Proceedings Paper:

Drummond, R. orcid.org/0000-0002-2586-1718, Taghavian, H. and Baldivieso-Monasterios, P.R. (2024) Learning-to-control relaxation systems with the step response. In: 2024 American Control Conference (ACC) Proceedings. 2024 American Control Conference (ACC), 10-12 Jul 2024, Toronto, ON, Canada. Institute of Electrical and Electronics Engineers (IEEE), pp. 3584-3589. ISSN: 0743-1619. EISSN: 2378-5861.

https://doi.org/10.23919/acc60939.2024.10644346

Learning-to-control relaxation systems with the step response

Ross Drummond, Pablo Rodolfo Baldivieso Monasterios and Hamed Taghavian

***Abstract*—The problem of learning-to-control relaxation systems from data is considered. The main results of the paper show that the equilibrium of a relaxation system's step response defines the solution of a class of robust control problems and provides a good suboptimal solution to a class of linear quadratic regulator problems involving relaxation systems. These results demonstrate the potential to efficiently learn policies for these control problems from a single, easy-to-implement trajectory data point, being the step response. More broadly, these results highlight how the system structure and problem definition of the control problem can be exploited to generate data efficient learning to control methods.**

***Index Terms*—Optimal control, linear quadratic regulator, relaxation systems, step response.**

I. Introduction

Consider the problem of designing a feedback control policy $u(x(t)) : \mathbb{R}^n \to \mathbb{R}^{n_u}$ for a linear time invariant systems of the form

$$\frac{d}{dt}x(t) = Ax(t) + Bu(x(t)) + w(t), \tag{1}$$

$$y(t) = Cx(t) + Du(t), \tag{2}$$

with state $x(t) \in \mathbb{R}^n$, disturbance $w(t) \in \mathbb{R}^n$, $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times n_u}$ and $C \in \mathbb{R}^{n_y\times n}$. Throughout this paper, it is assumed that this system is controllable and detectable. Particular attention is paid to the case where $A \prec 0$, as in the state transition matrix A is a negative definite matrix. Whilst being somewhat restrictive, systems satisfying $A \prec 0$ are quite common in practice, for example with the mechanical and electrical systems discussed in [17]. In fact, the *relaxation systems* studied in [17] are classical examples of systems with symmetric realisations.

Definition 1: The system (1) is said to be a relaxation system if there exists a realisation such that $A \prec 0$, $D \succeq 0$ and $B = C^\top$.

Remark 1: It is remarked that some systems (such as over-damped systems which are diagonalisable with real modes) may first have have a state transformation applied to them in order to satisfy $A \prec 0$.

In [17], it was shown that the structure imposed upon relaxation systems can be used to greatly simplify their passivity analysis. The results of this paper are of a similar spirit except the focus is on showing how some imposed system structure can simplify optimal control problems. Besides these theoretical motivations, the authors have also encountered relaxation systems in several other applications, including in equivalent circuit models of batteries and super-capacitors [7], externally positive systems results of [6] and the search for Zames-Falb multipliers in [16]. The breadth of these topics highlights the range of applications where relaxation systems are encountered.

Contributions: The results of this paper concern the problem of learning feedback policies for optimal control problems involving relaxation systems. Three main contributions are developed:

- A class of linear quadratic regulator problems for relaxation systems is introduced whose solution can be approximated by their step response (Theorem 2).
- A class of robust control problems for relaxation are studied whose solution is defined by their step response (Section III-A).
- Numerical simulations validate the results.

The overall theme of these results is to highlight how solutions to certain control problems for relaxation systems can be learned from a single easy-to-implement trajectory– the system's step response. More broadly, these results highlight the potential of exploiting the structure of solutions to control problems to improve the sample complexity of reinforcement learning (RL) methods. For example, Theorem 2 may be used to generate good initiliasations of RL algorithms for solving certain LQR problem and Section III-A provides a complete solution to the robust control problem of Problem 2 using the step response.

The results of this paper can be considered to lie at the intersection of model-based and data-driven control (with these approaches compared in [8]), as they exploit structural properties in the model dynamics to improve the learning of optimal feedback policies from data. A key reference for data-driven control is [4] which demonstrated that Willems' Fundamental lemma [18] could be used to learn the input-output map of discrete-time linear systems from a persistently exciting signal, and then numerical optimisation problems could be solved on this learned model, including for LQR problems. By contrast, the presented results focus on continuous-time relaxation systems and the step response; we show that for this class of problems, an optimal gain can be learned i) without a persistently exciting signal (the step response is not necessarily persistently exciting), ii) and without needing

R. Drummond and P. R. B Monasterios are with the Department of Automatic Control and Systems Engineering, University of Sheffield, Mappin St, Sheffield, S1 3JD, United Kingdom. Email: {ross.drummond,p.baldivieso}@sheffield.ac.uk.

Hamed Taghavian is with the Division of Decision and Control Systems, School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, Stockholm 100 44, Sweden. Email: hamedta@kth.se.

to numerically solve an optimisation problem (Theorem 2 shows that the gain can be obtained directly from the state-steady response). Although the proposed results are more restrictive than [4] since they focus on relaxation systems, they may contain some advantages. For example, they do not require persistently exciting learning signals (such signals can be challenging to implement on experimental rigs, especially for systems with fast and/or slow dynamics), and the optimal control can be solved without computational hardware, as the feedback gain is simply extracted from the step-response. Restricting the problem class to relaxation systems may then bring advantages for learning-to-control in the field.

Notation: A matrix $A \in \mathbb{R}^{n\times n}$ is negative-definite if $A \prec 0$. Similarly, positive definite matrices are defined as $A \succ 0$. The identity matrix of dimension n is I_n. The set of unitary matrices of dimension n is defined as $\mathbb{U}^n$. The eigenvalues of a matrix A are denoted $\lambda(A)$ and its singular values are denoted $\sigma(A)$. The k^{th} eigenvalue and singular value are denoted $\lambda_k(A)$ and $\sigma_k(A)$, respectively.

A. Step response definition

Relaxation systems in the form of Definition 1 are stable and passive, so properties such as their step response can be defined. The following gives the definition of the step response used in this work. The main point to note is that the multi-input case with $n_u \geq 2$ involves a step in each input component separately, rather than simultaneously.

Definition 2: The step response of the system (1) is defined as the solution of (1) from $x(0) = 0$ with $w(t) = 0 \forall t$ and where $u_j(t) = 1$, $\forall t \geq 0$ and 0 otherwise, with $u_j(t)$ corresponding to the j^{th} component of the input signal. The step response equilibrium matrix is defined as

$$X^* = -A^{-1}B. \tag{3}$$

With this definition, the columns of $X^* \in \mathbb{R}^{n\times n_u}$ correspond to the step responses from each individual component of $u(t) \in \mathbb{R}^{n_u}$. The equilibrium of the step response's output is $Y^* = D - CA^{-1}B$.

II. Linear quadratic regulators for relaxation systems

In this section, the following LQR problem is considered.

Problem 1 (LQR problem): Consider the linear time invariant system dynamics of (1) with initial condition $x(0) \in \mathbb{R}^n$. Assume $R \succ 0$ and $Q \succ 0$ are given. Find the optimal control policy $u(t) = Kx(t)$ that minimises the quadratic cost function

$$\min_u \int_0^\infty x(t)^\top Q x(t) + u(t)^\top R u(t) \; dt. \tag{4}$$

The solution to the LQR optimal control problem is well-known and defined by the positive-definite solution of a matrix Riccati equation [1]. Whilst stating the solution to the LQR solution in terms of the Riccati equation brings many advantages (notably, it enables fast computation using convex optimisation algorithms), its implicit natures makes it difficult to interpret how the system dynamics and problem statement characterises the optimal solution. Since the aims of this paper are to exploit the system structure to create more efficient ways to learn some optimal control policies, this limitation motivates the following formulation.

Theorem 1: The LQR problem of Problem 1 is solved by the feedback policy

$$u(t) = Kx(t) = -R^{-1}B^\top P x(t) \tag{5}$$

where $P \succ$ is defined by

$$P = (-AQ^{-1} + (AQ^{-1}A^\top + BR^{-1}B^\top)^{1/2} U Q^{-1/2})^{-1} \tag{6}$$

and $U \in \mathbb{U}^n$ is the unique unitary matrix for which $P \succ 0$.

Proof. From LQR theory, the optimal solution to Problem 1 subject to the controllable dynamics of (1) is obtained from the unique positive-definite solution $P \succ 0$ of the following Riccati equation

$$A^\top P + PA - PBR^{-1}B^\top P + Q = 0. \tag{7}$$

As $P \succ 0$, it is invertible and so it is possible to define $L = P^{-1} \succ 0$. Divide the Riccati equation of (7) on the left and right by P to give

$$LA^\top + AL - BR^{-1}B^\top + LQL = 0. \tag{8}$$

Rewrite the above as

$$\begin{aligned}(LQ^{1/2} + AQ^{-1/2})(LQ^{1/2} + AQ^{-1/2})^\top \\ = AQ^{-1}A^\top + BR^{-1}B^\top,\end{aligned} \tag{9}$$

which implies that (8) is solved by

$$L = -AQ^{-1} + (AQ^{-1}A^\top + BR^{-1}B^\top)^{1/2} U Q^{-1/2} \tag{10}$$

for any unitary $U \in \mathbb{U}^n$. The unique solution $P \succ 0$ of the Riccati equation must then be obtained by the U that makes $L \succ 0$. With this, the LQR gain of (5) is obtained. ■

Remark 2: Some special cases exist where (6) can be used to characterise the unitary matrix $U \in \mathbb{U}^n$ and hence also the matrix $P \succ 0$. Notably, when $Q = \rho I$ and $A \prec 0$ for some $\rho > 0$ then (6) implies that $U = I_n$.

Remark 3: As an aside, it it remarked that since U is a unitary matrix, Theorem 1 gives an immediate bound for the optimal solution of the LQR problem. To see this, use $L = P^{-1}$ and write

$$\begin{aligned}\sigma_k(LQ^{1/2}) &\in \sigma_k(L)[\min \lambda(Q)^{1/2}, \max \lambda(Q)^{1/2}], \\ &= \sigma_{n-k+1}(P)^{-1}[\min \lambda(Q)^{1/2}, \max \lambda(Q)^{1/2}].\end{aligned}$$

Let $q = n-k+1$ and note that $LQ^{1/2} = -AQ^{-1/2} + M^{1/2}U$ with $M = AQ^{-1}A^T + BR^{-1}B^T$. As P is positive definite, the above implies

$$\lambda_q(P) \in \frac{1}{\gamma_q}[\min \lambda(Q)^{1/2}, \max \lambda(Q)^{1/2}] \tag{11}$$

where

$$\begin{aligned}\gamma_q &= \sigma_{n-q+1}(-AQ^{-1/2} + M^{1/2}U),\\ &\leq \sigma_i(-AQ^{-1/2}) + \sigma_j(M^{1/2}U),\\ &= \sigma_i(AQ^{-1/2}) + \lambda_j(M)^{1/2},\end{aligned}$$

and $i + j = n - q + 2$. Choosing $i = j = 1$ gives the lower bound $\min \lambda(P) \geq \min \lambda(Q)^{1/2}/\gamma_n$ where

$$\gamma_n = \max \lambda(AQ^{-1}A^T)^{1/2} + \max \lambda(M)^{1/2}.$$

Bounding P may be used to restrict the search-space of RL methods involving reward functions– such as Q-learning– and so, potentially, make those methods more sample efficient for some problems.

A. Suboptimal LQR for relaxation systems from the step response

If the dynamics of (1) are relaxation systems with $A \prec 0$ (and these dynamics may be unknown), then Theorem 1 can be used to obtain suboptimal solutions to Problem (1) using the step response equilibrium matrix X^*.

Theorem 2: Assume $A \prec 0$, $Q = I_n$ and $\max \lambda(X^* R^{-1} X^{*\top}) < 1$. Consider the control policy

$$u(t) = -\frac{R^{-1}}{2} X^{*\top} x(t), \tag{12}$$

where $X^* = -A^{-1}B$ is the system's step response equilibrium. The optimality gap associated with this policy is bounded by

$$\begin{aligned}J_{step} - J_{LQR} &= x_0^\top \mathcal{O}(X^* R^{-1} X^{*\top}) x_0, &\text{(13a)}\\ &\leq \max \lambda(X^* R^{-1} X^{*\top}) ||x_0||^2, &\text{(13b)}\end{aligned}$$

where J_{LQR} is the cost of the optimal LQR policy and J_{step} is the cost obtained with (12). This bound implies that the policy (12) approximates the optimal one when $X^* R^{-1} X^{*\top}$ is sufficiently small.

Proof. From Theorem 1 and following Remark 2, the solution of the Riccati equation $P = {L_{LQR}}^{-1}$ for this LQR problem follows

$$\begin{aligned}L_{LQR} &= -A + (AA^\top + BR^{-1}B^\top)^{1/2},\\ &= -A - A(I_n + X^* R^{-1} X^{*\top})^{1/2},\end{aligned} \tag{14}$$

which, by the matrix inversion lemma, gives

$$P_{LQR} = -A^{-1} + ((I + X^* R^{-1} X^{*\top})^{-1/2} + I)^{-1} A^{-1}.$$

Assuming $\max \lambda(X^* R^{-1} X^{*\top}) < 1$, one may use the Laurent series coefficients of the function $f(z) = ((1 + z)^{-1/2} + 1)^{-1}$ about the origin to write the above as

$$\begin{aligned}P_{LQR} &= -A^{-1} + (I/2 + (X^* R^{-1} X^{*\top})/8 + \dots) A^{-1}\\ &= -\frac{1}{2} A^{-1} + ((X^* R^{-1} X^{*\top})/8 + \dots) A^{-1}\\ &= P_{step} - \mathcal{O}(X^* R^{-1} X^{*\top})\end{aligned} \tag{15}$$

where $P_{step} = -\frac{1}{2}A^{-1}$. Therefore when $X^* R^{-1} X^{*\top} \to 0$, the sub-optimal feedback policy choice of (12) approximates the optimal LQR control policy $u_{LQR}(t)$ as follows

$$u_{LQR}(t) = -R^{-1}B^\top P x(t) \simeq -\frac{R^{-1}}{2} X^{*\top} x(t).$$

The bound of (13) then follows from (15). ∎

Remark 4: Since the control policy of (12) is defined by the step response matrix X^*, it can be learned from a single trajectory data point. As such, approximate solutions in the sense of (13) for these control problems can be efficiently learned. Note that to compute X^*, and hence the feedback policy, it is not required to know the model dynamics, just that A is negative-definite and that the system is controllable. Whilst $A \prec 0$ is restrictive, [17] showed that it is satisfied by a broad class of physical systems, including many mechanical and electrical systems.

Remark 5: To obtain the step response matrix X^*, only n_u experiments are needed. During the learning phase, noise may corrupt the experiments, but the only relevant part of these experiments is their steady state value which is a single data point (information about the transients is not required). Depending upon the type of noise, different filters may have to be applied to the components of X^* and the feedback policy $u(t)$ if the measurements of $x(t)$ are noisy. Moreover, bias in the noise could also play an important role in corrupting the measurement of X^*. Again, data processing may have to be applied in this case.

Remark 6: Implementing data-driven control on continuous time problems, such as Problem 1, can introduce several issues. For example, there are issues following from the data being collected in discrete samples and derivatives having to be approximated. By contrast, learning the policy of (12) does not necessarily require discretising the state signal $x(t)$ in time– all that is required is the steady-state signal X^*. This may bring benefits for scalability and ease of implementation. Unlike existing data-driven control schemes such as those from [4] which optimise using a model built from the data itself, the policy (12) does not require numerically solving an optimisation problem. Instead, all that is required for (12) is X^*. Decoupling the learning from the state-dimension in this way may allow the method to scale to large systems, as long as the problem's assumptions hold.

III. Robust control of relaxation systems

The previous section contained the main result of this paper–an approximation of LQR control policies for relaxation systems in terms of their step responses. This result exploited the explicit characterisation of the optimal control problem's solution from Theorem 1 to characterise the solution of the LQR problem in terms of the system matrices, a feature which was then used to get an approximation in terms of the step response. In the control theory literature, similar results on explicit solutions for optimal control problems have already been derived and which elicit similar data-driven characterisations (as will be seen in Section III-A). In particular, [13] presented an explicit solution for the following robust control problem which was also formulated for relaxation systems.

Problem 2: Let $\alpha > 0$ and consider the LTI dynamics of (1) with $x(0) = 0$. Minimise

$$\sup_{w \in \mathcal{W}} \int_0^\infty y(t)^\top y(t) + \alpha^2 u(t)^\top u(t)\ dt, \tag{16}$$

where the disturbance set $\mathcal{W}$ is defined as the space of disturbance signals w with bounded $\mathcal{L}_2$ norm

$$\mathcal{W} := \left\{ w : \int_0^\infty w(t)^\top W w(t)\ dt \leq 1 \right\}, \tag{17}$$

for some positive-definite matrix $W \succ 0$.

The optimal solution to this robust control problem was obtained in [13].

Theorem 3 ([13]): If the LTI system (1) is a relaxation system (Definition 1) then $K = \alpha^{-1}(D - CA^{-1}B)$ solves Problem 2.

Remark 7: Similar results to [13] on the robust control of symmetric and relaxation systems can be found in [15], [11],[9],[2], [12] and, more recently, in [19].

A. Solution in terms of the step response

Even though Theorem 3 gives an explicit solution to the robust control problem of Problem 2, it is stated in terms of the system's A, B, C and D matrices. Implementing this optimal control policy using the theorem therefore requires knowledge about the matrices of the system model, which is not possible using the data-driven approach.

However, it is observed that the optimal solution of Theorem 3 is simply the equilibrium of the system output's step response, and so can be learned from a single trajectory data point. Specifically, the optimal policy of Theorem 3 is simply $K = \alpha^{-1} Y^*$ following Definition 2. The optimal feedback gain can then be obtained by simply applying the step input, waiting for the system's transients to decay and then using the settled output values within the feedback gain. In practice, the steady-signal signal Y^* may have to be filtered or de-biased when implementing this strategy, but this will depend upon the particular application being considered. The above analysis implies that when the system dynamics are relaxation systems, robust control problems in the form of Problem 2 can also be learned from the steady-state of the step response.

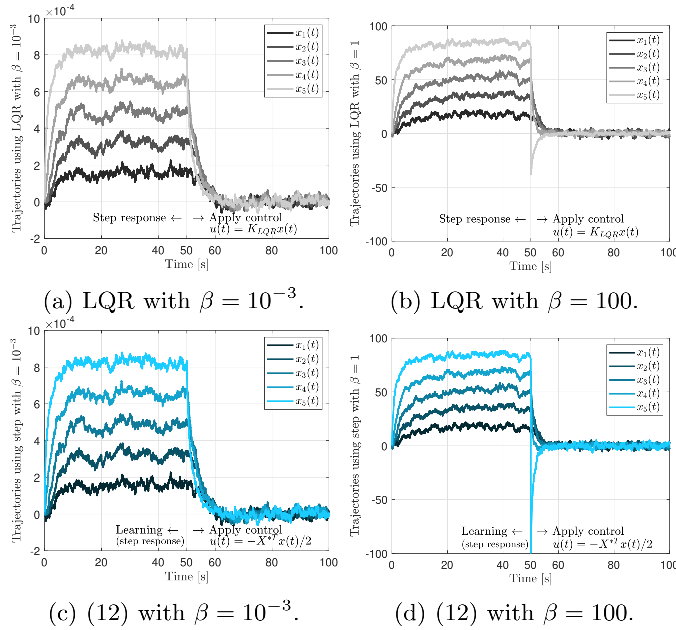
IV. Numerical example

Example 1

To demonstrate the validity of the approximation of Theorem 2, consider the problem of controlling the diffusion equation

$$\frac{\partial v(r,t)}{\partial t} = \frac{\partial^2 v(r,t)}{\partial r^2}, \tag{18}$$
$$\text{subject to } v(0,t) = 0, v(1,t) = \beta u(t),$$

defined on the spatial domain $r \in [0,1]$ and with $\beta > 0$ being a scaling term which scales the magnitude of the step response. The spatio-temporal variable that is diffusing is denoted as $v(r,t) : [0,1] \times [0,\infty) \to \mathcal{L}_{[0,1]\times[0,\infty)}$. The control actuation $u(t)$ is applied at the boundary at $r = 1$.



(a) LQR with $\beta = 10^{-3}$. (b) LQR with $\beta = 100$.

(c) (12) with $\beta = 10^{-3}$. (d) (12) with $\beta = 100$.

Figure 1: Comparison between the optimal LQR controller and that obtained from the step-response in Theorem 2.

To control this system, the spatial domain r is discretised into $n+2$ equally spaced grid points (with spacing $\Delta = 1/(n+1)$). These grid points are located at $\mathbf{r} \in \mathbb{R}^{n+2}$ with $\mathbf{r}_j = (j-1)\Delta$ for $j = 1, \ldots, n+2$. The value of the variable $v(r,t)$ on these grid points is defined as $\mathbf{v}_j(t) \in \mathbb{R}^{n+2} = v(\mathbf{r}_j, t)$ for $j = 1, \ldots, n+2$. With this discretisation, the diffusion operator is approximated by the central difference

$$\frac{\partial^2 v(r,t)}{\partial r^2} \approx \frac{\mathbf{v}_{k+1}(t) - 2\mathbf{v}_k(t) + \mathbf{v}_{k-1}(t)}{\Delta^2}, \tag{19}$$

for $k = 1, \ldots, n$, $\mathbf{v}_0 = 0$, $\mathbf{v}_{n+1} = u(t)$. The discretised version of the diffusion equation (18) then has the form of (1) with state $x(t) \in \mathbb{R}^n = \mathbf{v}_{2:n+1}(t)$, state transition matrix $A \prec 0$ defined by

$$A = \frac{1}{\Delta^2} \begin{bmatrix} -2 & 1 & & \\ 1 & -2 & \ddots & \\ & \ddots & & 1 \\ & & 1 & -2 \end{bmatrix} \tag{20}$$

and with $B = [0, \ldots, 0, \beta/\Delta^2]^\top$.

Figure 1 evaluates the performance of feedback policies to control this system, with a comparison between the optimal LQR gains (denoted K_{LQR}) obtained by solving the Riccati equation and that obtained by the approximation of Theorem 2. For the simulations of this figure, the parameters $n = 5$, $\Delta = 1$, $w(t) \sim \mathcal{N}(0, 10^{-5}\beta^2)$ and $Q = I_n$, $R = 1$ were used. Figures 1a & 1c compare the two controllers with $\beta = 10^{-3}$ where the step response is small in magnitude whereas 1b & 1d compare the controllers for the larger step response with $\beta = 100$. The simulations are separated into two distinct stages–each 50 seconds long. In the first stage, the step input of Definition 2 is applied and

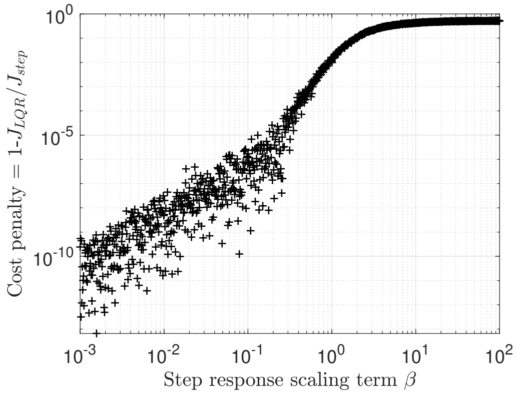
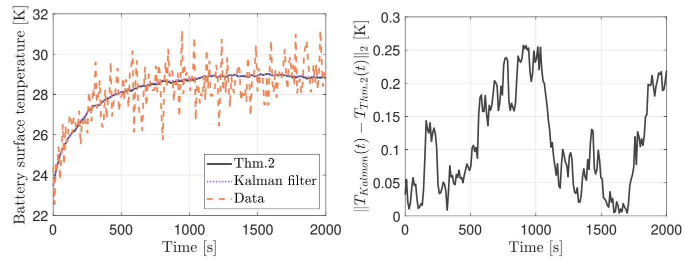


Figure 2: Cost comparison between the optimal LQR policy and Theorem 2 as a function of β for Example 1.



(a) Battery surface temperature during fast charging.

(b) Error between Kalman filter and Theorem 2.

Figure 3: Comparison between Kalman filter and Theorem 2 for Example 2. In (b), T_{Kalman} is the temperature estimate from the Kalman filter and $T_{Thm.2}$ is that based upon the gain (12) of Theorem 2.

the gains of (12) are then learned from the data. In the second stage, the feedback control action is applied.

The results of Figure 1 agree with the conclusions of Theorem 2; for the considered LQR problems involving relaxation systems, the optimal LQR gain can be approximated well by the learned policy (5) when the step response is small. This can be seen by comparing the responses of Figures 1a & 1c and Figures 1b & 1d. When the step response is small in magnitude, as in when $\beta = 10^{-3}$, then the two responses appear identical, as seen in Figures 1a & 1c. When the step response is bigger, as in when $\beta = 100$, then the approximation of (5) deteriorates, as seen in the difference between the responses of Figures 1b & 1d. This difference is especially noticeable in the response of the state x_5 at 50s.

Example 2

To demonstrate the results can be applied in practice, consider the problem of designing a Kalman filter to estimate the temperature distribution within large cylindrical lithium-ion batteries using only thermocouple measurements on the cell's surface. As batteries get increasingly employed in technologies where safety is critical, notably electric aircraft, there is a growing need to develop temperature estimators to detect short circuit faults and avoid thermal runaway events. To address this problem, it is noted that even though Theorem 2 solves a LQR problem, under the assumption that the battery thermal dynamics are linear, the separation principle implies that they can also be used to approximate Kalman filter gains.

For the battery model, we follow the approach of [3] and discretise the cylindrical cell into surface, middle and core components. The cell's A and B matrix for the state-estimation problem with surface temperature measurements are then

$$A = M^{-1}\begin{bmatrix} -\frac{1}{r_{c2m}} & \frac{1}{r_{c2m}} & 0 \\ \frac{1}{r_{c2m}} & -\frac{1}{r_{c2m}} - \frac{1}{r_{m2s}} & \frac{1}{r_{m2s}} \\ 0 & \frac{1}{r_{m2s}} & -\frac{1}{r_{s2a}} - \frac{1}{r_{m2s}} \end{bmatrix} \quad (21)$$

where $M = C_p \ \mathrm{diag}(m_{core}, m_{mid}, m_{surf})$ and $B = [0, \ldots, 0, 1]^\top$. Here, $C_p = 1100$ J kg^{-1}K^{-1} is the specific heat capacity of battery core, m_{surf} is the mass of the cell surface, m_{surf} is that of the middle, and m_{surf} is that of the core. It will be assumed that that cell is discretised evenly, with $m_{core} = m_{mid} = m_{surf} = 0.0346$ kg. $r_{c2m} = 3.18$ KW^{-1} is the thermal resistance between the cell core and mid, $r_{m2s} = 1.61$ KW^{-1} is that between the cell mid and surface and $r_{s2a} = 1$ K W^{-1} is that between the cell surface and the surrounding air. These parameters were obtained from [3]. The cost function of the Kalman filtering problem is given by (4) but with $Q = 10^{-3}I_3$ (a state-transformation translates this cost into that of Theorem 2) being the covariance matrix for the state noise and $R = 1$ being that for the sensor noise.

Figure 3 compares the results of Theorem 2 with the optimal Kalman filter gain under the condition of the cell being fast-charged with an Ohmic heating rate of I^2R_{ser} where $I = 4.5$ A is an applied 1C charging current and $R_{ser} = 0.0246\,\Omega$ is the cell's series resistance. To learn the gain of (12), the battery could be heated using external heaters, as in [5], and a thermal imaging camera (such as that developed in [10]) could then be applied to measure the equilibrium thermal state of the cell defining the gain of (12). The figure shows that during the fast charge, the thermocouple data is corrupted by noise but this noise is effectively filtered out by both the Kalman filter and the gain of Theorem 2. Figure 3b shows that the error between the two state estimators is small, providing support for the practical application of Theorem 2.

For this battery temperature estimation problem, the value of the developed approach is that state-estimator gains can be obtained quickly from a single experiment, e.g. by heating the battery using an external heat source and then measuring the equilibrium temperature distribution using a thermal camera to obtain the gain. This single-experiment gain-synthesis approach could accelerate the time needed to design battery management systems, as the estimator gains could be obtained without parameterising a model. Instead, with the proposed approach, the gains may be computed from the formation cycling experimental data. By posing the gain synthesis problem in the language of battery experimentalists, new avenues to implement and tune BMS algorithms may emerge.

Conclusions

The problem of learning feedback policies for a class of optimal control problems based upon relaxation systems was considered. A class of output-feedback robust control problems was shown to be solved by the policy whose feedback gain is simply the equilibrium of the relaxation system's step response. A class of LQR problem was also shown to be approximated by a policy defined by the step response's equilibrium. These two results highlight how feedback policies for these optimal control problems can be learned directly from a single easy-to-implement data point, being the step response. The potential of exploiting properties of the system dynamics and known solutions to optimal control problems to create more data efficient learning-to-control methods tailored for specific systems is highlighted by these results.

In terms of future work, one problem would be to develop conditions to check from data if a system is a relaxation system or not. The results of this paper demonstrate that if the system can be identified as a relaxation system, then the problem of learning the solutions to optimal control problems can be efficiently addressed. This line of work may follow the results on learning passivity properties of linear systems, as considered in [14], as relaxation systems and passive systems have close links [17]. Moreover, we see benefits in exploiting the structure observed in Theorem 1 for the general solution of the LQR problem based upon unitary matrices, especially in the context of vision-based control for robotics.

Acknowledgements

The authors would like to thank Hadriano Morales Escamilla for the insightful comments.

References

[1] K. J. Åström and R. M. Murray, *Feedback systems: An introduction for scientists and engineers.* Princeton University Press, 2021.

[2] V. Borukhov and D. Zelenyak, "Description of a set of solutions of a dissipation inequality for third-order relaxation systems," *J.of Engineering Physics and Thermophysics*, vol. 73, no. 5, pp. 1010–1018, 2000.

[3] T. Cai, A. G. Stefanopoulou, and J. B. Siegel, "Modeling Li-ion battery temperature and expansion force during the early stages of thermal runaway triggered by internal shorts," *J. of The Electrochemical Society*, vol. 166, no. 12, p. A2431, 2019.

[4] C. De Persis and P. Tesi, "Formulas for data-driven control: Stabilization, optimality, and robustness," *IEEE Trans on Aut. Cont.*, vol. 65, no. 3, pp. 909–924, 2019.

[5] S. Dey, H. E. Perez, and S. J. Moura, "Model-based battery thermal fault diagnostics: Algorithms, analysis, and experiments," *IEEE Trans. on Control Systems Technology*, vol. 27, no. 2, pp. 576–587, 2017.

[6] R. Drummond, M. C. Turner, and S. R. Duncan, "External positivity of linear systems by weak majorisation," in *Procs. of the American Control Conf.* IEEE, 2019, pp. 5191–5196.

[7] R. Drummond, S. Zhao, D. A. Howey, and S. R. Duncan, "Circuit synthesis of electrochemical supercapacitor models," *J. of Energy Storage*, vol. 10, pp. 48–55, 2017.

[8] S. Formentin, K. Van Heusden, and A. Karimi, "Model-based and data-driven model-reference control: a comparative analysis," in *Procs. of the European Control Conf.* IEEE, 2013, pp. 1410–1415.

[9] L. Fortuna, G. Muscato, and G. Nunnari, "On $\mathcal{H}_\infty$-control for symmetric systems," in *Procs. of the Conf. on Decision and Control.* IEEE, 1992, pp. 3723–3725.

[10] T. Heenan, I. Mombrini, A. Llewellyn, S. Checchia, C. Tan, M. Johnson, A. Jnawali, G. Garbarino, R. Jervis, D. Brett *et al.*, "Mapping internal temperatures during high-rate battery applications," *Nature*, vol. 617, no. 7961, pp. 507–512, 2023.

[11] M. Liu, J. Lam, B. Zhu, and K.-W. Kwok, "On positive realness, negative imaginariness, and $\mathcal{H}_\infty$ control of state-space symmetric systems," *Automatica*, vol. 101, pp. 190–196, 2019.

[12] R. Pates, "Passive and reciprocal networks: From simple models to simple optimal controllers," *IEEE Control Systems Magazine*, vol. 42, no. 3, pp. 73–92, 2022.

[13] R. Pates, C. Bergeling, and A. Rantzer, "On the optimal control of relaxation systems," in *Procs. of the Conf. on Decision and Control.* IEEE, 2019, pp. 6068–6073.

[14] A. Romer, J. Berberich, J. Köhler, and F. Allgöwer, "One-shot verification of dissipativity properties from input–output data," *IEEE Control Systems Letters*, vol. 3, no. 3, pp. 709–714, 2019.

[15] K. Tan and K. M. Grigoriadis, "Stabilization and $\mathcal{H}_\infty$ control of symmetric systems: an explicit solution," *Systems & Control Letters*, vol. 44, no. 1, pp. 57–72, 2001.

[16] M. C. Turner and R. Drummond, "Analysis of systems with slope restricted nonlinearities using externally positive Zames–Falb multipliers," *IEEE Trans. on Aut. Con.*, vol. 65, no. 4, pp. 1660–1667, 2019.

[17] J. C. Willems, "Realization of systems with internal passivity and symmetry constraints," *J. of the Franklin Institute*, vol. 301, no. 6, pp. 605–621, 1976.

[18] J. C. Willems, P. Rapisarda, I. Markovsky, and B. L. De Moor, "A note on persistency of excitation," *Systems & Control Letters*, vol. 54, no. 4, pp. 325–329, 2005.

[19] F. Yang, F. Gama, S. Sojoudi, and N. Matni, "Distributed optimal control of graph symmetric systems via graph filters," in *Conference on Decision and Control (CDC).* IEEE, 2022, pp. 5245–5252.