Available online at www.sciencedirect.com
ScienceDirect
Journal of Power Sources 173 (2007) 298–309
JOURNAL OF POWER SOURCES
www.elsevier.com/locate/jpowsour

Model-based control of fuel cells (2): Optimal efficiency

Joshua Golbert, Daniel R. Lewin *

PSE Research Group, Wolfson Department of Chemical Engineering, Technion IIT, Haifa 32000, Israel

Received 26 June 2006; received in revised form 13 April 2007; accepted 29 April 2007
Available online 5 May 2007

Abstract

A dynamic PEM fuel cell model has been developed, taking into account spatial dependencies of voltage, current, material flows, and temperatures. The voltage, current, and therefore, the efficiency are dependent on the temperature and other variables, which can be optimized on the fly to achieve optimal efficiency. In this paper, we demonstrate that a model predictive controller, relying on a reduced-order approximation of the dynamic PEM fuel cell model can satisfy setpoint changes in the power demand, while at the same time, minimize fuel consumption to maximize the efficiency. The main conclusion of the paper is that by appropriate formulation of the objective function, reliable optimization of the performance of a PEM fuel cell can be performed in which the main tunable parameter is the prediction and control horizons, V and U, respectively. We have demonstrated that increased fuel efficiency can be obtained at the expense of slower responses, by increasing the values of these parameters.

Keywords: Fuel cells; Model predictive control; Fuel efficiency

1. Introduction

Fuel cell models of various levels of complexity have been suggested, describing their performance under an array of conditions [1–3]. These models have been used to evaluate optimal schemes of external heating, water management and fuel composition. The regulation of the transient response of fuel cells is important for vehicular applications, since the power demands fluctuate, and the fuel cell will not be working at the optimal steady state designed by its manufacturer. It is desirable to control the fuel cell so that acceptable response time for the power demand is ensured, while achieving high efficiencies. Dynamic models facilitate the design and testing of control systems. To this end, a dynamic empirical model for the transient response of a fuel cell was developed by Amphlett et al. [4]. This is a lumped model with no spatial dependence. Kang et al. [5] present an analysis of a dynamic model for a molten-carbonate fuel cell (MCFC) where the system is modeled as a collection of first order transfer functions with dead times. More recently, load distribution control in hybrid vehicles has been investigated [6,7], motivated by the benefits in improved dynamic response of systems combining fuel cells and a battery. In addition, control of the power conditioning unit is addressed [8] along with control of the fuel delivery system [9]. These studies typically treat the fuel cell itself as a type of disturbance and its control is not addressed. Lauzze and Chmielewski [10] attempt to perform multivariable control using separate linear PID controllers. However, the system's high degree of interaction between the different control loops is evident.

In previous work [11], we described a framework to control the fuel cell using nonlinear model-based control. We proposed a transient model, based on the ideas presented by Yi and Nguyen [2], modeling a fuel cell along its channel. Both models account for heat transfer between the solid and the two gas channels, and between the solid and cooling water. In addition, the water content is modeled, accounting for condensation and evaporation, water drag through the membrane, and water generation at the cathode. However, we account for the transients in the energy balance on the solid, with all the other equations assumed to be at quasi-steady-state in equilibrium with a given solid temperature profile. This spatially dependent model is referred to as the "full-order" model, and is used to represent the true process in closed-loop simulations. Furthermore, a first-order, time dependent model of a fuel cell has been developed, which is fast enough to use for on-the-fly optimization of

* Corresponding author. Tel.: +972 4 8292006; fax: +972 4 8295672.
E-mail address: dlewin@tx.technion.ac.il (D.R. Lewin).
URL: http://pse.technion.ac.il (D.R. Lewin).

0378-7753/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.jpowsour.2007.04.062

Nomenclature

a	solid-gas heat transfer area per unit length along channel (cm)
A	matrix for linear constraints in NLP
b	solid-coolant heat transfer area per unit length along channel (cm)
B	vector for linear constraints in NLP
C	target function weight matrix of NLP
C_p	solid heat capacity (J (g K)$^{-1}$)
D	diffusion coefficient (cm^2 s^{-1})
e	area of current per unit length along channel (cm)
f	cross-section of solid in direction of reactant flow (cm^2)
F	Faraday constant (C s^{-1})
h	channel width (cm)
ΔH	enthalpy of overall reaction (J mol^{-1})
ΔH_{vap}	enthalpy of water condensation (J mol^{-1})
I	current density (A cm^{-2})
I_0	exchange current density (A cm^{-2})
J	objective function for MPC
L	channel length (cm)
m	vector of optimization variables
M	molar flow (mol s^{-1})
P	power density (W cm^{-2})
P_j	partial pressure of species j (atm)
R	gas constant (8.314 J (mol K)$^{-1}$)
S	weight coefficient matrix for control moves
t_m	membrane thickness (cm)
T	temperature (K)
u	control variable
Δu	change in control variable value
U	heat transfer coefficient (W (cm^2 K)$^{-1}$)
V	prediction horizon
V_{cell}	cell voltage (V)
V_{OC}	open circuit voltage (V)
$\underline{\underline{W}}_1$	weight coefficient matrix for setpoint tracking
y	output variable

Greek letters

δ	length of diffusion layer (cm)
γ	minimum value for gradient
η	efficiency
φ	minimum stoichiometric ratio between hydrogen and current
Λ	waste value (mol s^{-1})
ρ	solid density (g cm^{-3})
σ	membrane conductivity (Ω cm)$^{-1}$
ω_{eff}	weight coefficient matrix of efficiency

Subscripts

a	anode
avg	average
c	cathode
cool	coolant
g	gas
H_2	hydrogen
i	time step
in	inlet
inf	surroundings
N_2	nitrogen
O_2	oxygen
oc	open circuit
s	solid
set	set point
w	water

Superscripts

l	liquid
sat	saturation
v	vapor

operating parameters to ensure convergence to required power. This reduced model was used by a model-predictive-controller (described below) to regulate the power output of the fuel cell. As demonstrated in Golbert and Lewin [11], model predictive control (MPC) relying on this model is more robust than standard linear control, especially in regions of high power density. In addition, multiple degrees of freedom can be used to improve performance.

The objective of this paper is to demonstrate that MPC can be further exploited to achieve both robust performance as well as improved fuel efficiency. We begin by reviewing the model predictive control method and explain the solution procedure. Next, we review the reduced-order fuel cell model utilized by the controller. Using this reduced model we describe the MPC formulation for fuel efficiency. Finally, we present the results obtained, comparing the performance with and without accounting for optimal efficiency. The results show that advanced model-based control can improve efficiency by utilizing the degrees of freedom in the fuel cell operation. A novel definition of the MPC target function is presented which simultaneously achieves convergence and maximizes efficiency as opposed to a trade-off between the two concerns.

2. Modeling and analysis

2.1. Description of MPC

Model predictive control [12] is part of a family of optimization-based control methods, which are based on on-line optimization of future control moves. Using a process model, the optimizer predicts the effect of past inputs on future outputs. Then, using the same model, it computes a sequence of future control moves to minimize an objective function that including penalties on the trajectory of predicted tracking error. The first of the future control moves is implemented, and the entire optimization is repeated from the next step on, and so on, ad infinitum. Feedback is used to account for the model's inaccuracies and to ensure convergence.

The work described in this paper solves the optimization problem implicit in MPC simultaneously using the approach of Biegler [13], briefly reviewed next. A typical nonlinear discrete system is defined as

$$\underline{y}_{k+1} = \underline{f}(\underline{y}_k, \underline{u}_k) \tag{1}$$

To simplify nomenclature, we assume no system memory, that is, y_{k+1} depends explicitly only on the system values at time k (i.e., no time lag), noting that dependencies on historical data can be included with no loss of generality. For such a system, a nonlinear problem is defined as

$$\begin{aligned} &\min_{\substack{\Delta\underline{u}_1,\ldots,\Delta\underline{u}_U \\ \underline{y}_1,\ldots,\underline{y}_V}} J(\Delta\underline{u}_1,\ldots,\Delta\underline{u}_U,\underline{y}_1,\ldots,\underline{y}_V) \\ &\equiv \sum_{i=1}^{V} \underline{y}_i \underline{\underline{W}}_i \underline{y}_i + \sum_{i=1}^{U} \Delta\underline{u}_i \underline{\underline{S}}_i \Delta\underline{u}_i \\ &\text{s.t.} \quad \underline{y}_{k+1} - \underline{f}(\underline{y}_k, \underline{u}_k) = 0, \quad \underline{y}_{\min} \le \underline{y}_k \le \underline{y}_{\max}\ k = 1,\ldots,V \\ &\quad \underline{u}_{\min} \le \underline{u}_k \le \underline{u}_{\max}, \quad \Delta\underline{u}_{\min} \le \Delta\underline{u}_k \le \Delta\underline{u}_{\max}\ k = 1,\ldots U \end{aligned} \tag{2}$$

where the value of the control variables is simply the sum initial values and the changes in the variables:

$$\underline{u}_k = \underline{u}_0 + \sum_{i}^{k} \Delta\underline{u}_i \tag{3}$$

The target function, J, expresses the trade-off between the convergence to a desired trajectory in the outputs and the required moves of the control variables, u. As will be shown later in this paper, the target function determines the closed loop performance of the system.

In reality, the control variables (throughout the control horizon) are the only independent variables, whereas the optimization variables are defined as the control variables, Δu, *and* the output variables, y. The apparent discrepancy in the number of degrees of freedom is resolved by the nonlinear equality constraints. These constraints ensure that the output values of the optimization solution are feasible as defined by the system equations given as Eq. (1). The equality constraints reduce the degrees of freedom of the optimization to be identical to the number of control moves.

To achieve the classic form of an NLP, the optimization variable, m, is defined as the concatenation of the changes in the control variables (over the control horizon) and the values of the output variables (over the prediction horizon): $\underline{m} = [\Delta u_1, \ldots, \Delta u_U, y_1, \ldots, y_V]^T$. The replacement of the linear bounds in Eq. (2) by linear constraints and bounds on $\underline{m}$ is straightforward resulting in the final NLP which is solved at every time step:

$$\begin{aligned} &\min_{\underline{m}} J(\underline{m}) \equiv \underline{m}^T \underline{\underline{C}} \underline{m} \\ &\text{s.t.} \quad \underline{g}(\underline{m}) = 0, \quad \underline{\underline{A}}\underline{m} \le \underline{B}, \quad \underline{m}_{\min} \le \underline{m} \le \underline{m}_{\max} \end{aligned} \tag{4}$$

2.2. Reduced-order model

The model used in the MPC framework is a time dependant, lumped parameter model of the flow channels, and membrane, including temperatures and water content. Since this model has been detailed in Golbert and Lewin [11], we state only the key equations here. Assuming quasi-steady-state for most of the state variables leaves the temperature of the solid as the only dynamic variable:

$$\begin{aligned} \frac{dT_s}{dt} &= f_{T_s}(T_s, T_{cool}, ..., I_{avg}) \\ &= \frac{1}{\rho_s C_{p_s}} \left(\frac{U_g a}{f}(T_a + T_c - 2T_s) + \frac{U_{cool} b}{f}(T_{cool} - T_s) \right. \\ &\quad - \frac{e}{f}\left(\frac{\Delta H}{2F} + V_{cell}\right) I_{avg} + \frac{1}{fL}\Delta H_{vap}(T_s) \\ &\quad \times \left(M^l_{w,a} - M^l_{w,a,in} + M^l_{w,c} - M^l_{w,c,in}\right) \\ &\quad \left. - \frac{2U_{inf}}{L}(T_s - T_{inf})\right) \end{aligned} \tag{5}$$

The Nernst equation defines the dependence of the voltage on the current density accounting for overpotentials:

$$\begin{aligned} V_{cell} &= f_{V_{cell}}(T, P_{H_2,s}, \ldots) \\ &= V^o_{OC} + \frac{RT}{2F} \log\left(\frac{P_{H_2,s}\sqrt{P_{O_2,s}}}{P_{H_2O,s}}\right) \\ &\quad + \frac{RT}{F} \log\left(\frac{I_0}{I} P_{O_2,s}\right) - \frac{I t_m}{\sigma_m} \end{aligned} \tag{6}$$

where the concentration overpotential is accounted for by defining:

$$P_{H_2,s} = \left(P_{H_2,b} - \frac{\delta I}{2F D_{H_2}}\right),$$

$$P_{O_2,s} = \left(P_{O_2,b}(x) - \frac{\delta I}{4F D_{O_2}}\right) \quad \text{and}$$

$$P_{H_2O,s} = \left(P_{H_2O,b} + \frac{\delta I}{2F D_{H_2O}}\right)$$

The solution of Eq. (6) gives the current density. The concentrations of the reactant hydrogen and oxygen and the water product are governed by simple mass balances where their production/consumption rates are dependent on the current density. The spatially dependent model is simplified to enable rapid calculation for control and optimization purposes by lumping the spatial dependencies, which results in simple algebraic equations (see Golbert and Lewin [11], for details):

$$M_{H_2} = -\frac{hL}{2F} I_{avg} + M_{H_2,in} \tag{7}$$

$$M_{O_2} = -\frac{hL}{4F} I_{avg} + M_{O_2,in} \tag{8}$$

$$T_k = T_{k,in} + \frac{U_g a L (T_s - T_k)}{\sum_i^{N_k} C_{p,i} M_i}, \quad k = a, c \tag{9}$$

$$T_{cool} = T_{cool,in} + \frac{U_{cool} b L (T_s - T_{cool})}{C_{p,w} M_{cool}} \tag{10}$$

Since the system is assumed to be at quasi-steady-state, the water vapor in each channel is assumed to be at equilibrium with liquid water if present (i.e., at the saturation pressure), in which case the amount of liquid is a balance of water entering and exiting the channel:

$$M^{\mathrm{v}}_{\mathrm{w,a}} = \frac{M_{\mathrm{H_2}}}{P_{\mathrm{a}}/P_{\mathrm{sat}}(T_{\mathrm{a}}) - 1},$$

$$M^{\mathrm{l}}_{\mathrm{w,a}} = M^{\mathrm{v}}_{\mathrm{w,a,in}} + M^{\mathrm{l}}_{\mathrm{w,a,in}} - M^{\mathrm{v}}_{\mathrm{w,a}} - \alpha\frac{hL}{F}I \tag{11a}$$

If this results in a negative value for the liquid water, then the anode is not in equilibrium, and there is no liquid water present:

$$M^{\mathrm{v}}_{\mathrm{w,a}} = M^{\mathrm{v}}_{\mathrm{w,a,in}} + M^{\mathrm{l}}_{\mathrm{w,a,in}} - \alpha\frac{hL}{F}I, \qquad M^{\mathrm{l}}_{\mathrm{w,a}} = 0 \tag{11b}$$

The water content at the cathode is similarly calculated including the expression for water generation from the reaction. If the cathode water is at equilibrium:

$$M^{\mathrm{v}}_{\mathrm{w,c}} = \frac{M_{\mathrm{O_2}} + M_{\mathrm{N_2}}}{P_{\mathrm{c}}/P_{\mathrm{sat}}(T_{\mathrm{a}}) - 1},$$

$$M^{\mathrm{l}}_{\mathrm{w,c}} = M^{\mathrm{v}}_{\mathrm{w,c,in}} + M^{\mathrm{l}}_{\mathrm{w,c,in}} - M^{\mathrm{v}}_{\mathrm{w,c}} + \alpha\frac{hL}{F}I + \frac{hL}{2F}I \tag{12a}$$

Otherwise there is no liquid water present:

$$M^{\mathrm{v}}_{\mathrm{w,c}} = M^{\mathrm{v}}_{\mathrm{w,c,in}} + M^{\mathrm{l}}_{\mathrm{w,c,in}} + \alpha\frac{hL}{F}I + \frac{hL}{2F}I,$$

$$M^{\mathrm{l}}_{\mathrm{w,c}} = 0 \tag{12b}$$

In Eqs. (11a), (11b), (12a) and (12b), α, the water drag ratio, is a function of the temperature, pressures and water content [14]. Since the water content is, in turn, a function of α Eqs. (11a), (11b), (12a) and (12b) define an implicit equation, the solution of which results in the value of α:

$$\alpha = f_\alpha(T_{\mathrm{a}}, T_{\mathrm{c}}, M^{\mathrm{v}}_{\mathrm{w,a}}(\alpha), M^{\mathrm{v}}_{\mathrm{w,c}}(\alpha), P_{\mathrm{a}}, P_{\mathrm{c}})$$
$$= f_\alpha(T_{\mathrm{a}}, T_{\mathrm{c}}, \alpha, P_{\mathrm{a}}, P_{\mathrm{c}}) \tag{13}$$

2.3. Definition of optimization problem

In this paper we will address a system where the control variables are the dry hydrogen flow rate (assumed to be humidified), the coolant temperature and the average current density. Originally, the system was defined using the solid temperature, cell voltage and water drag ratio, α:

$$\underline{u}_i \equiv \begin{bmatrix} M_{\mathrm{H_2},i} \\ T_{\mathrm{cool},i} \\ I_{\mathrm{avg},i} \end{bmatrix}, \qquad \underline{y}_i \equiv \begin{bmatrix} T_{\mathrm{s},i} \\ P_i \\ \alpha_i \end{bmatrix} \tag{14}$$

2.3.1. Objective function

In Golbert and Lewin [11] we demonstrated that nonlinear MPC can satisfy changes in load demands robustly. We showed that the usage of multiple manipulated variables can improve the response time of the system, with the target function to be minimized being the sum of square errors from the setpoint, with a penalty on the moves required in the manipulated variables. Clearly, there is potential for the optimizer to exploit the degrees of freedom inherent in the fuel cell design to improve the fuel-efficiency. In this regard, efficiency is defined as the ratio between the actual power produced and the heat of formation of the water produced if all the hydrogen feed is consumed:

$$\eta = \frac{hLP}{\Delta H M_{\mathrm{H_2}}} \tag{15}$$

Since the MPC solves a minimization problem, a waste variable is defined, $1-\eta$, which is to be minimized by the optimizer. The most obvious way to improve the efficiency is to lower the feed flow rate. This is offset with the need for sufficient hydrogen concentration to achieve satisfactory voltage. This is the reason that the controller will not reduce the hydrogen too much, since the concentration overpotential will become unbearable and compromise the power output. The optimization problem is now defined as the weighted sum of the performance (the difference between the set point and the actual power), the size of the control steps and the local efficiency over time:

$$J(\underline{\Delta u}, \underline{y}) = (1-\omega_{\mathrm{eff}})(\underline{P} - \underline{P_{\mathrm{set}}})^{\mathrm{T}}\underline{\underline{W}}_1(\underline{P} - \underline{P_{\mathrm{set}}})$$
$$+ \omega_{\mathrm{eff}}(1-\underline{\eta})^{\mathrm{T}}(1-\underline{\eta}) + \underline{\Delta u}^{\mathrm{T}}\underline{\underline{S}}\underline{\Delta u} \tag{16}$$

where the power density, $\underline{P}$, is not calculated, but merely is a component of vector $\underline{y}$. The parameter ω_{eff} expresses the desired trade-off between performance and efficiency, while the coefficients of the diagonal matrix $\underline{\underline{W}}_1$ are selected to scale the terms in proportion to their importance.

Unfortunately, the definition of efficiency in Eq. (15) can lead to numerical problems, since it involves division by the hydrogen flow rate, which at low values, will lead to excessively large objective function gradients, and at high values to gradients in the efficiency contribution to J that approach zero. Consequently, either the optimizer will give excessive attention to efficiency, or will completely ignore it, depending on the local value of the hydrogen flow rate. To avoid these problems, the following expression for the efficiency loss, $\underline{\Lambda}$, is used for the optimization:

$$\underline{\Lambda} = \underline{M_{\mathrm{H_2}}} - \frac{hL}{\Delta \mathrm{H}}\underline{P} \tag{17}$$

The second term in Eq. (17) is the minimum hydrogen necessary for the actual amount of power produced based on the change in enthalpy due to the oxidation of hydrogen. Note that if the hydrogen flow rate exactly matches the amount needed to satisfy the power demand, $\underline{\Lambda} = 0$, and the influences of the power and hydrogen feed are retained in this formulation of the system inefficiency. Using this modified loss efficiency term, the objective function is now:

$$J(\underline{\Delta u}, \underline{y}) = (1-\omega_{\mathrm{eff}})(\underline{P} - \underline{P_{\mathrm{set}}})^{\mathrm{T}}\underline{\underline{W}}_1(\underline{P} - \underline{P_{\mathrm{set}}})$$
$$+ \omega_{\mathrm{eff}}\underline{\Lambda}^{\mathrm{T}}\underline{\underline{W}}_2\underline{\Lambda} + \underline{\Delta u}^{\mathrm{T}}\underline{\underline{S}}\underline{\Delta u} \tag{18}$$

In Eq. (18) it is noted that the coefficients of the diagonal matrix $\underline{\underline{W}}_2$ are set to 10^{10} to scale the second term with the others.

2.3.2. Equality constraints

The equality constraints, representing the discrete system, include the difference equations for the solid temperature based on Eq. (5) for each of the V time steps in the prediction horizon:

$$g_i(\underline{u}_i, \underline{y}_i) \equiv T_{s,i+1} - T_{s,i} - f_{T_s}(T_{s,i}, V_{cell,i}, \alpha_i, \ldots)\Delta t = 0, \quad i = 0, 1 \ldots, V-1 \tag{19}$$

Note that the numerical solution of Eq. (5) is implemented using the explicit Euler method, although increased accuracy is possible by invoking higher-order methods. Note that for simplicity, all variables assumed to be constant (for example the channel pressures and inlet temperatures) are not noted in the equations. The equality constraint for the power density requires that the assumed value for the power equals the calculated value, where the voltage, $V_{cell,i}$, is calculated using Eq. (6):

$$g_2(\underline{u}_i, \underline{y}_i) \equiv P_i - I_i V_{cell,i} = 0, \quad i = 0, 1 \ldots, V-1 \tag{20}$$

The nonlinear equation (13) for α for each time step is also represented by a constraint:

$$g_3(\underline{u}_i, \underline{y}_i) \equiv \alpha_i - f_\alpha(T_{s,i}, \alpha_i, \ldots) = 0 \tag{21}$$

It is well known that the fuel cell power–current curve exhibits a maximum value, while the voltage–current curve is nonlinear but monotonically decreasing. As a result of these characteristics, it is desirable to limit the operation region of fuel cells to always lie to the left of the peak in power, since this will ensure operating at high voltage and higher efficiencies. Unfortunately, at high power demands, the system can cross the maximum, and as was shown by Golbert and Lewin [11], model inaccuracies can lead to mistaken predictions of the sign of the gain between the power and current density. A controller relying on this prediction would lose the ability to control the system. On the other hand, using the voltage as a state variable is advantageous since the voltage–current curve is always monotonous. Therefore, even if the reduced model is inaccurate, the gain between the current and voltage is always negative. In this case, the target function calculates the predicted power by multiplying the voltage and current before comparing the result to the desired set point of the power:

$$J(\underline{\Delta u}, \underline{y}) = (1-\omega_{eff})(\underline{I_{avg}}\,\underline{V_{cell}} - \underline{P_{set}})^T \underline{\underline{W}}_1 (\underline{I_{avg}}\,\underline{V_{cell}} - \underline{P_{set}}) + \omega_{eff}\underline{\Lambda}^T \underline{\underline{W}}_2 \underline{\Lambda} + \underline{\Delta u}^T \underline{\underline{S}}\,\underline{\Delta u} \tag{22}$$

Again, I_{avg} and V_{cell} are not calculated in Eq. (22), rather simply extracted from $\underline{\Delta u}$ and $\underline{y}$, respectively.

The nonlinear equality constraint replacing Eq. (20) in this case is

$$g_2(\underline{u}_i, \underline{y}_i) \equiv V_{cell,i} - f_{V_{cell}}(T_{s,i}, M_{H_2,i}, \ldots) = 0, \quad i = 0, 1 \ldots, V-1 \tag{23}$$

2.3.3. Inequality constraints

The changes in the control variables, rather than the actual values themselves, are used for the optimization, so a number of linear constraints are necessary. First, each control variable has a maximum and minimum value. Since the actual variables for the optimization problem are defined in terms of changes from the previous value, we require, for each time step, that the sum of the initial value and the sum of all the preceding steps not exceed either the limits for each variable.

Another constraint places a lower limit on the ratio between the inlet fuel flow rate and the current density drawn from the fuel cell. Again, since the optimization variables are the changes in the control variables this requirement translates into linear inequality constraints. For specific details regarding these constraints see Appendix A.

3. Results and discussion

3.1. Trade-off between fuel efficiency and performance

We present results showing the closed-loop response subject to a setpoint change from 0.3 to 0.5 W cm^{-2} and back again. The control and prediction horizons are five and six time steps, respectively, using a time step of 0.5 s. As can be seen, the temperature of the coolant is damped as is the hydrogen inlet flow rate. The output variables are defined as the solid temperature and power density. In this example, the temperature is not controlled at all, so its weight in $\underline{\underline{W}}_1$ is set to zero. The selection of the coefficients of the scaling matrices enable the tuning of the overall performance, as will be seen. The MPC tuning parameters used in this study are listed in Table 1. The parameters defining the PEM system simulated are identical to those used in Golbert and Lewin [11].

Fig. 1 shows the closed-loop response obtained when the only desire is to ensure robust convergence to the desired setpoint, with the value of ω_{eff} set to zero. As can be seen, the initial efficiency is 15%, dropping to 11% as the power approaches 0.5 W cm^{-2}, due to the increase in the hydrogen inlet flow rate. Subsequently, the current is dropped to lower the power density to 0.3 W cm^{-2}, but the fuel flow rate is kept high and the efficiency drops to 6.5%. In contrast, Fig. 2 shows the response obtained when the efficiency is taken into consideration by setting ω_{eff} to 0.2. The first thing to notice is that up to 5 s the system is at steady state with an offset in the power density. This is due to the trade-off in the optimization target function between convergence and efficiency (note that the initial efficiency in Fig. 2 is greater that that of Fig. 1). In the same spirit, the efficiency

Table 1
MPC tuning parameters

Parameter	Values
$\underline{\underline{W}}_1$	Since the only output variable specified in the objective function is the tracking error from the power set point, the only relevant entry scales this error. A value of 5 is used
$\underline{\underline{W}}_2$	10^{10} on the main diagonal
$\underline{\underline{S}}$	The coefficients for the three manipulated variables are 30, 100, and 10, respectively
Δt	0.5 s
U	5
V	6

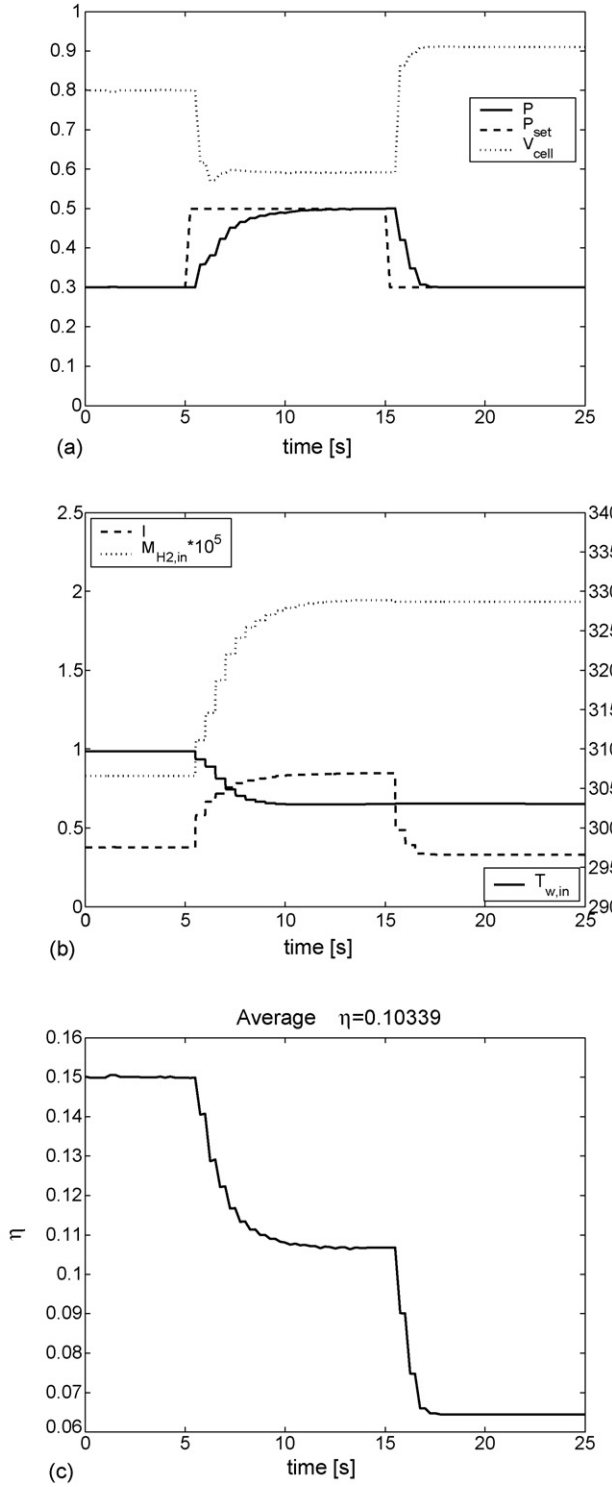

Fig. 1. Performance with no weight on efficiency and no effective limit on cell voltage: (a) Power and voltage trajectories; (b) manipulated variables; (c) instantaneous efficiency, η.

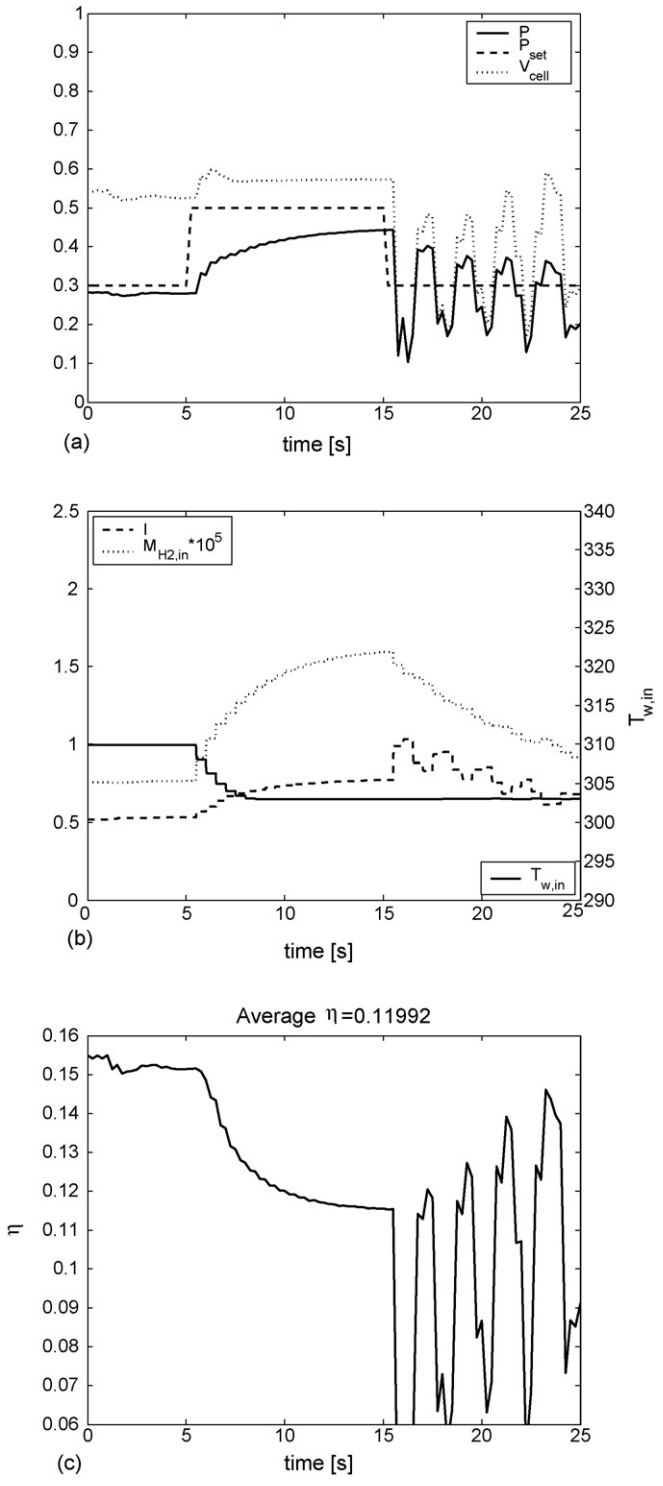

Fig. 2. Performance with weight on efficiency ($\omega_{\text{eff}} = 0.2$) and no effective limit on cell voltage: (a) Power and voltage trajectories; (b) manipulated variables; (c) instantaneous efficiency, η.

between 5 and 15 s is greater in Fig. 2 at the expense of slower convergence to the desired setpoint. This is achieved by using a lower fuel flow rate than the previous example.

Starting from $t = 15$, however, the fuel flow rate is lowered in an attempt to improve the efficiency, although it is obvious that the tracking ability of the controller is severely hampered. Close examination of the plot of the current density shows that the controller is *increasing* the current density in an attempt to *lower* the power density—meaning that the controller, either due to model inaccuracies or changes in the system conditions, has found itself on the right of the peak in the power–current plot (see Fig. 3).

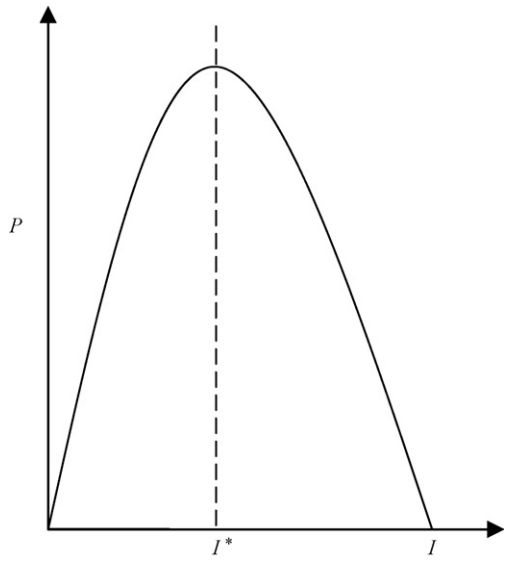

Fig. 3. Schematic power-current plot.

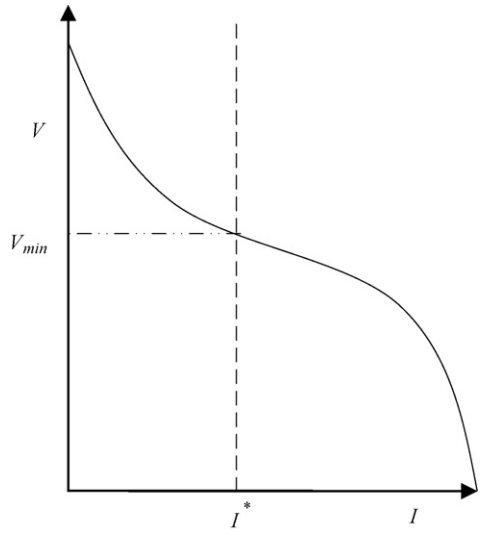

Fig. 4. Schematic voltage-current plot.

In this case, the controller will raise the current density in order to lower the power, as required by the change in P_{set}, in Fig. 2. This occurrence has been observed primarily when the setpoint is lowered (as opposed to being raised) since both considerations (convergence and efficiency) are satisfied by lowering the fuel flow rate. Since both criteria tend to lower the fuel flow rate, this can lead to excessive changes, which change the power–current curve and "trap" the system to the right of the peak. When the setpoint is raised, on the other hand, the convergence criterion tends to increase the fuel flow rate while the efficiency criterion tends to lower it. This conflict of interest moderates the change in the fuel flow rate. In an attempt to protect the system from crossing the maximum power density, we tried defining a lower limit for the voltage and introducing this as a boundary for the NLP (remember that the voltage itself is one of the optimization variables). The rational for this approach can be explained by observing the schematic diagrams of the current–power and current–voltage plots in Figs. 3 and 4.

In essence, the requirement is that the current not exceed the value corresponding to the peak of the power plot, I^*. Since I^* corresponds to a specific voltage, limiting the actual voltage to be above a limiting voltage, V_{min}, will limit the current to the left of the power peak. The results using this approach, for the case identical to that of Fig. 2, with the cell voltage limited to values above 0.5 V, are shown in Fig. 5.

It is clear that this approach does not solve the problem of crossing the power peak. Closer examination of the curves in Fig. 6, shown for varying amounts of hydrogen illustrate why this is so. Obviously an infinite number of hydrogen/current combinations correspond to a given voltage. Thus, setting a limit on the voltage implies no constraint on the current, and as seen in Fig. 5, the system can indeed jump to the left side of the power peak.

Although it is theoretically possible to define settings that could result in better performance, this is not a systematic approach and cannot be relied upon as a satisfactory solution. A more rigorous solution is to define a nonlinear constraint

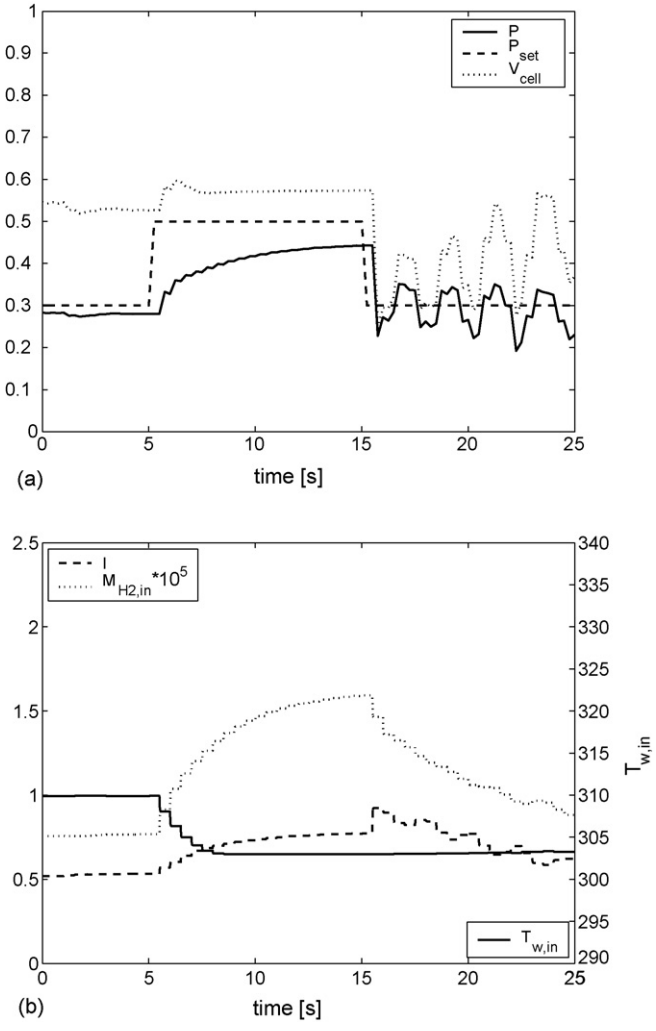

Fig. 5. Performance with weight on efficiency ($\omega_{eff} = 0.2$) and effective limit on cell voltage ($V_{cell} > 0.5$): (a) Power and voltage trajectories; (b) manipulated variables.

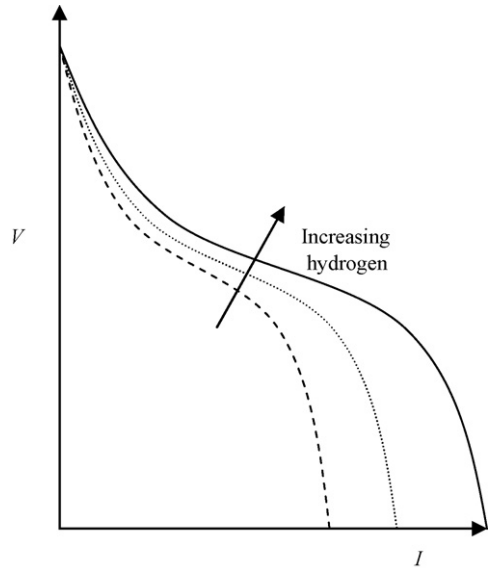

Fig. 6. Schematic voltage-current plot for varying hydrogen content.

whereby the partial derivative of the power to the current is above a set value:

$$\frac{\partial P}{\mathrm{d}I_{\mathrm{avg}}} > \gamma \tag{24}$$

By definition, Eq. (24) ensures that, to the controller's best knowledge, the system is always to the left of the power–current peak in Fig. 7. Clearly, if the controller model were perfectly accurate, it would be sufficient to ensure that the gradient in Eq. (24) is positive ($\gamma = 0$). However, due to model inaccuracies, a larger value for γ is used (we have used $\gamma = 0.2$). Fig. 8 shows the obtained closed-loop response when using the constraint in Eq. (24). Comparing the overall efficiency (12.8%) in Fig. 8 to that of Fig. 1 (10.3%) we see that using the efficiency in the NLP's target function results in a 25% efficiency improvement.

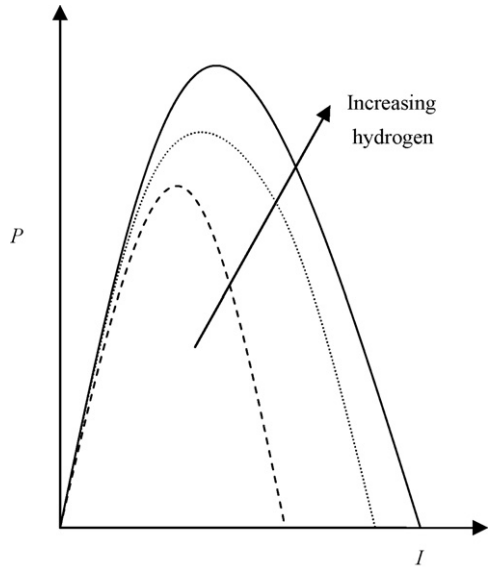

Fig. 7. Schematic power-current plot for varying hydrogen content.

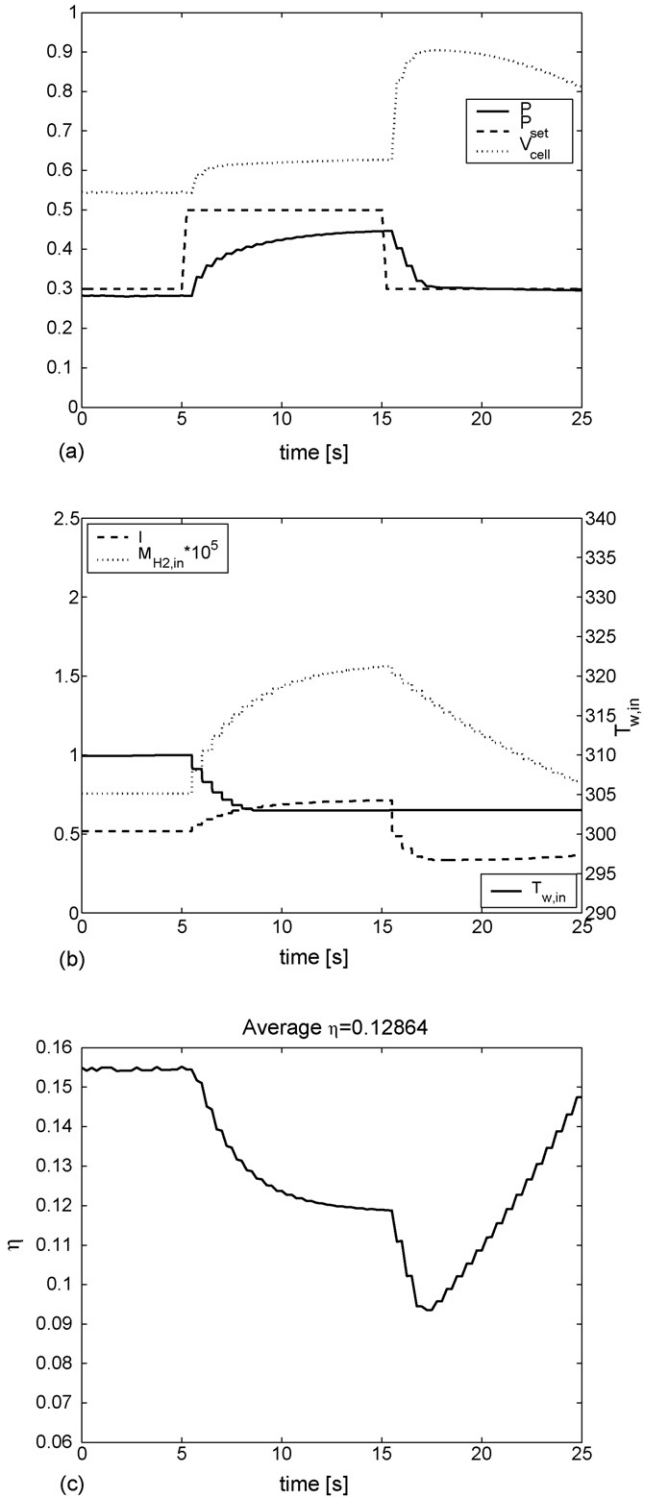

Fig. 8. Performance with weight on efficiency ($\omega_{\mathrm{eff}} = 0.2$) and limit on $\partial P/\mathrm{d}I_{\mathrm{avg}} > 0.2$: (a) Power and voltage trajectories; (b) manipulated variables; (c) instantaneous efficiency, η.

3.2. Ensuring offset-free response

Although the increase in efficiency is encouraging, the offset from the desired power setpoint is troubling. This is inherent in the formulation of the NLP, since at steady

state, $\Delta u = 0$ and the optimal solution is the minimal value of:

$$J(\underline{\Delta u}, \underline{y}) = (1 - \omega_{\text{eff}})(\underline{P} - \underline{P_{\text{set}}})^{\text{T}} \underline{\underline{W}}_1 (\underline{P} - \underline{P_{\text{set}}}) + \omega_{\text{eff}} \underline{\Lambda}^{\text{T}} \underline{\underline{W}}_2 \underline{\Lambda} \quad (25)$$

Is the case of no efficiency consideration $\omega_{\text{eff}} = 0$ and the optimal solution is complete offset-free convergence. However for any other value of ω_{eff}, the final value will always display offset since the optimum will be some trade-off between the two elements of Eq. (25).

The crux of the problem is the desire to satisfy two – often conflicting – criteria, convergence and efficiency. One attempt to solve this problem was to add extra weight to the last line of the $\underline{\underline{W}}_1$ matrix. With the intention of forcing the optimizer to converge by the end of the prediction horizon. The expectation was that the optimizer would find the fuel efficient path to convergence by the end of the prediction horizon, if not before that.

However, although the optimizer *did* propose sequences that converged at the end of the prediction horizon, it is important to recall that MPC, unlike optimal control, actually implements *the first step only*. Thus, at some point, the optimizer decided that, as defined by Eq. (18), it is best to do nothing at first, reap the benefits of efficiency for a while and only later on in the control horizon bring the system to convergence. If the system is at steady state, and the optimizer has decided that no control moves will be made in the following time step, the subsequent optimizations will be the same, ad infinitum, even with offset from the desired setpoint. Thus, even though the open loop NLP solution predicts offset-free convergence, closed loop performance will have offset.

In searching for a formulation of the NLP that will avoid this scenario, we realized that it is crucial to prevent the possibility of the optimal solution's first moves to be zero. This can be prevented by formulating a target function that consists only of a penalty for control moves, $\underline{\Delta u}^{\text{T}} \underline{\underline{S}} \underline{\Delta u}$, and imposing a constraint on convergence. Since a large number of small control moves incur a smaller penalty than a small number of large control moves, the optimizer will tend to use all of the moves available to it including the first step. With this in mind we defined the target function as follows:

$$J(\underline{\Delta u}, \underline{y}) = \omega_{\text{eff}} \Lambda(V)^2 + \underline{\Delta u}^T \underline{\underline{S}} \underline{\Delta u} \quad (26)$$

In addition to the equality constraints, Eqs. (19), (21) and, (23), we require that:

$$P(V) - P_{\text{set}} = 0 \quad (27)$$

In Eq. (26) only the value of the waste variable, Λ, at time step V (the end of the prediction horizon) appears, and the target function is dominated by the penalty on the control moves. Satisfying Eq. (27) ensures that the optimizer plans a sequence of control moves that converges at the end of the prediction horizon to the power setpoint.

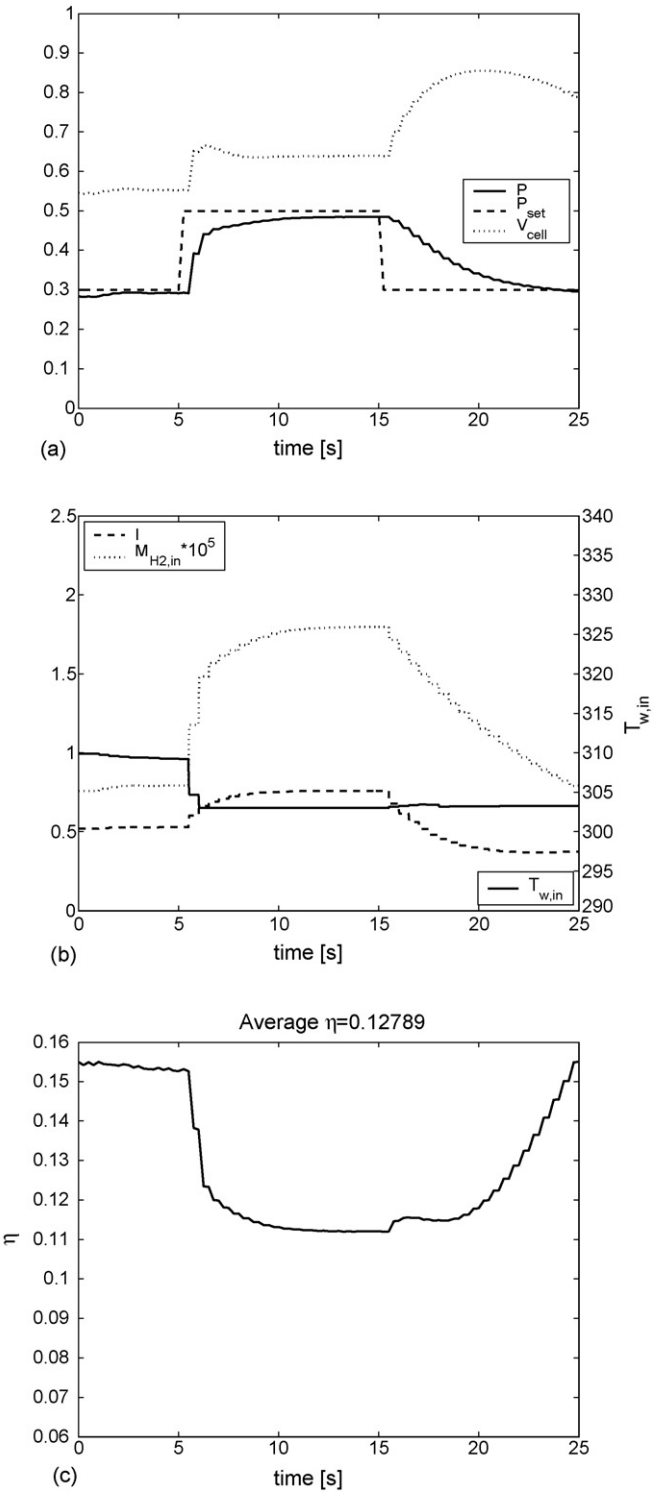

Fig. 9. Performance with weight on final efficiency, $U = 5$, $V = 6$, ($\omega_{\text{eff}} = 0.2$) and limit on $\partial P/\text{d}I_{\text{avg}} > 0.2$: (a) Power and voltage trajectories; (b) manipulated variables; (c) instantaneous efficiency, η.

To understand the closed-loop behavior when using this formulation, we examine the case where $\omega_{\text{eff}} = 0$. In this case, the solution seeks to minimize the penalty on the control moves and results in a monotonous sequence of control moves, achieving convergence at the end of the

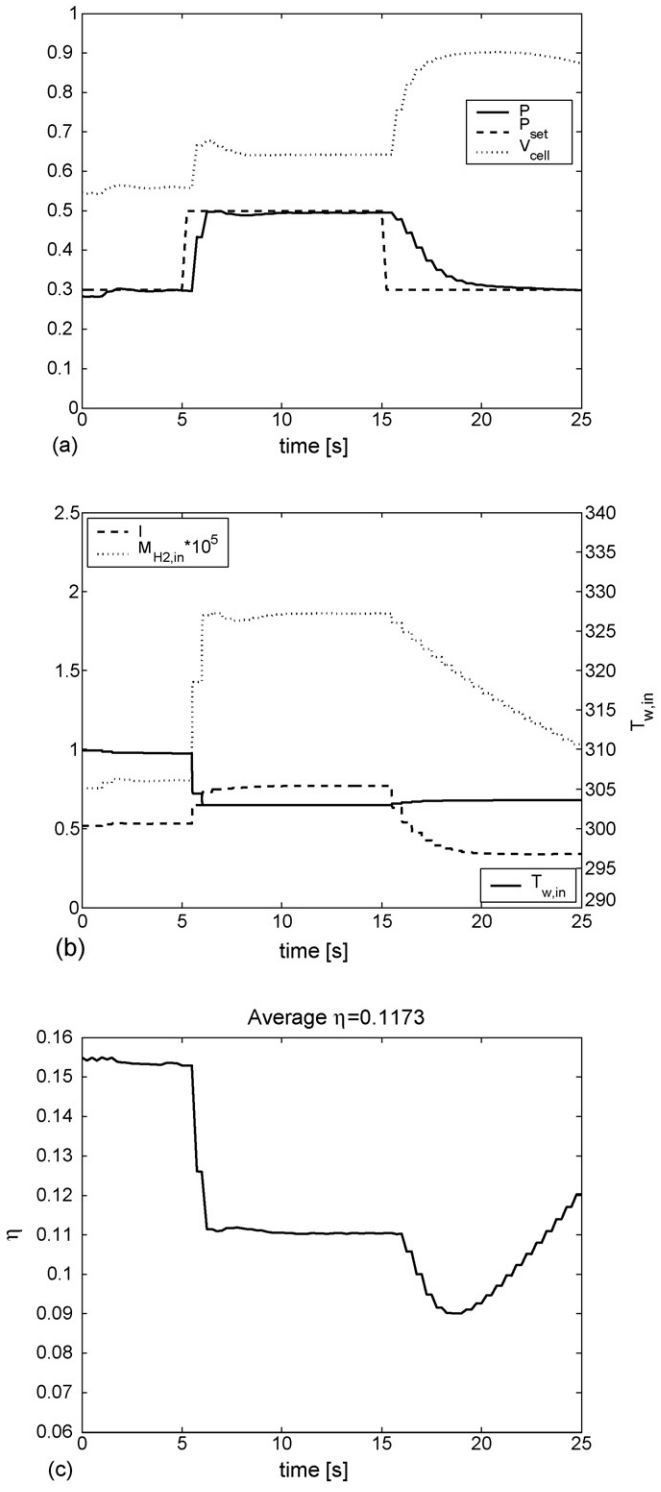

Fig. 10. Performance with weight on final efficiency $U=3$, $V=4$, ($\omega_{\text{eff}}=0.2$) and limit on $\partial P/\mathrm{d}I_{\text{avg}} > 0.2$: (a) Power and voltage trajectories; (b) manipulated variables; (c) instantaneous efficiency, η.

control horizon. It is important to note that for systems with more inputs than outputs there are numerous combinations of control that satisfy the convergence. For example, in Fig. 1(b), the system converges to a power setting of $0.3\,\mathrm{W\,cm^{-2}}$ twice, initially with low hydrogen flow rate and subsequently using a higher hydrogen flow rate and a lower current.

For $\omega_{\text{eff}}=0$, any of the possible steady states is satisfactory. If ω_{eff} is non-zero, however, the system will converge to the steady state that satisfies Eq. (27) and has the highest efficiency. Unlike before, however, since the target function is not affected by the efficiency during the prediction horizon the optimal control sequence is monotonous and the discrepancy between the open and closed loop performance is avoided.

To avoid defining unfeasible constraints, Eq. (27) is combined with the objective function of Eq. (26), giving

$$J(\underline{\Delta u}, \underline{y}) = \varphi(P(V) - P_{\text{set}})^2 + \omega_{\text{eff}}\Lambda(V)^2 + \underline{\Delta u}^{\mathrm{T}}\underline{\underline{S}}\underline{\Delta u} \qquad (28)$$

where φ is a very large weight (in the following example, $\varphi=1000$). As φ approaches infinity, we approach the original formulation of Eqs. (26) and (27). Fig. 9 illustrates the performance obtained using this new formulation, for the same conditions as those in Fig. 8.

Comparing these results with Fig. 8 clearly shows the improved convergence with very little offset (which would be completely eliminated if φ were infinite). Also note that the overall efficiencies for both cases are almost identical. Another advantage of this formulation is its ease of tuning. Here, the coefficients of $\underline{\underline{S}}$ have no affect on convergence to the desired setpoint, but only affect the degree to which each of the manipulated variables is utilized. Likewise, since the convergence is ensured as a constraint, there is no trade-off between convergence and efficiency, so tuning ω_{eff} is also irrelevant. Instead, the trade-off between the speed of convergence and efficiency is tuned only by determining the length of the prediction horizon, V. This is due to the fact that if V is large, there is plenty of time to converge, so the optimizer can allow itself to aim for an efficient solution. If, however, V is shorter, all resources are marshalled to converging quickly at the expense of efficiency. Fig. 10 shows the performance of a system identical to that in Fig. 9, with the control and prediction horizons reduced to 3 and 4, respectively. As expected, the convergence to the desired setpoint exhibited in Fig. 10 is faster than that in Fig. 9, at the expense of efficiency (11.7% compared with 12.8%).

4. Conclusions

As has been previously established, model-based control scheme of a PEM fuel cell, relying on a reduced-order, nonlinear model of the process, can be used for robust regulation. In addition, as demonstrated in this contribution, since the controller adjusts a number of manipulated variables, it takes advantage of all of the degrees of freedom to simultaneously satisfy power demands while optimizing the fuel efficiency of the entire system. The use of appropriate constraints results in significant improvements in fuel efficiency. We have shown that by appropriate formulation of the objective function, reliable optimization of the performance of a PEM fuel cell can be performed in

which the main tunable parameter is the prediction and control horizons, V and U, respectively. We have demonstrated that increased fuel efficiency can be obtained at the expense of slower responses, by increasing the values of these parameters.

Acknowledgements

This research was supported by the Israel Science Foundation (Grant No. 25/03-15.4). The support of a Rieger Foundation Scholarship to Josh Golbert for the academic year 2002–2003 is acknowledged with thanks.

Appendix A. Linear inequality constraints

Each control variable is bounded by a maximum value allowed:

$$\begin{bmatrix} u_{1,k+1} \\ \vdots \\ u_{1,k+U} \\ \vdots \\ u_{n,k+1} \\ \vdots \\ u_{n,k+U} \end{bmatrix} \leq \begin{bmatrix} u_{\max,1} \\ \vdots \\ u_{\max,1} \\ \vdots \\ u_{\max,n} \\ \vdots \\ u_{\max,n} \end{bmatrix} \tag{29}$$

However

$$\begin{bmatrix} u_{1,k+1} \\ \vdots \\ u_{1,k+U} \\ \vdots \\ u_{n,k+1} \\ \vdots \\ u_{n,k+U} \end{bmatrix} = \begin{bmatrix} u_{1,k} \\ \vdots \\ u_{1,k} \\ \vdots \\ u_{n,k} \\ \vdots \\ u_{n,k} \end{bmatrix} + \begin{bmatrix} 1 & 0 & \cdots & & & & 0 \\ \vdots & \ddots & \ddots & & & & \vdots \\ 1 & \cdots & 1 & 0 & \cdots & & \vdots \\ 0 & \cdots & 0 & \ddots & \ddots & & \vdots \\ \vdots & \cdots & \cdots & 0 & 1 & 0 & 0 \\ \vdots & \cdots & \cdots & 0 & \vdots & 1 & \vdots \\ 0 & \vdots & \cdots & 0 & 1 & \cdots & 1 \end{bmatrix}$$

$$\cdot \begin{bmatrix} \mathrm{d}u_{1,k+1} \\ \vdots \\ \mathrm{d}u_{1,k+U} \\ \vdots \\ \mathrm{d}u_{n,k+1} \\ \vdots \\ \mathrm{d}u_{n,k+U} \end{bmatrix} \equiv \begin{bmatrix} u_{1,k} \\ \vdots \\ u_{1,k} \\ \vdots \\ u_{n,k} \\ \vdots \\ u_{n,k} \end{bmatrix} + \underline{\underline{A_1}}\,\underline{\mathrm{d}u} \tag{30}$$

Hence,

$$\underline{\underline{A_1}}\,\underline{\mathrm{d}u} \leq \begin{bmatrix} u_{\max,1} \\ \vdots \\ u_{\max,1} \\ \vdots \\ u_{\max,n} \\ \vdots \\ u_{\max,n} \end{bmatrix} - \begin{bmatrix} u_{1,k} \\ \vdots \\ u_{1,k} \\ \vdots \\ u_{n,k} \\ \vdots \\ u_{n,k} \end{bmatrix} = \underline{b_1} \tag{31}$$

In a similar fashion linear constraints are defined for the lower boundaries:

$$-\underline{\underline{A_1}}\,\underline{\mathrm{d}u} \leq -\left(\begin{bmatrix} u_{\min,1} \\ \vdots \\ u_{\min,1} \\ \vdots \\ u_{\min,n} \\ \vdots \\ u_{\min,n} \end{bmatrix} - \begin{bmatrix} u_{1,k} \\ \vdots \\ u_{1,k} \\ \vdots \\ u_{n,k} \\ \vdots \\ u_{n,k} \end{bmatrix} \right) = \underline{b_2} \tag{32}$$

So far, the constraints ensure that the maximal values of the variables will not be exceeded at any step. If the fuel flow rate and the current density are to be manipulated there is a danger of the optimizer requesting an infeasible current density (above the limiting current density, which is largely influenced by concentration overpotential). Thus, for the sake of feasibility (as long as the current density is the input to the fuel cell model) a minimum ratio between the fuel and the current density must be enforced at all times:

$$M_{\mathrm{H_2}} \geq \varphi \frac{hL}{2F} I = \phi I \tag{33}$$

where φ is a tunable variable, which, in essence, ensures sufficient saturation of hydrogen. Translating from the values of u to the changes in u at each step and defining:

$$\underline{A_3} \equiv \begin{bmatrix} -1 & 0 & 0 & 0 & \cdots & 0 & \phi & 0 & 0 \\ \vdots & \ddots & 0 & \vdots & \ddots & \vdots & \vdots & \ddots & 0 \\ -1 & \cdots & -1 & 0 & \cdots & 0 & \phi & \cdots & \phi \end{bmatrix}$$

and

$$\underline{b_3} \equiv (u_{M_{\mathrm{H_2}},k} - \phi u_{I,k}) \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

gives

$$\underline{A_3}\,\underline{\mathrm{d}u} \leq \underline{b_3} \tag{34}$$

Note that when defining the matrix, the actual indices depend on the number of input variables being used.

Similar constraints must be defined for the oxygen/current ratio. Combining all of the constraints gives

$$\begin{bmatrix} \underline{\underline{A_1}} \\ -\underline{\underline{A_1}} \\ \underline{\underline{A_3}} \end{bmatrix} \underline{\mathrm{d}u} \leq \begin{bmatrix} \underline{b_1} \\ \underline{b_2} \\ \underline{b_3} \end{bmatrix} \tag{35}$$

This defines all of the constraints. For the sake of sensitivity, the variables are all scaled by their respective values entering the optimization:

$$\mathrm{d}u_i = \mathrm{d}u_i^* u_{i,\mathrm{nom}} \tag{36}$$

or

$$\underline{\mathrm{d}u} = \mathrm{diag} \begin{pmatrix} u_{1,\mathrm{nom}} \\ \vdots \\ u_{1,\mathrm{nom}} \\ \vdots \\ u_{n,\mathrm{nom}} \\ \vdots \\ u_{n,\mathrm{nom}} \end{pmatrix} \underline{\mathrm{d}u^*} \tag{37}$$

Substituting Eq. (37) into Eq. (35) gives

$$\begin{bmatrix} \underline{\underline{A_1}} \\ -\underline{\underline{A_1}} \\ \underline{\underline{A_3}} \end{bmatrix} \cdot \mathrm{diag} \left(\begin{bmatrix} u_{1,\mathrm{nom}} \\ \vdots \\ u_{1,\mathrm{nom}} \\ \vdots \\ u_{n,\mathrm{nom}} \\ \vdots \\ u_{n,\mathrm{nom}} \end{bmatrix} \right) \overline{\mathrm{d}u^*} \leq \begin{bmatrix} \underline{b_1} \\ \underline{b_2} \\ \underline{b_3} \end{bmatrix} \tag{38}$$

These are the linear constraints on the optimization variables. Furthermore, in cases where the hydrogen or oxygen flow rates are not optimized, the current is limited to the permitted ratio between the current and the constant value of the reactant flow rate. In this case, the upper limit on the current is either the set maximum current density (imposed by the user) or the value determined by the permitted current/reactant flow ratio, the lower of the two.

Note that since the optimization variables include the state and output variables as well, the constraint matrix needs to be padded with zeros to match the dimensions of the optimization variable.

References

[1] P. Costamagna, Chem. Eng. Sci. 56 (2) (2001) 323–332.
[2] J.S. Yi, T.V. Nguyen, J. Electrochem. Soc. 145 (4) (1998) 1149–1159.
[3] J.H. Lee, T.R. Lalk, J. Power Sources 73 (2) (1998) 229–241.
[4] J.C. Amphlett, R.F. Mann, et al., J. Power Sources 61 (1–2) (1996) 183–188.
[5] B.S. Kang, J.H. Koh, et al., J. Power Sources 94 (1) (2001) 51–62.
[6] C.C. Lin, M.J. Kim, et al., J. Dyn. Syst. Meas. Control-Trans. ASME 128 (4) (2006) 878–890.
[7] A. Vahidi, A. Stefanopoulou, et al., IEEE Trans. Control Syst. Technol. 14 (6) (2006) 1047–1057.
[8] C. Wang, M.H. Nehrir, et al., IEEE Trans. Energy Convers. 21 (2) (2006) 586–595.
[9] J.T. Pukrushpan, A.G. Stefanopoulou, H. Peng, Proceedings of the American Control Conference, 2002, Vol. 4, pp. 3117–3122 (2002).
[10] K.C. Lauzze, D.J. Chmielewski, Ind. Eng. Chem. Res. 45 (13) (2006) 4661–4670.
[11] J. Golbert, D.R. Lewin, J. Power Sources 135 (1–2) (2004) 135–151.
[12] M.A. Henson, Comput. Chem. Eng. 23 (2) (1998) 187–202.
[13] L.T. Biegler, J. Process Control 8 (5–6) (1998) 301–311.
[14] T.E. Springer, T.A. Zawodzinski, et al., J. Electrochem. Soc. 138 (8) (1991) 2334–2342.