View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Missouri University of Science and Technology (Missouri S&T): Scholars' Mine

Missouri University of Science and Technology
Scholars' Mine

Electrical and Computer Engineering Faculty Research & Creative Works

Electrical and Computer Engineering

01 Jan 2004

Choice of Utility Functions for Adaptive Critic Designs

Ganesh K. Venayagamoorthy
Missouri University of Science and Technology

Follow this and additional works at: https://scholarsmine.mst.edu/ele_comeng_facwork

Part of the Electrical and Computer Engineering Commons

Recommended Citation

G. K. Venayagamoorthy, "Choice of Utility Functions for Adaptive Critic Designs," *Proceedings of the IEEE International Symposium on Intelligent Control, 2003*, Institute of Electrical and Electronics Engineers (IEEE), Jan 2004.
The definitive version is available at https://doi.org/10.1109/ISIC.2003.1254689

This Article - Conference proceedings is brought to you for free and open access by Scholars' Mine. It has been accepted for inclusion in Electrical and Computer Engineering Faculty Research & Creative Works by an authorized administrator of Scholars' Mine. This work is protected by U. S. Copyright Law. Unauthorized use including reproduction for redistribution requires the permission of the copyright holder. For more information, please contact scholarsmine@mst.edu.

Proceedings of the 2003 IEEE
International Symposium on Intelligent Control
Houston, Texas • October 5-8, 2003

Choice of Utility Functions for Adaptive Critic Designs

Ganesh K Venayagamoorthy, *Senior Member*, IEEE
Department of Electrical and Computer Engineering
University of Missouri-Rolla, MO 65409, USA
gkumar@ieee.org

Abstract - This paper first presents a general overview of Adaptive Critic Designs (ACDs) and their existing control applications. It describes the importance of the *right* choice of utility functions for the development of critic networks and their convergence to the cost-to-go function *J*. A closer look into the step by step derivation of an utility function for the design of an ACD nonlinear optimal neurocontroller to replace/augment the conventional controllers, the automatic voltage regulator and governor, in a power system consisting of a generator connected to the power grid is described and some results are presented.

I. Introduction

There are numerous applications of artificial neural networks (ANNs) in control and hence the field of *neurocontrol* which is defined as the use of ANNs to emit control signals for dynamic systems. There are five basic schemes as argued by Werbos [1], namely: the supervised control, direct inverse control, neural adaptive control, backpropagation through time (BPTT) and adaptive critic design (ACD). Of these schemes, only the last two are able to address the problem of planning or optimization over time. BPTT is simple and easy to implement provided an exact model of the dynamic system to be controlled exists. The method does not allow for residual errors in system identification. In addition to that, BPTT requires a flow of information backwards from time *T* to time *T-1*, to *T-2* and, so on. This requires exact storage of the entire time series of observations, therefore not suitable for real time adaptation especially if the truncation depth is high.

Adaptive critic designs have recently emerged as a promising technique for nonlinear optimal control [2 – 6]. ACDs are based on the combined concepts of reinforcement learning and dynamic programming. In dynamic programming, the user supplies a utility function, *U(.)*, and a stochastic model of the external environment and using the Bellman's equation of dynamic programming, given in (1), a secondary or strategic utility function *J* is calculated. In (1), γ is called the discount factor and has a value between 0 and 1.

$$J(t) = \sum_{k=0}^{\infty} \gamma^k U(t+k) \qquad (1)$$

The basic theorem of dynamic programming is that maximizing *J* in the short term is equivalent to maximizing *U* in the long term and vice versa. Unfortunately, the number of computations increases exponentially with number of states in the environment; therefore, true dynamic programming is not feasible. With ACDs, it is possible to find an approximate solution to dynamic programming which is a difficult task to accomplish for complex problems. A network called the 'Critic' in ACDs is used to approximate the *J* function in (1). A family of ACDs was proposed by Werbos [7] and these include the model dependent ones: Heuristic Dynamic Programming (HDP) that approximates *J*, Dual Heuristic Programming (DHP) that approximates the derivatives of *J*, and Global Dual Heuristic Programming (GDHP) that approximates *J* and its derivatives, listed in order of increasing complexity and power [8]. Other versions of ACDs that do not use models called the action-dependent designs exist as well [8].

The selection of the utility function *U(.)* and the discount factor γ play an important role in getting the secondary function J to converge over time. The convergence of the critic network is critical in obtaining an optimal neurocontroller. This paper describes how the utility function for a power system adaptive critic's based neurocontroller is obtained. Results are presented to show the impact of different discount factors on the critic network training.

Section II of the paper describes adaptive critic designs further with emphasis given to HDP and DHP [3, 4]. Section III describes the power system considered to demonstrate the adaptive critic based neurocontrol. Section IV describes the steps involved in deriving the utility function. Section V presents some critic convergence results and some power system control results.

II. Adaptive Critic Designs

The adaptive critic method determines optimal control laws for a system by successively adapting two neural networks, namely an Action neural network (which dispenses control signals) and a Critic neural network (which learns the desired performance index for some function associated with the performance index). These two neural networks approximate the Hamilton-Jacobi-Bellman equation associated with optimal control theory. The following subsections describe briefly the HDP and DHP techniques [3, 4].

A. Heuristic Dynamic Programming

The HDP critic network estimates the function *J* (cost-to-go) in (1). The configuration for training the critic network

0-7803-7891-1/03/$17.00 © 2003 IEEE

Authorized licensed use limited to: Missouri University of Science and Technology. Downloaded on October 23, 2009 at 12:03 from IEEE Xplore. Restrictions apply.

is shown in Fig. 1. The critic network is a neural network trained forward in time, which is of great importance for real-time operation. The inputs to the critic are the *estimated* states (outputs of the model neural network), and their two time delayed values respectively, forming the six inputs. The critic network tries to minimize the following error measure over time

$$\|E_{C1}\| = \frac{1}{2}\sum_t e_{C1}^2(t) \qquad (2)$$

$$e_{C1}(t) = J[\Delta Y(t)] - \gamma J[\Delta Y(t+1)] - U(t) \qquad (3)$$

where $\Delta Y(t)$ is a vector of observables of the plant (or the states, if available). The necessary condition for (2) to be minimal is given in (4).

$$\frac{1}{2}\frac{\partial}{\partial W_{C1}}\left\langle e_{C1}^2(t)\right\rangle = \left\langle e_{C1}\frac{\partial e_{C1}(t)}{\partial W_{C1}}\right\rangle = 0 \qquad (4)$$

The expression for the weights' update for the critic neural network is as follows:

$$\Delta W_{C1} = -\eta_2\{J[\Delta Y(t)] - \gamma J[\Delta Y(t+1)] - U(t)\}\frac{\partial\{J[\Delta Y(t)] - \gamma J[\Delta Y(t+1)] - U(t)\}}{\partial W_{C1}} \qquad (5)$$

where η_2 is a positive learning rate and W_{C1} is the weights of the HDP critic neural network.

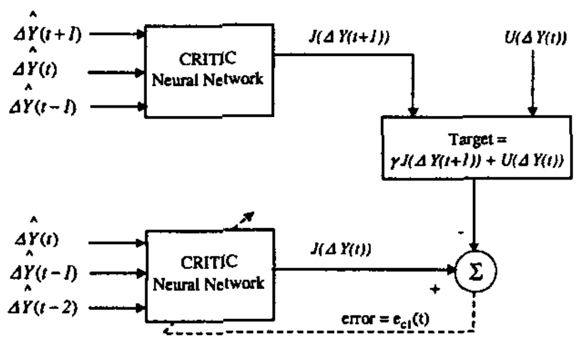

Fig. 1 Critic Adaptation in HDP. The same critic network is shown for two consecutive times, t and t+1. The critic's output J(t+1) at time t+1, is necessary to generate a target signal γJ(t+1) + U(t), for training the critic network. The discount factor γ is chosen to be 0.5. The backpropagation path is shown by the dashed line

The same critic neural network is shown in two consecutive moments in time in Fig. 1. The critic neural network's output $J(t+1)$ is necessary in order to provide the training signal $\gamma J(t+1) + U(t)$, which is the target value for $J(t)$. The objective of the action neural network in HDP setup, is to minimize J in the immediate future, thereby optimizing the overall cost expressed as a sum of all $U(t)$ over the horizon of the problem. This is achieved by training the action neural network with an error signal $\partial J(t)/\partial A(t)$. The gradient of the cost function $J(t)$, with respect to the outputs $A(t)$, of the action network, is obtained by backpropagating $\partial J/\partial J$ (i.e. the constant 1) through the critic neural network and then through the pretrained model neural network to the action neural network. This gives $\partial J(t)/\partial A(t)$ and $\partial J(t)/\partial W_{A1}(t)$ for all the outputs of the action neural network, and all the action neural network's weights W_{A1}, respectively. The action neural network is trained to minimize (6). The expression for the weights' update in the action neural network is based on an error feedback from the critic neural network backpropagated through the model neural network using the backpropagation algorithm, and is given by (7) [4].

$$\|E_{A1}\| = \frac{1}{2}\sum_t e_{A1}^2(t) \qquad (6)$$

$$\Delta W_{A1} = -\eta_3\frac{\partial J(t)}{\partial A(t)}\frac{\partial}{\partial W_{A1}}\left(\frac{\partial J(t)}{\partial A(t)}\right) \qquad (7)$$

where η_3 is a positive learning rate and W_{A1} is the weights of the HDP action neural network.

$U(t)$ is the local utility function which plays a direct role in the HDP critic training (5) and indirectly through the critic network on the action network training (7). Here, the value of J is a scalar and on the critic convergence J takes a value depending on (1), which is dependent on $U(t)$ and γ. It is critical that $U(t)$ is a true evaluation of the states of the plant and this is dealt with in detail in Section IV. J(t) is expected for a trained critic and action network to ideally be zero. But most of the time a non-zero value is observed.

B. Dual Heuristic Programming

The DHP critic neural network in Fig. 2, estimates the derivatives of J with respect to the vector ΔY, and learns minimization of the following error measure over time:

$$\|E_{C2}\| = \sum_t e_{C2}^T(t)e_{C2}(t) \qquad (8)$$

where

$$e_{C2}(t) = \frac{\partial J[\Delta Y(t)]}{\partial \Delta Y(t)} - \gamma\frac{\partial J[\Delta Y(t+1)]}{\partial \Delta Y(t)} - \frac{\partial U(t)}{\partial \Delta Y(t)} \qquad (9)$$

where $\partial(.)/\partial\Delta Y(t)$ is a vector containing partial derivatives of the scalar (.) with respect to the components of the vector ΔY. The critic neural network's training is more complicated than in HDP, since there is a need to take into account all relevant pathways of backpropagation as shown in Fig. 2, where the paths of derivatives and adaptation of the critic are depicted by dashed lines. The adaptation of the DHP critic network weights based on the error given by (5), can be further expressed using chained derivatives by (10).

$$e_{C2}(t) = \frac{\partial J[\Delta Y(t)]}{\partial \Delta Y_j(t)} - \gamma\frac{\partial J[\Delta Y(t+1)]}{\partial \Delta Y_j(t)} - \frac{\partial U(t)}{\partial \Delta Y_j(t)} - \sum_{k=1}^{m}\frac{\partial U(t)}{\partial A_k(t)}\frac{\partial A_k(t)}{\partial \Delta Y_j(t)} \qquad (10)$$

The adaptation of the action neural network in Fig. 2 is to satisfy the goal [8] and is given by (11). The weights' update expression [8], when applying backpropagation, is given by (12).

Authorized licensed use limited to: Missouri University of Science and Technology. Downloaded on October 23, 2009 at 12:03 from IEEE Xplore. Restrictions apply.

$$\frac{\partial U(t)}{\partial A(t)}+\gamma\frac{\partial J(t+1)}{\partial A(t)}=0 \quad \forall\ t \tag{11}$$

$$\Delta W_{A2}=-\eta_4\left[\frac{\partial U(t)}{\partial A(t)}+\frac{\partial J(t+1)}{\partial A(t)}\right]^T\frac{\partial A(t)}{\partial W_{A2}} \tag{12}$$

where η_4 is a positive learning rate and W_{A2} is the weights of the action neural network in the DHP scheme. The structure of the action neural network is identical to that of the action network in the HDP scheme. The general derivation of the equations in this section is given in detail in [4].

In DHP, the utility function has a direct effect in both the critic and action neural networks' training. DHP critic output is the derivatives of J with respect to the model outputs; therefore, if the trained action is giving the optimal right control signal, the critic outputs should be zero.

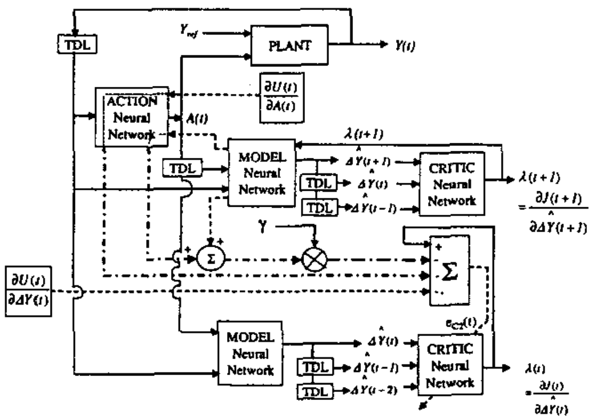

Fig. 2 Critic adaptation in DHP. This diagram shows the implementation of eq. (10). The same critic network is shown for two consecutive times, t and t+1. Backpropagation paths are shown by dashed lines. The output of the critic network λ(t+1) is backpropagated through the model network from its outputs to its inputs. Backpropagation of the vector ∂U(t)/∂A(t) through the action network results in a vector with components computed as the last term of eq. (10). The summation produces the error vector $e_{c2}(t)$ used for training the critic network

III. Power System

The synchronous generator, turbine, exciter and transmission system connected to an infinite bus form the power system called the *plant* (dotted block) in Fig. 3 that has to be controlled. Nonlinear equations are used to describe and simulate the dynamics of the plant in order to generate the data for the optimal neurocontrollers [3]. On a physical plant, this data would be measured. In the plant, P_t and Q_t are the real and reactive power at the generator terminal, respectively, Z_e is the transmission line impedance, P_m is the mechanical input power to the generator, V_{fd} is the exciter field voltage, V_b is the infinite bus voltage, $\Delta\omega$ is the speed deviation, ΔV_t is the terminal voltage deviation, V_t is the terminal voltage, ΔV_{ref} is the reference voltage deviation, V_{ref} is the reference voltage, ΔP_{in} is the input power deviation, and P_{in} is the turbine input power. The position of the switches *S1* and *S2* in Fig. 3 determines whether the optimal neurocontroller, or the conventional controller (CONVC) consisting of governor and AVR, is controlling the plant. Block diagrams, time constants, and gains of the CONVC (AVR/exciter and turbine/governor systems) are given in [3].

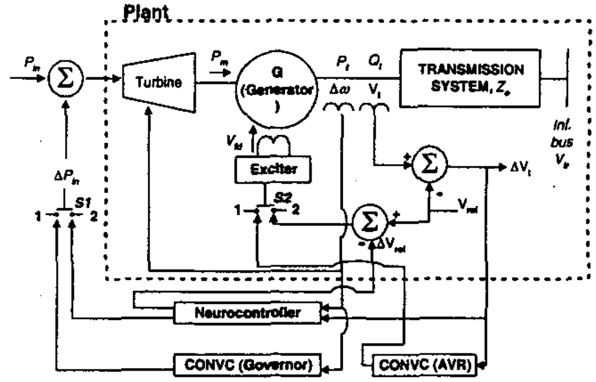

Fig. 3 A power system with a synchronous generator connected to an infinite bus with conventional automatic voltage regulator and governor controllers (CONVC) that are augmented/replaced by a neurocontroller

IV. Design of Utility Function

The design of the utility function $U(t)$in (1) is described in [9] and is repeated here. $U(t)$ is designed based on a desired response predictor which has the following characteristics:

(a) It must be flexible enough to modify the performance of the generator.
(b) The desired response signal must ensure that the generator is inherently stable at all times. In other words, the predictor must be stable.
(c) The desired response signal must incorporate the effects of a power system stabiliser.

The optimal predictor is designed on the basis of guiding the disturbed output variables, in this case the terminal voltage and speed, of the generator to a desired steady operating point or set point, in a step by step fashion. In other words, a desired trace of outputs from t_i to t_{i+j} can be predicted, based on the present and past-time values of the outputs. Optimal here refers to predictions of the desired response for the generator and ensuring its stability over a wide range of operating conditions. The prediction equation of the optimal predictor is given in (13).

$$\hat{X}(k+1)=B_0X(k)+B_1X(k-1)+...+B_NX(k-N) \tag{13}$$

$B_i(i$ = 0,1,...,N) are chosen so that any disturbed output variable always transfers towards the desired steady operating point, that is the optimal predictor is always globally asymptotically stable. $\hat{X}$ is the value predicted for the next immediate time step and X can be either the

Authorized licensed use limited to: Missouri University of Science and Technology. Downloaded on October 23, 2009 at 12:03 from IEEE Xplore. Restrictions apply.

terminal voltage deviation ΔV_t or speed deviation $\Delta\omega$.

In (13), it is assumed that each output variable of the optimal predictor is a linear combination of the independently predicted output variables of the dynamic system. The magnitude of the coefficients, A_i, determine the magnitude of the error signal between the identifier output and the desired response signal (or predictor), and therefore, the magnitude of the error backpropagated to the controller to adapt its weights.

If the output $X(t)$ is bounded for $0 < t < \infty$ and

$$\lim_{t\to\infty}\left(X(t) - \hat{X}(t) \right) = 0 \tag{14}$$

then a predictor can be designed which forces the generator, by means of the neurocontroller, back to desired setpoints [9]. The magnitude of the forcing signal depends on the coefficients A_i.

The conditions defined by (14) are necessary because it is not possible to damp the generator to take up the required setpoints if its outputs are unbounded. If (14) does not hold then the outputs of the generator will not return to their setpoints after a given disturbance. The fundamental assumption made in this design is that it is possible for a controller to return a generator to its setpoints after a disturbance.

Equation (13) can be re-written in the following way:

$$\hat{X}(k+1) = \alpha_i \hat{x}_1(k+1) + \beta_i \hat{x}_2(k+1) + \ldots + \gamma_i \hat{x}_i(k+1) \tag{15}$$

where

$$\hat{x}_i(k+1) = a_{i0}x_i(k) + a_{i1}x_i(k-1) + \ldots + a_{iN}x_i(k-N) \qquad i = 1, 2, \ldots, h \tag{16}$$

The eigenfunction polynomial of (16) is

$$Z - a_{i0} - a_{i1}Z^{-1} - a_{i2}Z^{-2} - \ldots - a_{iN}Z^{-N} = 0 \tag{17}$$

or

$$\begin{aligned} &Z^{N+1} - a_{i0}Z^{N} - a_{i1}Z^{N-1} - \ldots - a_{iN} \\ &= (Z - S_{i0})(Z - S_{i1})\ldots(Z - S_{iN}) = 0 \end{aligned} \tag{18}$$

If S_{i0}, S_{i1}, ..., S_{iN} ($i = 1,2,\ldots, h$) are chosen inside a unit circle, then (15) will be globally asymptotically stable. It should be pointed out that α_i, β_i, ..., γ_i ($i = 1, 2, \ldots, h$) in (15) are the qualitative coefficients, and are not relevant to the stability of the dynamic system. These coefficients describe the relationship between the desired outputs of the optimal predictor and the outputs of the dynamic system, and may be chosen according to the qualitative requirements of the controlled generator system.

An optimal predictor for the generator is designed as follows:

$$\begin{aligned} \hat{x}_1(k+1) &= \sum_{i=0}^{2} a_{1i}x_1(k-i) + \beta_1\{\sum_{i=0}^{2} a_{2i}x_2(k-i)\} \\ \hat{x}_2(k+1) &= \sum_{i=0}^{2} a_{2i}x_2(k-i) \end{aligned} \tag{19}$$

where $a_{1i}(i = 0, 1, 2)$ can be obtained by

$$Z^3 - a_{10}Z^2 - a_{11}Z - a_{12} = (Z^2 - S_{11})(Z - S_{12}) \tag{20}$$

S_{11} and S_{12} are real and inside a unit circle. a_{2i} ($i = 0, 1, 2$) can be obtained in the same way, $0 < \beta_1 < 1$.

In (19), $\hat{x}_1(k+1)$ and $\hat{x}_2(k+1)$ refer to the predicted terminal voltage and speed deviation respectively. The predicted terminal voltage deviation depends on both the terminal voltage and speed deviation signals. The weighting of the speed deviation on the predicted terminal voltage deviation depends on the value of β_1. The inclusion of the speed deviation signal for predicting the terminal voltage deviation brings in the effects of power system stabilisers.

To find suitable values for the coefficients in (20), several simulations are carried out starting with small values for S_{11} and S_{12} and the response of the controller (to disturbances such as step changes in terminal voltage, and three phase short circuits) is evaluated. Small values of S_{11} and S_{12} give better damped responses in generator speed and voltage. The values S_{11} and S_{12} are increased in steps until acceptable voltage and speed responses are achieved. If too large values of S_{11} and S_{12} are used then the voltage and speed of the generator overshoot their setpoints.

The predicted terminal voltage and speed deviation are given by (21).

$$\begin{aligned} \hat{x}_1(k+1) &= 4\Delta V(k) + 4\Delta V(k-1) + 16\Delta V(k-2) \\ &+0.01\{0.4\Delta\omega(k) + 0.4\Delta\omega(k-1) + 0.16\Delta\omega(k-2)\} \\ \hat{x}_2(k+1) &= 0.4\Delta\omega(k) + 0.4\Delta\omega(k-1) + 0.16\Delta\omega(k-2) \end{aligned} \tag{21}$$

It can be seen from (21) that the coefficients (4, 4, and 16) used for the terminal voltage deviation fall outside the unit circle. Nevertheless, the results for the neurocontrollers (HDP and DHP) in the section V show that the large values of the coefficients used for voltage deviation (a_{1i}) do not cause instability. The reasons for this are as follows:

(a) The limit in (14) applies to the terminal voltage deviations. The neurocontroller creates a damping signal only when there is a difference between the generator's setpoint terminal voltage and the instantaneous terminal voltage. The controller ensures that this difference becomes zero over a period of time and the output of (21) will then be zero even with large coefficients ($a_{1i}>1$) for the voltage deviation terms.

(b) The generator used in this study has an open loop frequency response of about 0.3 Hz to changes in terminal voltage setpoints, which is considered to be slow. Therefore the damping signal mentioned in (a) does not cause any oscillation about the setpoint.

The utility function $U(k)$ is taken to be the sum of squares of $\hat{x}_1(k+1)$ (with $\beta_1 = 0$) and $\hat{x}_2(k+1)$ as shown in (22). The squares are taken to ensure a positive utility function when deviations exist.

$$\begin{aligned} U(k) &= [4\Delta V(k) + 4\Delta V(k-1) + 16\Delta V(k-2)]^2 \\ &+[0.4\Delta\omega(k) + 0.4\Delta\omega(k-1) + 0.16\Delta\omega(k-2)]^2 \end{aligned} \tag{22}$$

Authorized licensed use limited to: Missouri University of Science and Technology. Downloaded on October 23, 2009 at 12:03 from IEEE Xplore. Restrictions apply.

V. HDP/DHP Critic and Action Network Results

The HDP critic output (*J(t)*) are shown in Figs. 4 and 5 during training with discount factors of 0.5 and 0.9 respectively for pseudorandom binary signals (PRBS) [3, 4] applied to the power system at 4 seconds. As the discount factor is increased the value of *J(t)* increases and vice-versa. As the discount factor is varied between 0 and 1, it is observed that for lower values of γ, the *J(t)* flows *U(t)*.

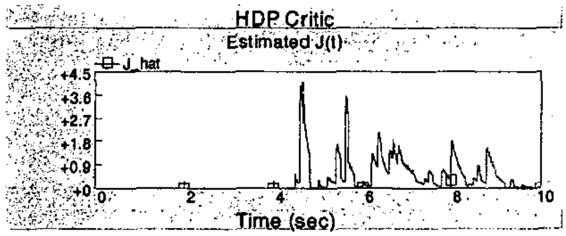

Fig. 4 HDP critic's output for the utility function in (22) discount factor of 0.5 for the system undergoing constant perturbations

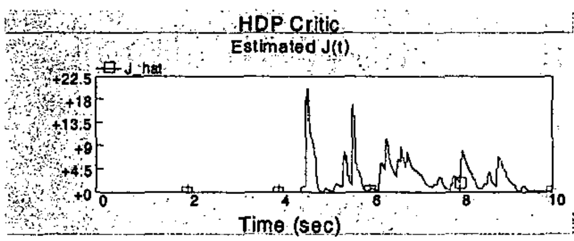

Fig. 5 HDP critic's output for the utility function in (22) discount factor of 0.9 for the system undergoing constant perturbations

A U(t) given in (23) with coefficients less than one is used to train the critic network and the estimated J(t) by the network for two different discount factors, 0.5 and 0.9, are shown in Figs. 6 and 7. The direct impact of this *U(t)* is less cost-to-go, *J(t,)* but the control signal damping was not good enough compared to that obtained with *U(t)* in (22) with both discount factors (0.5 and 0.9).

$$U(k) = [0.4\Delta V(k) + 0.4\Delta V(k-1) + 0.16\Delta V(k-2)]^2 + [0.4\Delta\omega(k) + 0.4\Delta\omega(k-1) + 0.16\Delta\omega(k-2)]^2 \quad (23)$$

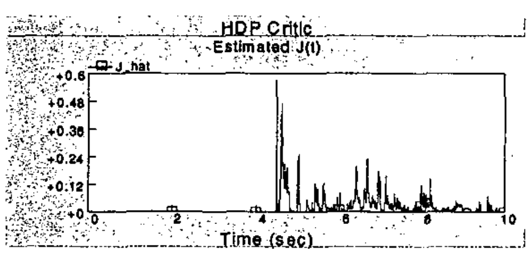

Fig. 6 HDP critic's output for the utility function in (23) discount factor of 0.5 for the system undergoing constant perturbations

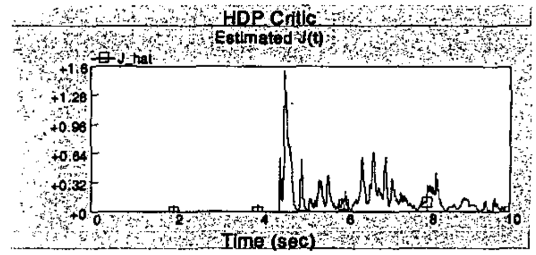

Fig. 7 HDP critic's output for the utility function in (23) discount factor of 0.9 for the system undergoing constant perturbations

Fig. 8 shows that the HDP critic network converges eventually after undergoing training with PRBS signals for 50 seconds for *U(t)* in (22) and with a discount factor of 0.5.

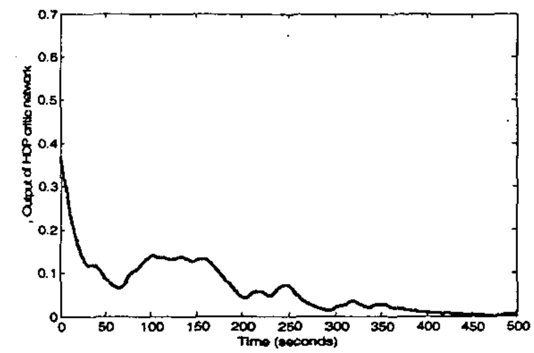

Fig. 8 HDP critic output for the utility function in (22) and discount factor of 0.5 after undergoing training for 50s with perturbations

The DHP critic network is trained with PRBS signals first. The trained DHP critic's outputs (two) are shown in Figs. 9 and 10 below.

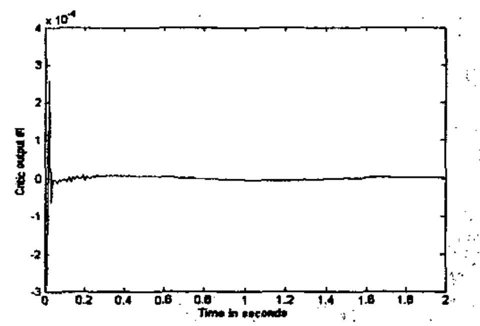

Fig. 9 Trained DHP critic network output #1 for the utility function *U(t)* in (22) and for a discount factor of 0.5

The dynamic and transient operation of the HDP and DHP action neural network (neurocontroller) is compared with the operation of a conventional PID controller (AVR and turbine governor), under two different conditions: ± 5% step changes in the terminal voltage setpoint, and a three phase short circuit at the infinite bus. Fig. 11 shows the

Authorized licensed use limited to: Missouri University of Science and Technology. Downloaded on October 23, 2009 at 12:03 from IEEE Xplore. Restrictions apply.

performance of the different controllers for ± 5% desired step changes in the terminal voltage with the turbogenerator operating at 1 pu real power (P) and 0.85 lagging power factor (pf) (at the generator terminals), with the transmission line impedance $Z_1 = 0.02 + j\,0.4$ pu.

Fig. 12 shows a generator operating under the same conditions but experiencing a temporary 50 ms three phase short circuit at the infinite bus, with an increased transmission line impedance $Z_2 = 0.025 + j\,0.6$ pu. The results with the conventional AVR and governor controllers, and those of the HDP and the DHP neurocontrollers, are labeled as CONV, HDP and DHP respectively in these figures.

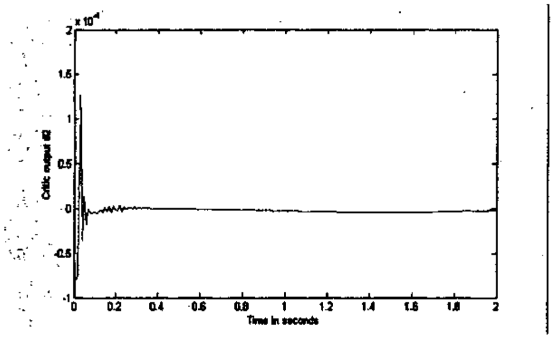

Fig. 10 Trained DHP critic network output #2 for the utility function *U(t)* in (22) and for a discount factor of 0.5

VI. Conclusions

This paper has summarized the important steps in obtaining utility function for an adaptive critic based neurocontroller on a power system control problem. It is necessary to go through an analytical approach rather just taking a binary utility function with a '1' or '0' for respective plant outputs as reported in many ACD applications. It is noticed that having *U(t)* that progressively increases or decreases with the plant outputs over time is desirable for plant and controller stability. It is also critical to pay attention to the discount factor since it determines the slope of the critic function. It is desirable for γ not to be too small or too large in order to initiate *efficient* action network learning. It is important in the design of ACD based controllers to have insight into the cost-to-go and utility functions as the critic network is trained to convergence. It is anticipated that an adaptive *U(t)* will enhance critic learning and this needs to be investigated.

VII. References

[1] P.J. Werbos, "Consistency of HDP Applied to a Simple Reinforcement Learning Problem", *Neural Networks*, vol. 3, 1990, pp. 179 – 189.

[2] P. J. Werbos, A Menu of Designs for Reinforcement Learning Over Time, In R S Sutton, W T Miller, III and P J Werbos, editors, *Neural Networks for Control*, MIT Press, 1990.

[3] G.K. Venayagamoorthy, R.G. Harley, D.C. Wunsch, "Implementation of Adaptive Critic Based Neurocontroller for Turbogenerators in a Multimachine Power System", *IEEE Trans. on Neural Networks*, vol. 14, no. 5, September 2003.

[4] G.K. Venayagamoorthy, R.G. Harley, D.C. Wunsch, "Comparison of Heuristic Dynamic Programming and Dual Heuristic Programming Adaptive Critics for Neurocontrol of a Turbogenerator", *IEEE Trans. on Neural Networks*, vol. 13, no. 3, May 2002, pp. 764 -773.

[5] D.Z. Han, S.N. Balakrishnan, "State-Constrained Agile Missile Control with Adaptive-Critic-Based Neural Networks", *IEEE Transactions on Control Systems Technology*, vol. 10, no. 4, July 2002, pp. 481 – 489.

[6] D. Liu, "Adaptive Critic Designs for Problems with Known Analytical Form of Cost Function", *IEEE International Conference on Neural Networks*, vol. 2, 2002, pp. 1808 – 1813.

[7] P.J. Werbos, *Approximate Dynamic Programming for Real-Time Control and Neural Modeling*, in Handbook of Intelligent Control, White and Sofge, Eds., Van Nostrand Reinhold, ISBN 0-442-30857-4, pp 493 – 525.

[8] D. Prokhorov, D. Wunsch, "Adaptive Critic Designs", *IEEE Trans. on Neural Networks*, vol. 8, no.5, 1997, pp 997-1007.

[9] G.K. Venayagamoorthy, R.G. Harley, D.C. Wunsch, "Dual Heuristic Programming Excitation Neurocontrol for Generators in a Multimachine Power System", *IEEE Trans. on Industry Applications*, vol. 39, no. 2, March 2003, pp. 382 to 394.

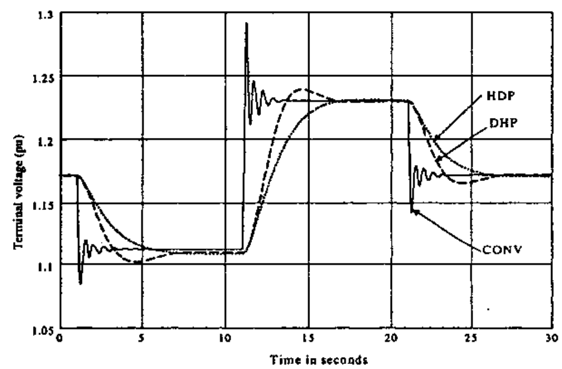

Fig. 11 Terminal voltage variations for ± 5% step changes in the desired terminal voltage [4]

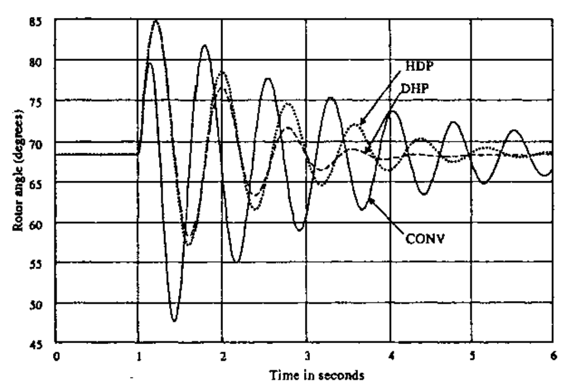

Fig. 12 Rotor angle variation for a temporary 50 ms three phase short circuit at the infinite bus with increased transmission line impedance (P = 1 pu, pf = 0.85 lagging, Z = 0.025 + j 0.6 pu) [4]

Authorized licensed use limited to: Missouri University of Science and Technology. Downloaded on October 23, 2009 at 12:03 from IEEE Xplore. Restrictions apply.