

Damping Control by Fusion of Reinforcement Learning and Control Lyapunov Functions

Mevludin Glavic, *Member, IEEE*, Damien Ernst, *Member, IEEE*, Louis Wehenkel, *Member, IEEE*

Abstract—The main idea behind the concept, proposed in the paper, is the opportunity to make control systems with increased capabilities by synergetic fusion of the domain-specific knowledge and the methodologies from control theory and artificial intelligence. The particular approach considered combines Control Lyapunov Functions (CLF), a constructive control technique, and Reinforcement Learning (RL) in attempt to optimize a mix of system stability and performance. Two control schemes are proposed and the capabilities of the resulting controllers are illustrated on a control problem involving a Thyristor Controlled Series Capacitor (TCSC) for damping oscillations in a four-machine power system.

Index Terms — Reinforcement learning, Control Lyapunov functions, Power system damping control.

I. INTRODUCTION

RECENT theoretical and technological advances such as availability of powerful constructive control techniques [1]-[3], new control devices [4], better communication and data acquisition techniques [5], and better computational capabilities open the possibilities to implement advanced control schemes in power systems [4], including a possibility to endow power system controllers with the methods to learn and update their decision-making capability [6]-[10].

The constructive control techniques [1]-[3], in particular the concept of CLF, provide powerful tools for studying stabilization problems, both as a basis of theoretical developments and as the methods for actual feedback design.

RL emerges as an attractive learning paradigm that offers a panel of methods that allow controllers to learn a goal oriented control law from interactions (by trial-and-error) with a system or its simulation model. RL driven controllers (agents) observe the system state, take actions, and observe the effects of these actions. By processing the experience they progressively learn an appropriate control law (so called “control policy”) in order to fulfill a pre-specified objective.

Power system community started getting interested in application of the concepts of CLF and RL to control power system quite recently [6]-[14]. Different types of practical problems in using RL methods to solve power system control problems were discussed in [6]-[10], while the concept of CLF was extensively studied in [11]-[14] in the context of

power oscillations damping.

Most nonlinear control design methods based on CLF provide strong guarantees of stability but do not directly address important issue of control performances. On the other hand, the learning by trial-and-error is not justified when one intends to apply it on-line and when the primary concern is the system stability. In this case the “exploration-stability” tradeoff is to be resolved [15].

In principle, a feedback control system should try to optimize some mix of stability and performance [15]. This was recognized and the research efforts were undertaken to tackle this issue in the context of inverse optimality techniques [16], a time-varying/dynamic compensation strategy [17], adaptive critic design techniques [15], and adaptive dynamic programming [18].

In this paper a constructive control based on the fusion of CLF and RL techniques is considered for oscillations damping, a phenomenon of paramount importance in many real-life systems [4] related to the growth of extensive power systems and especially with the interconnection of these systems with ties of limited capacity, by controlling a TCSC device in a four-machine power system [19]. We extend our previous work presented in [6] and propose two control schemes, provide theoretical basics of the two frameworks (CLF and RL) and provide more in-depth analysis of the proposed control schemes.

The paper is organized as follows: theoretical basics of the CLF concept are provided in section 2; CLF-based stabilizing control laws for a TCSC device are presented in section 3, the theoretical foundation of RL is shortly presented in section 4; two control schemes are introduced in section 5; in section 6 test power system model is described; sections 7, 8, and 9 provide results and discussion of the two case studies; section 10 and 11 discuss some future research opportunities and provide conclusions.

II. CONTROL LYAPUNOV FUNCTIONS: THEORETICAL BASICS

The concept of CLF was introduced by Artstein [2] and Sontag [3] and had tremendous impact on stabilization theory [1]-[3]. Its main significance is the fact it converted stability descriptions into tools for solving stabilization tasks. To make this text self-contained, some results about the CLF from control theory [1]-[3] are briefly presented.

The discussion in this section largely follows that in [1], [2], [11]-[14]. Consider a nonlinear, affine in control, system,

$$\dot{x} = f(x, u) = f(x) + g(x) \cdot u, \quad x \in \Omega \subseteq R^n. \quad (1)$$

The authors are with the Electrical Engineering and Computer Science Department, the University of Liège, Sart Tilman B28, 4000 Liège, BELGIUM. E-mails: {glavic, ernst, lwh}@montefiore.ulg.ac.be

One way to stabilize this system is to select a Lyapunov function $v(x)$ and then to find a feedback control law $u(x)$ that renders $\dot{v}(x, u(x))$ negative definite [1], [2], [12]. If there exists a feedback control $u = u(x)$ defined in a neighborhood q of the origin such that the closed-loop system $\dot{x} = f(x, u(x))$ has a locally asymptotically stable equilibrium point at the origin $f(0, u(0))$ then the system (1) is stabilizable at the origin and the function $u = u(x)$ is called a stabilizing feedback law. The function $v(x)$, for the nonlinear system (1), is called CLF if [1], for all $x \neq 0$,

$$L_g v(x) = 0 \Rightarrow L_f v(x) < 0 \quad (2)$$

where $L_g v(x) := \left(\frac{\partial v}{\partial x}\right) \cdot g(x)$ and $L_f v(x) := \left(\frac{\partial v}{\partial x}\right) \cdot f(x)$.

The existence of a CLF is necessary and sufficient condition for the stabilizability of a system with a control input. When $v(x)$ is a CLF, there are many control laws that render $\dot{v}(x)$ negative definite [1]. This fact is considered in the paper as an opportunity rather than difficulty and more than one stabilizing control law is used for performance improvement purpose. The construction of a CLF is a hard problem, which has been solved for some special classes of systems [1]. Although a difficult task and still the subject of ongoing research [20], [21], the issue of finding Lyapunov functions candidates and construction of a CLF for a power system are not addressed in the paper. Instead, the results from [11]-[14] are largely followed.

III. CLF-BASED STABILIZING CONTROL LAWS FOR A TCSC

The energy function of uncontrolled system, for reduced network model, was used as a Lyapunov function candidate in all [11]-[14]. This approach is known, in control theory, as Jurdjevic-Quinn control [22] or damping control for stable systems where the uncontrolled system is stable and the task of the control is to provide additional damping.

Suppose a TCSC device is installed in the line between the buses l and m . The TCSC is considered as a variable reactance placed in series with a transmission line. Following the results from [11]-[14] the stabilizing control laws of the TCSC's reactance reference can be formalized as,

$$X_{ref} = F(V_{lm} \dot{V}_{lm}), \quad (3)$$

where $F(V_{lm} \dot{V}_{lm}) > 0$ for $V_{lm} \dot{V}_{lm} > 0$, $F(V_{lm} \dot{V}_{lm}) < 0$ for $V_{lm} \dot{V}_{lm} < 0$, and $F(0) = 0$. In (3) V_{lm} is the magnitude of the voltage drop across the line l - m where the TCSC is installed. Voltage V_{lm} is locally measurable. If the line has impedance X_{lm} then,

$$V_{lm} \cong X_{lm} I_{lm} + V_{TCSC}, \quad (4)$$

where V_{TCSC} is the magnitude of the voltage drop across the TCSC. \dot{V}_{lm} can be inferred based upon tracking of V_{lm} . A wide range of possible choices for the control function $F(V_{lm} \dot{V}_{lm})$ can be used [13]. Different possibilities for the control function were addressed in [13]. In this paper we consider only linear, continuous control function $k \cdot V_{lm} \dot{V}_{lm}$.

An important observation is that the control laws are based on locally measurable quantities.

IV. REINFORCEMENT LEARNING

A. Theoretical framework

RL is presented here in the context of a *deterministic* time-invariant system, sampled at constant rate. If x_t and u_t denote the system state and control at time t , the state at time $t+1$ is given by,

$$x_{t+1} = f(x_t, u_t), \quad (5)$$

where assumed that $u_t \in U, \forall t \geq 0$ and that U is finite.

To formulate the problem mathematically the framework of discounted infinite time horizon optimal control is used here. Let $r(x, u) \leq B$ be a *reward* function, $\gamma \in (0, 1)$ a discount factor, and denote by $u_{\{t\}} = (u_0, u_1, u_2, \dots)$ a sequence of control actions applied to the system. The objective is to define, for every possible initial state x_0 , an optimal control sequence $u_{\{t\}}^*(x_0)$ maximizing the discounted return,

$$R(x_0, u_{\{t\}}) = \sum_{t=0}^{\infty} \gamma^t r(x_t, u_t). \quad (6)$$

One can show that these optimal control sequences can be expressed in the form of a single time-invariant closed-loop control policy $u^*(\cdot)$, i.e. $u_t^*(x_0) = u^*(x_t), \forall x_0 \forall t \geq 0$. In order to determine this policy one can define so-called the Q function [23], [24] by,

$$Q(x, u) = r(x, u) + \gamma V(f(x, u)), \quad (7)$$

where $V(f(x, u))$ is the value function [23], [24], and deduce the optimal control policy,

$$u^*(x) = \arg \max_{u \in U} Q(x, u), \quad (8)$$

by solving Bellman equation [25]. Equation (8) provides a straightforward way to determine the optimal control law from the knowledge of Q . RL algorithms estimate the Q function by interacting with the system.

B. State-space discretization and learning Q function

When the state-space is infinite (as in most power system

problems) the Q function has to be approximated [23]-[26]. In the considered applications, a state space discretization technique is used to this end. It consists in dividing the state space into a finite number of regions and considering that on each region the Q function depends only on u . Then, in the RL algorithms, the notion of state used is not the real state of the system x but rather the region of the state space to which x belongs. Let the letter s to denote a discretized state, $s(x)$ the region to which the (true) state x belongs, and S the finite set of discretized states. Notice that the sole knowledge of the region $s(x_t)$ at some time instant t together with the control value u is not sufficient (in general) to predict with certainty the region $s(x_{t+1})$ to which the system will move at time $t+1$. This uncertainty is modeled by assuming that the sequence of discretized states followed by the system under a certain control sequence is a time-invariant Markov chain characterized by transition probabilities $p(s'|s, u)$. Given these transition probabilities and a discretized reward signal, i.e. a function $r(s, u)$, the initial control problem can be reformulated as a Markov Decision Process (MDP).

The corresponding Q function is characterized by the following Bellman equation,

$$Q(s, u) = r(s, u) + \gamma \sum_{s' \in S} p(s'|s, u) \max_{u \in U} Q(s', u), \quad (9)$$

the solution of which can be estimated by a classical dynamic programming algorithm like the value iteration or the policy iteration [23]-[26]. The corresponding optimal control policy is extended to the original control problem, which is thus defined by the following rule,

$$\hat{u}^*(x) = u^*(s(x)) = \arg \max_{u \in U} Q(s(x), u). \quad (10)$$

RL methods either estimate the transition probabilities and the associated rewards (model-based learning methods) and then compute the Q function, or learn directly the Q function without learning any model (non-model-based learning methods). For the purpose of this paper a model-based algorithm, known as prioritized sweeping [26], with ε -greedy policy, is used. This policy consists in choosing with a probability ε a control action at random and with probability $\varepsilon-1$ the “optimal” control action associated with the current state by the current estimate of Q function. The value of ε defines the so-called “exploration-exploitation” tradeoff used by the algorithm [23]-[26].

V. PROPOSED CONTROLS: THE PRINCIPLE

The conceptual diagrams of the proposed control schemes are illustrated in Fig. 1. In the first control scheme (Fig. 1a) a RL algorithm works in usual setting [6]-[10] and at each time step suggests a control action in pursuit for the control sequence that minimizes pre-specified cost function (maximizes the discounted reward (4)). Each suggested control action is then checked for “safety” using qualitative

knowledge about stabilizing control laws derived from the concept of CLF. If suggested control action satisfies certain criterion the action is taken, otherwise the control action is set to zero. The control action taken (indicated as u^r in Fig. 1a), at each time step, is passed back to the RL agent and together with other variables constitutes the RL agent state.

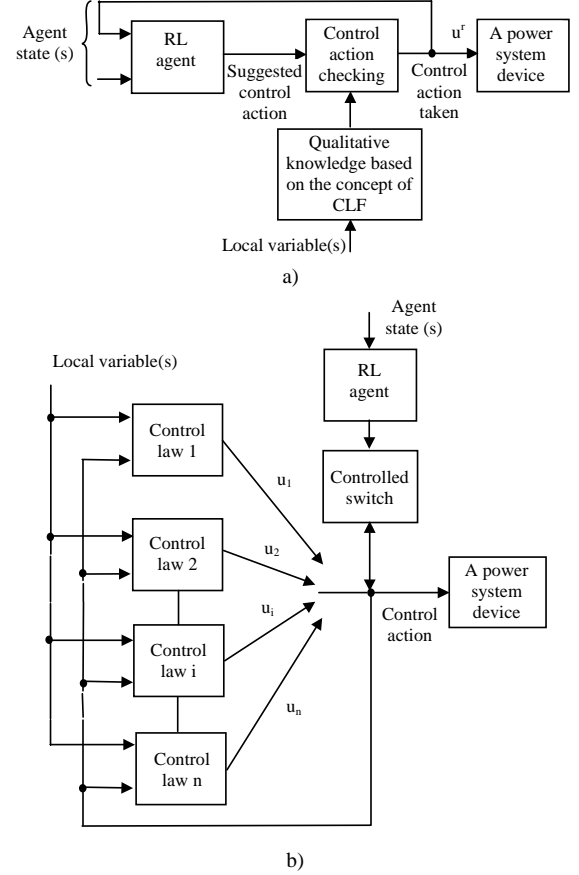


Fig. 1 Conceptual diagram of the proposed controls

In the second control scheme (Fig. 1b) the system, as a whole, reveals itself as a switched system in which a switched controller is used to control the system. Two major tasks should be accomplished as a switched controller is designed [27]: (i) design of basic controllers (or control laws), and (ii) definition of switching law of the basic controllers. To solve the first task the concept of CLF is used and for the second one a RL algorithm is employed to determine a switching sequence of the basic control laws that minimizes a pre-specified cost function. Each individual control law u_1, u_2, \dots, u_n , is derived as a stabilizing continuous control law that renders a common (global) Lyapunov function candidate decreasing.

The aim of having active control signal as an input to each of the control laws is to allow that all the control laws always track each other such that “bumpless” switch from one control law to another is assured [27]. Smoothing the control in this manner eliminates chattering that may excite high frequency dynamics.

VI. DESCRIPTION OF THE TEST POWER SYSTEM MODEL

To illustrate capabilities of the proposed control schemes to control a TCSC the four-machine power system model [19] described in Fig. 2, is used.

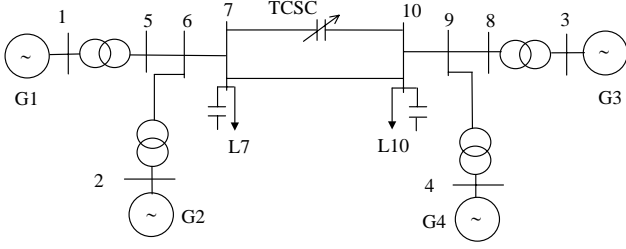


Fig. 2 A four-machine power system

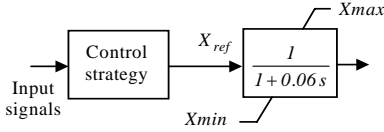


Fig. 3 Block diagram of the TCSC

All the machines are modeled with a detailed generator model, slow direct current exciter, automatic voltage regulator, and speed regulator. The loads are modeled as constant current (active) and constant impedance (reactive). The TCSC is considered as a variable reactance placed in series with a transmission line. The block diagram of TCSC considered is illustrated in Fig. 3. The control variable is X_{ref} and is supposed to belong to the interval $[-61.57, 0] \Omega$. The value of -61.57Ω for TCSC reactance corresponds approximately to 30% compensation of the line on which the device is installed. The results of two case studies are presented.

VII. CASE STUDY I

This case study focuses on how to control a TCSC using the control scheme illustrated in Fig. 1a.

A. State and reward definition

The control scheme, as a whole, is supposed to rely on strictly locally measurable information and a minimal set of a single local measurements (in addition to those required by CLF control) is chosen, namely of the active power flow through the line in which the device is installed. This quantity is obtained at each time step of $50ms$. It is used to define the rewards and pseudo-states used by the RL algorithm. To cope with partial observability of the system the approach considered in [9] is adopted and a pseudo-state from the history of the measurements and actions taken is defined. The pseudo-state at time t is defined by the following expression,

$$s_t = (P_{et}, P_{et-1}, u_{t-1}, u_t^r). \quad (11)$$

The aim of the control is to maximize damping of the electrical power oscillations in the line with as less as possible

control efforts. In this respect, the reward is defined by,

$$r_t = \begin{cases} -|P_{et} - \bar{P}_e|/10^6 - c \cdot |u_t| & \text{if } u_t = u_t^r \\ -1000 & \text{if } u_t \neq u_t^r \end{cases}. \quad (12)$$

where \bar{P}_e represents an estimate of the steady-state value of the electric power transmitted through the line.

Note that the steady state value of the electrical power is dependent on several aspects (operating point, steady state value of X_{ref} indicated in Fig. 3) and so cannot be fixed beforehand. Thus, rather than to use a fixed value of \bar{P}_e , its value is estimated on-line using the following equation,

$$\bar{P}_e = \frac{1}{1200} \sum_{i=0}^{1199} P_{et+1-i}, \quad (13)$$

which is a moving average over the last $1200 \cdot 50ms = 60s$ and provides the algorithm with some adaptive behavior.

B. The rule for suggested control actions checking

This rule is derived directly from the results discussed in Section III. The results reveal that the TCSC's reactance has to be controlled in such a way that it matches the sign of the product $V_{lm} \dot{V}_{lm}$ and the rule is formalized as follows (assuming linear continuous CLF-based control law),

$$u_t^r = \begin{cases} u_t & \text{if } u_t > 0 \text{ and } V_{lm} \dot{V}_{lm} > 0 \text{ and } u_t \geq 5 \cdot (V_{lm} \dot{V}_{lm}) \\ & \text{or} \\ & u_t < 0 \text{ and } V_{lm} \dot{V}_{lm} < 0 \text{ and } u_t \leq 5 \cdot (V_{lm} \dot{V}_{lm}) \\ 0 & \text{otherwise} \end{cases} \quad (14)$$

Although mathematically any positive gain ($k \cdot V_{lm} \dot{V}_{lm}$) should stabilize the system there are some limitations on the value of the parameter and in [11] it has been observed that the value $k \geq 5$ is to be considered (this is the reason why the term $V_{lm} \dot{V}_{lm}$ is multiplied by 5). Each "accepted" control action is passed through a low-pass filter, before being applied, with the purpose of chattering problem elimination.

C. The value of parameters

The control set is discretized in four values equal to $U = \{-61.57, -41.05, -20.54, 0\}$ while electrical power transmitted in the line is discretized in 100 values within interval $[-250, 250]MW$. The parameters are set to $\gamma = 0.98$, $\varepsilon = 0.05$ (preliminary simulations and results from [9] revealed these values as appropriate), and $c = 0.5$.

D. Simulation results

The system response to the $100ms$ duration, three-phase self-cleared short circuit near bus 7 on the lower line (Fig. 2) between the buses 7 and 10, is shown in Fig. 4 (the speed of G2 relative to G3). From Fig. 4 it is clear that the system exhibits poorly damped inter-area oscillations of $0.72Hz$.

The learning period is partitioned into different scenarios. Each scenario starts with the power system being at rest and is such that at $1s$ the self-cleared short-circuit near bus 7 occurs. The simulation then proceeds in the post-fault configuration until t is greater than $60s$. The short circuit duration is drawn at random from interval $[50-200\text{ ms}]$. The progressive learning in terms of the system response to 100 ms fault duration, is illustrated in Fig. 5a.

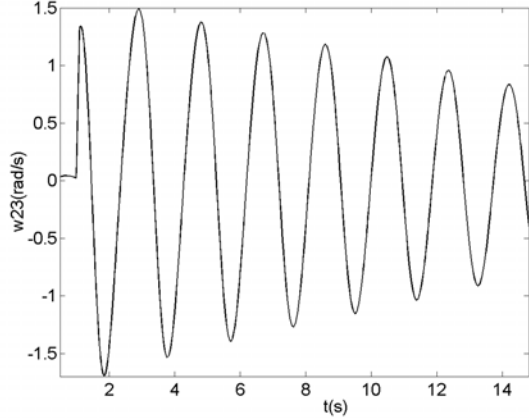


Fig. 4 The response of the uncontrolled system

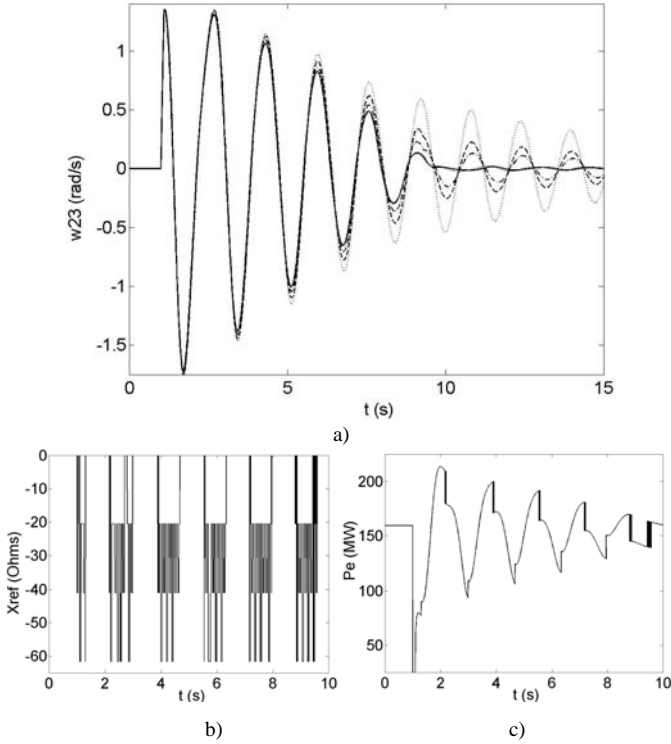


Fig. 5 The response of the controlled system (a), control actions u_t^r (b), and electrical power in the line (c)

The controlled system responses after the first learning scenario (dotted), after 100 (dot-dashed), 300 (solid), and 350 (solid) learning scenarios together with the response under a standard CLF-based linear control law (dashed, $X_{ref} = k \cdot (V_{km} \dot{V}_{km})$, $k = 5.0$), are shown.

When the fault is met for the first time, the controller

succeeds to stabilize the system but oscillations are rather poorly damped. After 100 learning scenarios the controller outperforms standard CLF-based control. The results are more pronounced after 300 and 350 scenarios. The system responses after 300 and 350 scenarios are almost identical indicating that the RL algorithm converged to a good approximation of an optimal control policy. The variation of the control variable and electrical power in the line, after 350 learning scenarios, is illustrated in Fig. 5b and Fig. 5c.

VIII. CASE STUDY II

This section illustrates the capabilities of the control scheme depicted in Fig. 1b in controlling the same device.

A. Basic control laws

For the simulation purposes four linear, continuous control laws are used (see section III). The slope of the control laws is chosen based on preliminary simulations and observations given in [11], [12], and for the control laws,

$$u = X_{ref} = F(V_{lm} \dot{V}_{lm}) = k \cdot (V_{lm} \dot{V}_{lm}), \quad (15)$$

four values of parameter $k=(5.0,6.0,7.0,8.0)$ are considered. In this setting an approximation of the optimal sequence of parameter k is to be learned. Stability guarantees hold for any sequence of k within the stability domain determined by limiting values of the control.

B. State and reward definition

To retain locality of the overall control scheme the measurements of the active power flow in the line where the TCSC is installed are again used as inputs to the RL algorithm. The pseudo-state at time t is defined as,

$$s_t = (P_{et}, P_{et-1}, P_{et-2}, k_{t-1}, k_{t-2}). \quad (16)$$

Since the aim of the control is the same as in the case study I, the reward is defined as,

$$r_t = -|P_{et} - \bar{P}_e|/10^6 - c \cdot |u_t| \quad (17)$$

The same parameters of the learning algorithm and discretization of the power flow, as in case study I, are used.

C. Simulation results

The controlled system responses when subjected to the same fault as in the previous case study, with the TCSC controlled by CLF-based control for $k=5.0$ (dotted) and $k=8.0$ (dashed) are given in Fig. 6a. In the same figure the response of the system being controlled by the proposed control scheme after 350 learning scenarios (100 ms fault duration), is presented (solid). The learning scenarios are the same as defined in the case study I. Note that the proposed control scheme considerably improves the oscillations damping and the settling time is quite smaller than in the case of standard CLF-based controls.

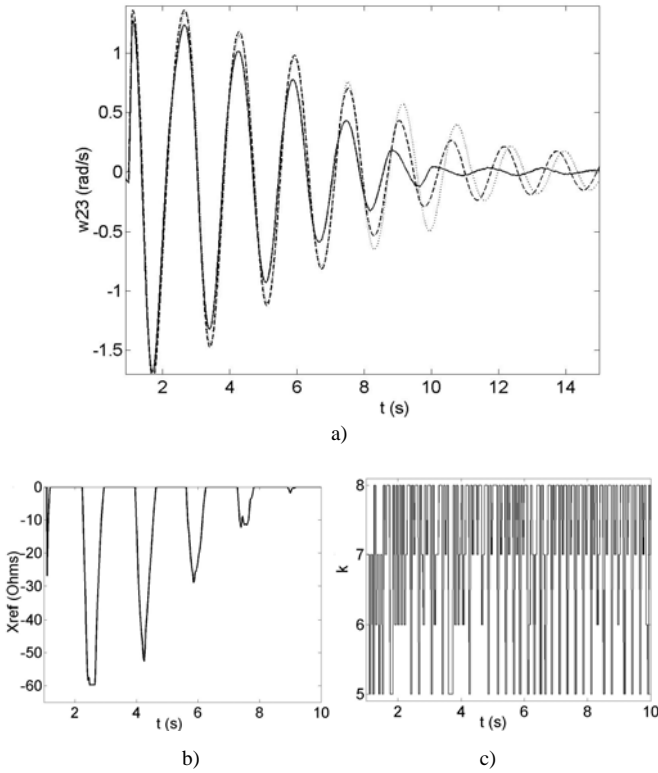


Fig. 6 The controlled system response: CLF with $k=5.0$ (dotted), $k=8.0$ (dashed), and after 350 learning scenarios (solid) (a), control efforts (b) and sequence of k (c).

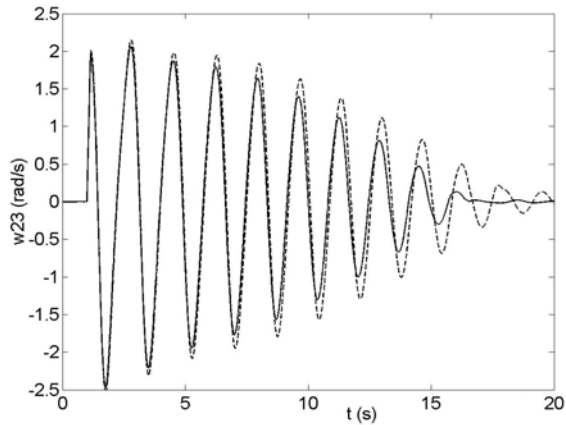


Fig. 7 The controlled system responses to an “unseen” fault

The “bumpless” non-excessive control law is achieved (Fig. 6b). Assuring “bumpless” switch from one control law to another is, in general nonlinear systems, a rather formidable task. However, in this case study it is extremely simplified due to the fact that all stabilizing control laws have the same analytical form and “bumpless” switch is simply realized through the use of a low-pass filter placed behind the switch. The sequence of the values of k is illustrated in Fig.6c.

To illustrate how the control scheme accommodates changing system operating conditions, the system is subjected to the same fault type with the fault duration increased to 120 ms and slightly modified pre-fault conditions (active power

productions of the generators G1 and G2 are increased by the amount of 10 MW). The system response is shown in Fig. 7 (dashed). The controller successfully stabilized the system. Moreover, the control scheme benefits from additional scenarios of the same type and the system response after 10 additional scenarios is shown in Fig. 7 (solid).

IX. DISCUSSION

Stability guarantees in the proposed control schemes are ensured by design, i.e. by imposing Lyapunov-based constraints on the available control choices to the RL algorithm. These guarantees hold regardless of particular RL algorithm used. The CLF-based basic control laws for the TCSC do not include any parameter, which is dependent on the network conditions, and the control response is robust with respect to system loading, network topology, and fault type and location [11]-[14]. However, the RL algorithm is used to learn an approximation of the optimal sequence of the basic control laws and to each power system configuration corresponds an optimal sequence. Thus, each time the controller faces a structural change in the system it must learn a new approximation, and may “forget” the learned policy for other system configurations. This is the reason prioritized sweeping algorithm is used because this algorithm, according to our experience [6], [9], [10], makes more efficient use of data gathered, finds better policies, and handle changes in the environment more efficiently. Prioritized sweeping RL algorithm solves the problem just partly and this is certainly the issue to be tackled in the future research. Fortunately, the research in the field of RL is very intensive and more and more powerful algorithms are at our disposal [28] and should be considered in the future research to find those able to cope with the problem more efficiently.

Observe that any meaningful cost function, by appropriate reward definition, can be used in the proposed control.

Derivation of CLF-based control laws [11]-[14] were based on domain-specific knowledge rather than Sontag’s universal formulae [29]. Moreover, by imposing Lyapunov-based constraints on available controls the domain-specific knowledge is incorporated into RL method.

X. SOME FUTURE RESEARCH DIRECTIONS

Highlighting the benefits of combining methodologies from control theory and artificial intelligence in solving power system control problems is primary aim of this paper. Some important issues are not considered and will be tackled within future research efforts.

There is a space for improvement in both CLF and RL application in power system control. Research opportunities in RL applications are discussed briefly in previous section. As of the CLF concept research possibilities may include: extension of the concept from [11]-[14] by considering recently introduced extended Lyapunov function [20] for power system, using the concept of robust CLF [16], and casting the problem as an inverse optimal control [16]. Furthermore, all control laws derived in [11]-[14], and used in

this paper, are derived without explicit account for bounds on control signals. Theoretical results given in [29] offer a sound solution to this problem and are worth further consideration. In addition, Jurdjevic-Quinn approach [22] used in [11]-[14] has a limitation [31] that stems from the fact that Lyapunov function candidate is chosen for the uncontrolled system in complete disregard of the flexibility that may be offered by the control term $g(x) \cdot u$ in (1). More systematic approaches are now becoming available to resolve this problem [31].

XI. CONCLUSIONS

An appropriate combination of the methodologies from control theory and artificial intelligence, together with the domain-specific knowledge, is a promising way to implement advanced control techniques to solve power system control problems. In this paper, it is demonstrated by the combination of the RL algorithm and the concept of the CLF. The two control schemes were presented and their capabilities illustrated on the control problem involving a TCSC for damping oscillations in the system.

The main advantages of the proposed control schemes to either standard RL or CLF-based methods alone are: by imposing Lyapunov-based constraints on the control set it has been made possible to apply RL in on-line mode and by using a RL method the system performances over a set of stabilizing control laws have been improved.

In principle, any control with stability guarantees can be combined with the RL methods and any heuristic search technique can be combined with the CLF-based control in a similar way.

ACKNOWLEDGMENT

M. Glavic and D. Ernst gratefully acknowledge the support from Belgian National Fund for Scientific Research (FNRS).

REFERENCES

- [1] P. Kokotovic, M. Arcak, "Constructive Control: A Historical Perspective", *Automatica*, Vol. 37, pp. 637-662, 2001.
- [2] Z. Artstein, "Stabilization with Relaxed Controls", *Nonlinear Analysis*, Vol. 7, No. 11, pp. 1163-1173, 1983.
- [3] E. D. Sontag, "A 'Universal' Construction of Artstein's Theorem on Nonlinear Stabilization", *System and Control Letters*, Vol. 13, No. 2, pp. 117-123, 1989.
- [4] C. Taylor (Convener), "Advanced Angle Stability Controls", CIGRE Technical Brochure No. 155, 1999.
- [5] A. G. Phadke, "Synchronized Phasor Measurements in Power", *IEEE Computer Applications in Power*, Vol. 6, No. 2, pp 10-15, 1993
- [6] M. Glavic, D. Ernst, L. Wehenkel, "Combining a Stability and a Performance Oriented Control in Power Systems", *IEEE Transactions on Power Systems*, vol. 20, no. 1, pp. 525-527, 2005.
- [7] J. Jung, C. C. Liu, S. L. Tanimoto, V. Vittal, "Adaptation in Load Shedding Under Vulnerable Operating Conditions", *IEEE Transactions on Power Systems*, Vol. 17, No. 4, pp. 1199-1205, 2002.
- [8] C. C. Liu, W. Jung, G. T. Heydt, V. Vittal, A. G. Phadge, "The Strategic Power Infrastructure Defense (SPID) System", *IEEE Control System Magazine*, Vol. 20, pp 40-52, 2000.
- [9] D. Ernst, M. Glavic, L. Wehenkel, "Power System Stability Control: Reinforcement Learning Framework", *IEEE Transactions On Power Systems*, Vol. 19, No. 1, 2004.
- [10] M. Glavic, D. Ernst, L. Wehenkel, "A Reinforcement Learning based Discrete Supplementary Control for Power System Transient Stability

- Enhancement", *International Journal of Engineering Intelligent Systems*, vol. 13, no. 2, pp. 81-89, 2005.
- [11] M. Ghandhari, G. Andersson, M. Pavella, D. Ernst, "A control strategy for controllable series capacitor in electric power systems", *Automatica*, Vol. 37, pp. 1575-1583, 2001.
- [12] M. Ghandhari, G. Andersson, I. A. Hiskens, "Control Lyapunov Functions for Controllable Series Devices", *IEEE Trans. on Power Systems*, Vol. 16, No. 4, pp. 689-694, 2001.
- [13] J. F. Gronquist, W. A. Sethares, F. L. Alvarado, R. H. Lasseter, "Power oscillations damping control strategies for FACTS devices using locally measurable quantities", *IEEE Transactions on Power Systems*, Vol. 10, No. 3, pp. 1598-1605, 1995.
- [14] M. Noroozian, M. Ghandhari, G. Andersson, J. Gronquist, I. Hiskens, "A Robust Control Strategy for Shunt and Series Reactive Compensators to Damp Electromechanical Oscillations", *IEEE Transactions on Power Delivery*, Vol. 16, No. 4, pp. 812-817, 2001.
- [15] P. J. Werbos, "Stable Adaptive Control Using New Critic Designs", [Online]. Available: <http://arxiv.org/html/adap-org/9810001>, 1998.
- [16] R. A. Freeman, P. V. Kokotovic, "Inverse Optimality in Robust Stabilization", *SIAM Journal of Control and Optimization*, Vol. 34, pp. 307-312, 1996.
- [17] V. L. Rawls, R. A. Freeman, "Adaptive Minimum Cost-Lyapunov-Descent Control of Nonlinear Systems", *In Proc. of American Control Conference*, Chicago, USA, pp. 1649-1653, 2000.
- [18] J. J. Murray, C. J. Cox, G. G. Lendaris, "Adaptive Dynamic Programming", *IEEE Trans. on Systems, Man, and Cybernetics – Part C: Applications and Reviews*, Vol. 32, No. 2, pp. 140-153, 2002.
- [19] P. Kundur, *Power System Stability and Control*, McGraw Hill, 1994
- [20] N. G. Bretas, L. F. C. Alberto, "Lyapunov Function for Power Systems with Transfer Conductances: Extension of the Invariance Principle", *IEEE Trans. on Power Systems*, Vol. 18, No. 2, pp. 769-777, 2003.
- [21] R. J. Davy, I. A. Hiskens, "Lyapunov Functions for Multimachine Power Systems with Dynamic Loads", *IEEE Trans. on Circuits and Systems – I: Fundamental Theory and Applications*, Vol. 44, No. 9, 1997.
- [22] V. Jurdjevic, J. P. Quinn, "Controllability and Stability", *Journal of Differential Equations*, Vol. 28, pp. 381-389, 1978.
- [23] R. S. Sutton, A. G. Barto, *Reinforcement Learning: An Introduction*, MIT Press, 1998.
- [24] D. Bertsekas, *Dynamic Programming and Optimal Control*, vol. 1, 2nd ed., Belmont, MA, Athena Scientific, 2000.
- [25] R. Bellman, *Dynamic Programming*, Princeton University Press, 1957
- [26] A. W. Moore, C. G. Atkeson, "Prioritized sweeping: Reinforcement learning with less data and less real time", *Machine Learning*, vol. 13, pp. 103-130, 1993.
- [27] Z. G. Li, C. Y. Wen, Y. C. Soh, "Switched Controllers and Their Applications in Bilinear Systems", *Automatica*, Vol. 37, pp. 477-481, 2001.
- [28] D. Ernst, P. Guerts, L. Wehenkel, "Tree-based batch mode reinforcement learning", *Journal of Machine Learning Research*, vol. 6, pp. 503-556, 2005.Y.
- [29] Lin, E. Sontag, "Control Lyapunov Univesal Formulae for restricted inputs", *International Journal of Control Theory and Advanced Technology*, Vol. 12, pp. 1981-2004, 1995.
- [30] T. J. Perkins, A. G. Barto, "Lyapunov Design for Safe Reinforcement Learning", *Journal of Machine Learning Research*, Vol. 3, pp. 803-832, 2002.
- [31] R. Sepulchre, M. Jankovic, P. Kokotovic, *Constructive Nonlinear Control*, Springer-Verlag, 1997.