial functions (5). In Figs. 2 and 3, the node center is in the coordinate origin, and the first quadrant defines

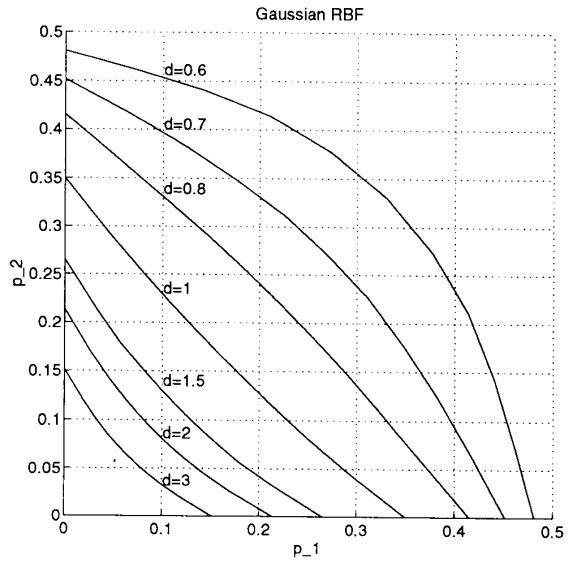


Fig. 2. Diagonal dominance domain borders for Gaussian RBF and different Gaussian widths d. Border of only one quadrant respective to the RBF center is shown.

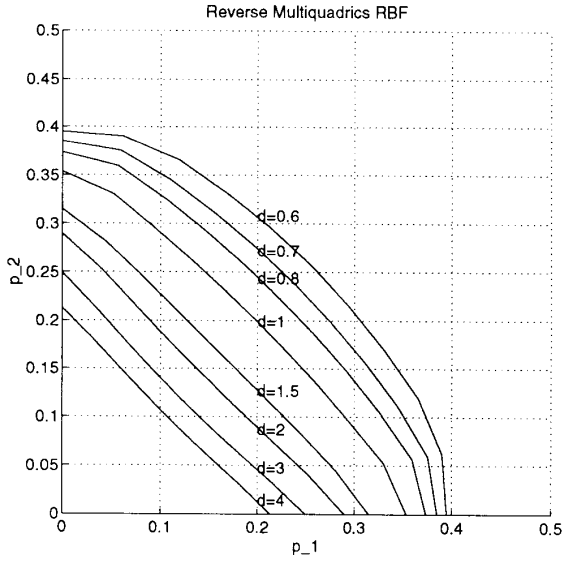


Fig. 3. Diagonal dominance domain borders for reverse multiquadrics RBF and different widths d of the multiquadrics. The border of only one quadrant respective to the RBF center is shown.

the shape of the symmetrical domain $\mathcal{A}_j^0$. The border depends on the radial function width parameter d, which is displayed near the curve. For smaller values of d, the radial functions (5) are narrower and their values at the neighboring nodes smaller. Thus, each radial function dominates other radial functions in a larger neighborhood of its own center.

The borders in Figs. 2 and 3 are plotted for the widths d in the range normally used in practice. We do not discuss an issue of optimal choice of d here. Note also that Theorem 1 actually gives only a sufficient condition of the persistent excitation. Thus, the width d should be chosen to provide the best accuracy of approximation rather than a maximal size of the diagonal dominance domain.

Optimal choice of d for the problem of scattered data RBF approximation of manipulator dynamics is studied in [2]. By normalizing the average distance between the node centers in the results of [2] to unit, one can derive that the best accuracy is obtained with $d \approx 2$ for Gaussian RBF and $d \approx 3$–6 for reverse multiquadrics. Yet, these results strongly depend on the scaling of the input data. Our experience suggests $d \approx 1.5$ for Gaussian RBF.

One can see that the diagonal dominance domains in Fig. 1 cover about 25% of the plane area. Thus, for randomly generated inputs $p^{(j)}$, about 25% of the inputs will add to the persistent excitation as guaranteed by Theorem 1. This guarantees fast convergence of the considered recursive algorithms for the RBF network identification.

From the results of Figs. 2 and 3, some estimates can be made regarding the size of the diagonal dominance domains for the dimension of the input vector p higher than two.[1] Let us assume that the domain $\mathcal{A}_j^0$ could be approximated by hypercube that has unit diagonals and the side $\sqrt{2}/2$, and occupies $(\sqrt{2}/2)^{N_p}$ of the unit hypercube volume. This makes 50% for $N_p = 2$, 25% for $N_p = 4$, and 12.5% for $N_p = 6$. RBF approximation is rarely feasible for larger input dimension N_p, as the number of the network nodes grows with N_p exponentially. Still this estimate shows that it becomes increasingly more difficult to guarantee persistency of the excitation for large dimension N_p of the input space.

IV. Identification and Persistency of Excitation in Affine RBF Network

Let us now consider a mapping that describes a nonlinear system and has the form

$$Y = S(U, p) = f(p) + G(p)U$$
$$Y \in \Re^{N_y}: \quad U \in \Re^{N_u}: \quad p \in \mathcal{P} \subset \Re^{N_p} \tag{23}$$

where vector p has the meaning of the system parameters (state variables) and U is the control vector. Affine nonlinear dynamical systems, for which the right-hand side of the governing equations has the form (23), have attracted much attention in modern nonlinear control theory [21]. For such systems, in (23) p is the state-space vector, vector Y is the state derivative, and U is a control input. Affine mappings of the same form are considered in a number of papers on ANN-based adaptive control, e.g., [15], [27], [28], and [35] among many others. Nonlinear mappings of the form (23) are used in [12]–[14], [17], and [18] to describe open-loop feedforward control of dynamical systems, where Y is the task-related system output, U describes the open-loop control program over a certain time interval, and p is the task parameter vector.

By applying the RBF network approximation of the form (3), (9) to the vector-valued mapping $f(\cdot)$ and the matrix-valued mapping $G(\cdot)$ and combining them to approximate the

[1] The author owes this remark to an anonymous reviewer.

system (23), we obtain an affine RBF network of the form

$$Y = \hat{S}(U,p) = \hat{f}(p) + \hat{G}(p)U = \sum_{j=1}^{N_a} h(p - Q^{(j)})[Z^{(j)} + G^{(j)}U] \quad (24)$$

where $Z^{(j)}$ and $G^{(j)}$ are the parameter matrixes of appropriate dimensions.

As in Section III, we assume that the network node centers $Q^{(j)}$ are fixed parameters of the approximation (24). By introducing the regressor vector $\overline{\Phi}(p,U)$ and abusing Section II notations for the network parameter matrix Θ and the weights $Z^{(j)}$, we can write (24) in the regression form

$$Y = \hat{S}(U,p) = \Theta\overline{\Phi}(p,U)$$
$$\Theta = [Z^{(1)}\ G^{(1)} \cdots Z^{(N_a)}\ G^{(N_a)}]. \quad (25)$$

In accordance with (24), the regressor vector in (25) has the form

$$\overline{\Phi}(p,U) = \Phi(p) \otimes W, \quad W = \begin{bmatrix} 1 \\ U \end{bmatrix} \quad (26)$$

where $\otimes$ denotes direct (Kronecker) matrix product and $\Phi(p)$ is given by (13). The regression representation (25) of the network (24) has the same form as the regression representation (10) of the network (3). Therefore, we can apply one of the recursive identification algorithms discussed in Section III for estimating Θ. By using the notation of (10), we can write the estimation algorithms in the same form as presented in Section III.

For clarity, let us consider an estimation algorithm discussed in Section III—the projection estimation—in more detail here. For network (25), the projection algorithm can be written in the form similar to (16)

$$\hat{\Theta}^{(k+1)} = \hat{\Theta}^{(k)} + e^{(k)}\overline{\Phi}^{(k)^T}/(\alpha + \|\overline{\Phi}^{(k)}\|^2)$$
$$e^{(k)} = Y^{(k)} - \hat{\Theta}^{(k)}\overline{\Phi}^{(k)} \quad (27)$$

where $\hat{\Theta}^{(k)}$ is an estimate of the network parameter (weight) matrix Θ (25) at the iteration k $\overline{\Phi}^{(k)} = \overline{\Phi}(p^{(k)}, U^{(k)})$ is the regressor vector (26) at the same iteration and $p^{(k)}, U^{(k)}$, and $Y^{(k)}$ are, respectively, the task parameter, input, and output vector at the iteration k.

Estimation algorithm (27) converges, if the PE condition of the form (17) is satisfied. We follow the outline of Section III and prove a sufficient condition for all finite pieces of the sequence $\overline{\Phi}^{(k)}$ of length $K_e \geq N_u(N_a + 1)$ to be exciting in the sense similar to (18).

Unlike the standard RBF network considered in Section III, in the affine RBF network (25) the regressor vector $\overline{\Phi}^{(k)} = \overline{\Phi}(p^{(k)}, U^{(k)})$ depends on two input vectors: $p^{(k)}$ and $U^{(k)}$. Thus, the PE condition of this section is more involved than that of Section III.

Let us now formulate a generalization of the Theorem 1 for the affine network identification.

Theorem 2: Let us consider an infinite sequence of the affine RBF network inputs $\{p^{(j)}, U^{(j)}\}_{j=1}^{\infty}$. Assume that $K_e \geq N_a(N_u + 1)$ and $\varepsilon > 0$ exist, such that for any k and any node number $j, 1 \leq j \leq N_a$ we can find $N_u + 1$ different input pairs $\{p^{(l_i)}, U^{(l_i)}\}_{i=1}^{N_u+1}, (k < l_i \leq k + K_e)$ for which the following inequality holds

$$|\chi_j(p^{(l_i)})|/\mathrm{cond}(\overline{W}) - \sum_{k=1, k\neq j}^{N_a} |\chi_k(p^{(l_i)})| > \varepsilon. \quad (28)$$

Here $\chi_j(p)$ are the cardinal functions (19) of the RBF interpolation, and the square matrix $\overline{W} \in \Re^{N_u+1, N_u+1}$ has the columns

$$W^{(i)} = \begin{bmatrix} 1 \\ U^{(l_i)} \end{bmatrix}. \quad (29)$$

We also assume that the matrix $\overline{W}$ satisfies the inequality

$$\underline{\sigma}(\overline{W}) > \sigma_0 \quad (30)$$

where σ_0 is a given constant, and $\underline{\sigma}(\cdot)$ denotes the minimal singular value of the matrix. Then the sequence $\overline{\Phi}^{(k)} = \overline{\Phi}(p^{(k)}, U^{(k)})$ provides PE.

The proof of Theorem 2 is given in the Appendix.

Remark 2: For $N_u = 0$, $\mathrm{cond}(\overline{W}) = 1$ in (28), and we obtain the condition of the Theorem 1.

Remark 3: Condition (28) describes the diagonal dominance set that is smaller than the set (22) to the degree depending on the excitation properties of the control vector sequence $U^{(j)}$.

V. Conclusions

We have considered linear algorithms for RBF network identification of nonlinear systems. The algorithms converge, if the regressor vectors (13) or (26) satisfy the PE condition. The regressor vector is a nonlinear function of the input variables. We have proved that PE is achieved, if the input variables belong to certain neighborhoods of the RBF network nodes. By means of a numerical example, we have further demonstrated that these neighborhoods are sufficiently large.

We have further considered affine RBF network identification. Such a network has two different types of inputs. The system output depends in a nonlinear way on the input vector p of the first type and linearly on the input vector U of the second type. Affine RBF networks approximate an important class of nonlinear systems that are affine in control components. For such systems, the vector p represents the system state variables, and the vector U comprises the control inputs.

The formulated conditions represent the necessary basis for convergence analysis of the RBF network-based identification and adaptive control algorithms for nonlinear dynamical and controlled systems. In this paper, we have formulated the PE conditions for the RBF network identification of static mappings. The derived conditions can be used in analysis of RBF network identification of dynamical systems. This usage, however, requires an additional study. Clearly, the particular application of the derived PE conditions will depend on the form of the dynamical system representation via a

static mapping, such as the form of difference or differential governing equation used in the analysis.

APPENDIX

Proof of Theorem 1: The proof is by a contradiction.

It is sufficient to prove that the finite sequences $\{p^{(j+k)}\}_{j=1}^{K_e}$ satisfy (21) for some ε_1 and any k. Let us suppose that this is not true and that $\underline{\sigma}(\Psi_{K_e}) < \varepsilon_1$ for any fixed ε_1 and some k. Then for some $x \in \Re^{N_a}$, $\|x\| = 1, x = \mathrm{col}(\{x_j\}_{j=1}^{N_a})$. we have $x^T \Psi_{K_e} x < \varepsilon_1$. Therefore, for any j. $k < j \le k + K_e$ we obtain

$$\|X(p^{(j)})x\| \le \sqrt{\varepsilon_1}. \tag{31}$$

Without a loss of generality we can assume that the first component of the vector x is one with the largest absolute value

$$|x_1| \ge |x_j|, \ (j = 2, \cdots, N_a): \text{ and } |x_1| \ge 1/\sqrt{N_a}. \tag{32}$$

The second inequality in (32) follows from the first one and the fact that $\|x\| = 1$. By the theorem condition, for some i. $(k < i \le k + K_e)$ we have $p^{(i)} \in \mathcal{A}_1^{\varepsilon}$. Combining (22) and (32) we obtain

$$|\chi_1(p^{(i)})x_1| \ge \sum_{l=2}^{N_a} |\chi_l(p^{(i)})x_l| + \varepsilon/\sqrt{N_a}. \tag{33}$$

On the other hand, we can write (31) in the form

$$\sqrt{\varepsilon_1} > \sum_{l=1}^{N_a} |\chi_l(p^{(i)})x_l| \ge \left| |\chi_1(p^{(i)})x_1| - |\sum_{l=2}^{N_a} \chi_l(p^{(i)})x_l| \right| \ge |\chi_1(p^{(i)})x_1| - \sum_{l=2}^{N_a} |\chi_l(p^{(i)})x_l|. \tag{34}$$

By taking $\varepsilon_1 = \varepsilon^2/N_a$. we come to a contradiction between (33) and (34). □

Proof of Theorem 2: The proof follows the scheme of the proof of Theorem 1, but is somewhat more involved.

Let us assume that the conditions of Theorem 2 hold, but for any $\delta > 0$ some k exists so that

$$\underline{\sigma}(\overline{\Psi}_{K_e}) = \underline{\sigma}\left(\sum_{j=k+1}^{j=k+K_e} \overline{\Phi}(p^{(j)}, U^{(j)})\overline{\Phi}^T(p^{(j)}, U^{(j)})\right) < \delta. \tag{35}$$

Inequality (35) implies that a unit vector $\overline{z} \in \Re^{N_a(N_u+1)}$ exists such that

$$\overline{z}^T \overline{\Phi}(p^{(j)}, U^{(j)}) < \sqrt{\delta}, \qquad (j = k+1, \cdots, k+K_e). \tag{36}$$

Let us now introduce an extended interpolation matrix $\overline{H} \in \Re^{N_a(N_u+1), N_a(N_u+1)}$ and the new regressor vector $\overline{X}(p, U)$ of the form

$$\overline{H} = H \otimes I_{N_u+1}, \quad \overline{X}(p, U) = \overline{H}^{-1}\overline{\Phi}(p, U) = X(p) \otimes W \tag{37}$$

where I_{N_u+1} is the unit matrix of dimension $N_u + 1$, $X(p)$ is a vector of the form (19), $\overline{\Phi}(p, U)$ and W have the form (26), $\otimes$ is the Kronecker product, and the last equality in (37) follows from the block matrix properties [10].

By substituting (37) into (36), we obtain that for a unit vector $\overline{x} = \overline{H}\overline{z}/\|\overline{H}\overline{z}\|$ and some k the following inequality holds

$$\overline{x}^T \overline{X}(p^{(j)}, U^{(j)}) < \varepsilon_1. \qquad (j = k+1, \cdots, k+K_e) \tag{38}$$

where $\varepsilon_1 = \sqrt{\delta}/\|\overline{H}\overline{z}\| \le \sqrt{\delta}/\underline{\sigma}(\overline{H})$. Let us cut vector $\overline{x}$ into N_a vectors $x^{(j)}$ of length $N_u + 1$ each

$$\overline{x} = [{x^{(1)}}^T \cdots {x^{(N_a)}}^T]^T. \tag{39}$$

Without a loss of generality we can assume that

$$\|x^{(1)}\| \ge \|x^{(j)}\|. \qquad (j = 2, \cdots, N_a)$$
$$\|x^{(1)}\| \ge 1/\sqrt{N_a} \tag{40}$$

where the second inequality follows from the first one and the fact that $\|\overline{x}\| = 1$.

Now let us take $N_u + 1$ vectors $W^{(l_i)}$ so that the Theorem 2 condition (28) is satisfied for the first node, $j = 1$. Again, without a loss of generality we assume that $l_i = k + i$. By writing inequality (38) for each of these vectors and computing a linear combination of the inequalities with weights s_j, we obtain

$$\left| \sum_{j=1}^{N_u+1} \overline{x}^T \overline{X}(p^{(j+k)}, U^{(j+k)}) s_j \right| \le \varepsilon_1 \sum_{j=1}^{N_u+1} |s_j|. \tag{41}$$

By denoting $\chi_i^{(j)} = \chi_i(p^{(j+k)})$ and using (37), we can write (41) in the form

$$\left| \sum_{j=1}^{N_u+1} s_j {W^{(j)}}^T x^{(1)} \chi_1^{(j)} + \sum_{j=1}^{N_u+1} s_j {W^{(j)}}^T \sum_{i=2}^{N_a} x^{(i)} \chi_i^{(j)} \right| \le \varepsilon_1 \sum_{j=1}^{N_u+1} |s_j|. \tag{42}$$

We are considering a finite number of the network nodes, and the radial functions (5) have infinite support. Therefore, we can always find constants C and C_1 so that in the domain (22)

$$0 < C \le \chi_j(p) \le C_1 \quad \text{for} \quad p \in \mathcal{A}_j^0. \tag{43}$$

In (28) $\mathrm{cond}(\overline{W}) \ge 1$. thus $p^{(l_i)} \in \mathcal{A}_1^0$, and by using (43) we obtain from (28)

$$\sum_{i=2}^{N_a} |\chi_i^{(j)})|/|\chi_1^{(j)}| \le 1/\mathrm{cond}(\overline{W}) - \varepsilon/C_1. \tag{44}$$

Let us also denote $s = \mathrm{col}(\{s_j\}_{j=1}^{N_u+1}) \in \Re^{N_u+1}$. If we divide (42) by $|\chi_1^{(j)}|$ and use (30), (40), (43), and (44), we obtain the following chain of the inequalities

$$\frac{\varepsilon_1}{C}\|s\|\sqrt{N_u+1} \ge \varepsilon_1 \sum_{j=1}^{N_u+1} |s_j| \ge \left| \|\overline{W}s\| \cdot \|x^{(1)}\| - \|\overline{W}s\| \left(\frac{1}{\mathrm{cond}(\overline{W})} - \frac{\varepsilon}{C_1} \right) \|x^{(1)}\| \right| \ge \|s\|\underline{\sigma}(\overline{W})\|x^{(1)}\|\varepsilon/C_1 \ge \|s\|\sigma_0\varepsilon/(C_1\sqrt{N_a}). \tag{45}$$

Since δ in (28) and ε_1 in (38) can be taken arbitrary small, and ε in the Theorem 2 condition (28) is fixed, by choosing $\varepsilon_1 < C\sigma_0/(C_1\sqrt{N_a(N_u+1)})$ we come to the contradiction. □

REFERENCES

[1] C. Bishop, "Improving the generalization properties of radial basis function neural networks," *Neural Computa.*, vol. 3, pp. 579–88, 1991.

[2] M. Botros and C. G. Atkeson, "Generalization properties of radial basis functions," in *Advances in Neural Information Processing Systems 3*, R. P. Lipmann, J. E. Moody, and D. S. Touretzky, Eds. San Mateo, CA: Morgan Kaufmann, vol. 3, pp. 707–713.

[3] D. S. Broomhead and D. Lowe, "Multivariable functional interpolation and adaptive networks," *Complex Syst.*, no. 2, pp. 321–355, 1988.

[4] S. Chen and S. A. Billings, "Neural networks for nonlinear dynamic system modeling and identifications," *Int. J. Contr.*, vol. 56, no. 2, pp. 319–346, 1992.

[5] S. Chen, S. A. Billings, and P. M. Grant, "Recursive hybrid algorithm for nonlinear systems identification using radial basis function networks," *Int. J. Contr.*, vol. 55, no. 5, pp. 1051–1070, 1992.

[6] S. Chen, C. F. N. Cowan, and P. M. Grant, "Ortogonal least squares learning algorithm for radial basis function networks," *IEEE Trans. Neural Networks*, vol. 2, no. 2, pp. 837–863, 1991.

[7] N. Dyn, "Interpolation of scattered data by radial functions," in *Topics in Multivariable Approximation*, L. L. Schumaker, C. K. Chui, and F. I. Utreras, F. I., Eds. Boston: Academic, 1987, pp. 47–61.

[8] R. Franke, "Scattered data interpolation: Test of some methods," *Math. Computa.*, vol. 38, no. 157, pp. 181–200, 1982.

[9] ——, "Recent advances in the approximation of surfaces from scattered data," in *Topics in Multivariable Approximation*, L. L. Schumaker, C. K. Chui, and F. I. Utreras, Eds. Boston: Academic, 1987, pp. 79–98.

[10] F. R. Gantmacher, *The Theory of Matrices*. New York: Chelsea, 1959.

[11] G. C. Goodwin and K. S. Sin, *Adaptive Filtering, Prediction, and Control*. Englewood Cliffs, NJ: Prentice-Hall, 1984.

[12] D. M. Gorinevsky, "Experiments in direct learning of feedforward control for manipulator path tracking," *Robotersysteme*, vol. 8, pp. 139–147, 1992.

[13] ——, "Modeling of direct motor program learning in fast human arm motions," *Biol. Cybern.*, vol. 69, pp. 219–228, 1993.

[14] ——, "Adaptive learning of feedforward control using radial basis function network approximation," in *Proc. 1993 IEEE Int. Symp. Intell. Contr.*, Chicago, IL, Aug. 1993, pp. 762–767.

[15] ——, "Multirate indirect adaptive control of a bioreactor using a novel radial basis function network architecture," in *Intelligent Control Systems*, C. W. de Silva and R. Shoureshi, Eds. ASME, DSC-vol. 48, 1993, pp. 67–74

[16] D. M. Gorinevsky and T. H. Connolly, "Comparison of some neural network and scattered data approximations: The inverse manipulator kinematics example," *Neural Computa.*, vol. 6, no. 3, pp. 519–540, 1994.

[17] D. M. Gorinevsky, A. Kapitanovsky, and A. A. Goldenberg, "Network architecture and controller design for automated car parking," in *Proc. 1994 Amer. Contr. Conf.*, Baltimore, MD, June 1994.

[18] ——, "On-line approximation of control programs for a class of nonlinear free-flying systems," in *Proc. 1994 Amer. Contr. Conf.*, Baltimore, MD, June 1994.

[19] E. Hartman and D. Keeler, "Predicting the future: Advantages of semilocal units," *Neural Computa.*, no. 3, pp. 566–578, 1991.

[20] K. J. Hunt, D. Sbarbaro, R. Zbikowski, and P. J. Gawthorp, "Neural networks for control systems—A survey," *Automatica*, vol. 28, no. 6, pp. 1083–1112, 1992.

[21] A. Isidori, *Nonlinear Control Systems*. Berlin: Springer-Verlag, 1989.

[22] V. Kadirkamanathan, M. Niranjan, and F. Fallside, "Sequential adaptation of radial basis function neural networks and its application to time-series prediction," in *Advances in Neural Information Processing Systems 3*, J. E. Moody, R. P. Lipmann, and D. S. Touretzky, Eds. San Mateo, CA: Morgan Kaufmann, 1991, vol. 3, pp. 721–727.

[23] E. J. Kansa, "Multiquadrics—A scattered data approximation scheme with applications to computational fluid dynamics—i," *Computers Math. Applic.*, vol. 19, no. 8, pp. 127–145, 1990.

[24] J. A. Leonard, M. A. Kramer, and L. H. Ungar, "Using radial basis functions to approximate a function and its error bounds," *IEEE Trans. Neural Networks*, vol. 3, no. 4, pp. 624–627, 1992.

[25] C. A. Micchelli, "Interpolation of scattered data: Distance matrices and conditionally positive definite functions," *Const. Approx.*, vol. 2, pp. 11–22, 1986.

[26] J. Moody and C. J. Darken, "Fast learning in networks of locally-tuned processing units," *Neural Computa.*, vol. 2, pp. 281–284, 1989.

[27] S. Mukhopadhyay and K. S. Narendra, "Disturbance rejection in nonlinear systems using neural networks," *IEEE Trans. Neural Networks*, vol. 4, no. 1, pp. 63–72, 1993.

[28] K. S. Narendra and K. Parthasarathy, "Identification and control of dynamical systems using neural networks," *IEEE Trans. Neural Networks*, vol. 1, no. 1, pp. 4–26, 1990.

[29] Y.-H. Pao, *Adaptive Pattern Recognition and Neural Networks*. Reading, MA: Addison-Wesley, 1989.

[30] J. Park and I. W. Sandberg, "Universal approximation using radial-basis-function networks," *Neural Computa.*, vol. 3, pp. 246–257, 1991.

[31] J. Platt, "A resource-allocating network for function interpolation," *Neural Computa.*, vol. 3, no. 3, pp. 213–225, 1991.

[32] T. Poggio and F. Girosi, "Networks for approximation and learning," *Proc. IEEE*, vol. 78, no. 9, pp. 1481–1497, Sept. 1990.

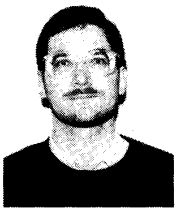
[33] M. J. D. Powell, "Radial basis functions for multivariable interpolation: A review," in *Algorithms for Approximation*, J. C. Mason and M. G. Cox, Eds. Oxford, UK: Clarendon, 1987, pp. 143–168.

[34] ——, "The theory of radial basis function approximation in 1990," in *Advances in Numerical Analysis*, W. Light, Ed. Oxford: Clarendon Press, vol. 2, 1992, pp. 102–205.

[35] R. M. Sanner and J.-J. E. Slotine, "Gaussian networks for direct adaptive control," *IEEE Trans. Neural Networks*, vol. 3, no. 6, pp. 837–863, 1992.

[36] ——, "Stable recursive identification using radial basis function networks," in *Proc. Amer. Contr. Conf.*, Chicago, IL, June 1992, pp. 1829–1833.

[37] A. N. Tikhonov and V. Ya. Arsenin, *Solutions of Ill-Posed Problems*. Washington: Halsted, 1977.



Dimitry M. Gorinevsky (M'91) received the M.Sc. degree with highest honors from Moscow Institute of Physics and Technology (Phystech) in 1982 and the Ph.D. degree from the Department of Mechanics and Mathematics at Moscow State University, Moscow, Russia in 1986.

During 1983–1990 he was with the Institute for Problems of Information Transmission, the Russian Academy of Sciences, Moscow. From 1991–1992 he was with Munich University of Technology, and from 1992–1994 he was with the University of Toronto, Canada. He also worked as an engineering consultant for the Canadian Space Agency. He worked in biomechanics, robotics, legged locomotion, flexible dynamics, control of distributed-parameter systems, automotive, satellite control, and process control. He is currently employed as a Senior Control Engineer by Measurex Devron Inc., North Vancouver, BC, Canada, where he develops advanced control systems for paper machines. His current research is in application of nonlinear approximation and optimization methods, including neural networks, to control, identification, and signal processing. He has authored and co-authored more than 50 technical papers in these areas, co-authored a book on robotics, and holds five patents.

Dr. Gorinevsky was awarded the USSR Academy of Sciences Honorary Diploma for the Young Scientist's Work in the Field of Mathematics, Computing, Mechanics, and Control in 1987 and an International Research Fellowship from the Alexander von Humboldt Foundation, Bonn, Germany in 1990. He is listed in the 1993–1994 *Marquis' Who's Who in the World*. He is a member of several professional organizations including ASME and SAE and is a registered professional engineer in Ontario.