Adaptive Biomedical Treatment and Robust Control ⋆

Q. Clairon * **E. D. Wilson** ** **R. Henderson** *** **C. J. Taylor** ****

* *School of Mathematics and Statistics, University of Newcastle, UK (e-mail: Quentin.Clairon@ncl.ac.uk)*
** *Engineering Department, Lancaster University, UK (e-mail: e.d.wilson1@lancaster.ac.uk)*
*** *School of Mathematics and Statistics, University of Newcastle, UK (e-mail: Robin.Henderson@ncl.ac.uk)*
**** *Engineering Department, Lancaster University, UK (e-mail: c.taylor@lancaster.ac.uk)*

Abstract: An adaptive treatment strategy is a set of rules for choosing effective medical treatments for individual patients. In the statistical literature, methods for optimal dynamic treatment (ODT) include *Q-learning* and *A-learning* methods, which are linked to machine learning in engineering and computer science. The research project behind this article aims to develop new methodology for both ODT and engineering control, through the integration of techniques and approaches that have been developed in both fields, with a particular focus on the problem of robustness. The methodological framework is based on a regret-regression approach from the statistical literature and non-minimal state-space methods from control. This article provides an introduction to some of these concepts and presents preliminary novel contributions based on the application of robust H^∞ methods to ODT problems.

Keywords: Linear control systems (TC2.2); optimal control (TC2.4); control of physiological and clinical variables (TC8.2)

1. INTRODUCTION

There is growing interest in the use of control for biomedical applications. These typically have greater stochastic uncertainty and weaker repeatability than found in classical engineering application areas. Although control theory has been connected to biological systems for decades (see e.g. Doyle III et al., 2011, for a review), developments in sensor technology mean that it is now possible to measure physiological variables such as the heart rate of animals on–line. Selected examples include the real–time control of heart rate (Taylor and Aerts, 2014) and targeted growth curves in poultry farming (Cangar et al., 2007). Such control problems are stimulating links to related challenges in biostatistics, particularly in personalised medicine. One example is warfarin dosing strategies for long-term anticoagulation, in which the output is blood clotting speed and the control input is the dose (Henderson et al., 2011; Rich et al., 2016). Another is maintenance chemotherapy for childhood leukaemia, in which the input is doseage of cytostatics and the output is white blood cell count, measured weekly (Rosthoj et al., 2012).

In the statistical literature, research in optimal dynamic treatment, ODT, has developed rapidly since the seminal papers of Murphy (2003) and Robins (2004). Approaches include *Q*- and *A-learning* (Chakraborty and Moodie, 2013; Schulte et al., 2014), which are linked to machine learning in engineering and computer science. The aim is to derive adaptive decision rules for medical treatments or other interventions. In many applications there are few decision times, a low number of possible treatments and a finite follow-up period. However, the methods have begun to be used for adaptive treatment of chronic conditions, for dose selection and under infinite horizons (Henderson et al., 2010; Rosthoj et al., 2012). Consequently there is clear overlap with the scenarios typically considered in modern control theory.

In control, a major concern is to ensure satisfactory behaviour in the presence of modelling uncertainty, sampling problems, external disturbances and sensor noise, considerations that have close parallels in ODT regimes. In this context, the H^∞ philosophy is particularly pertinent, since it complements stochastic approaches, and seeks designs that minimise the effect of the disturbance that produces the largest effect on the system output (Mustafa and Glover, 1990), or that maximises the size of the uncertainty region for which a single controller can guarantee a satisfactory performance (Zhou et al., 1995).

The present authors are aiming to combine ideas from well–established approaches in control, including H^∞ methods, with the statistical theory of ODT regimes, with a particular focus on the problems of irregular sampling and robustness. This article provides a tutorial introduction to some of the relevant methods and describes preliminary novel contributions in respect to the application of robust H^∞ methods to both ODT and engineering control.

⋆ This work is supported by the UK Engineering & Physical Sciences Research Council (EPSRC) grant EP/M015637/1 (2015–2018).

The article is motivated by the *Control and Data Driven Modelling in Biomedicine* open track theme at the IFAC Congress, and aims to stimulate discussion and research links under this topic.

Section 2 introduces the problem from a statistical perspective, concentrating on the regret version of A-learning advocated by Murphy (2003). In section 3 we develop a novel state-space representation of the statistical problem and in section 4 we recommend a combined approach in which firmly based statistical estimation methods are used in modelling but H^∞ ideas are used in subsequent control decisions. The state space model used here is based on a non-minimal state space (NMSS) form that has previously been utilised in engineering applications for proportional-integral-plus (PIP) control (Taylor et al., 2000, 2013). In section 5, we present an expression for the H^∞ norm of the NMSS/PIP control system and subsequently use this to evaluate the closed-loop robustness when applied to the model defined by the earlier regret analysis.

2. ADAPTIVE TREATMENT AND A-LEARNING

In biostatistical applications, data will be available in the form of short and noisy sequences of observations on many subjects, and there will be no opportunity to collect repeat data. Most of the emphasis in the statistical literature has been on parameter estimation, the properties of estimators and the assumptions required for causal inference (Zhang et al., 2013). By contrast, in engineering, usually a single subject is under study but it is closely monitored, with frequent observations and the emphasis is on performance and robustness of the controller. Nonetheless the generic problem is the same and can be described as follows. At time k an output s_k is observed and an input u_k is determined. The input is directly controllable by an experimenter but the output is not. The output can be a vector or functional response. Given the history of all previous inputs, outputs and other information leading up to time k, the purpose is to choose u_k so as to achieve some objective measured in terms of future outputs. The same problem of course is of interest in a large number of other areas, including machine learning, scheduling and other sequential decision problems in operational research.

In the statistical literature there is usually a fixed number K of decision times and the objective is to maximise some quantity $Y(\mathbf{u})$ whose properties depend on the vector $\mathbf{u}$ of potential inputs. For chronic conditions the objective function might accrue over time, such as the proportion of time the biomarker is within a target range, whereas for acute conditions it may not be available until after the Kth input, for example the health of a patient at the end of a course of treatment. We will concentrate on the chronic, accruing information, situation in this note.

Two main strategies are available: Q-learning and A-learning (Chakraborty and Moodie, 2013; Schulte et al., 2014). The former, *quality learning*, attempts to relate directly the objective function to the inputs and outputs, either nonparametrically or through a statistical model. Recursive procedures akin to dynamic programming are then used to determine the optimal input at each decision time. Such an approach has the advantage that any modelling assumptions can in principle be checked against the data, but the disadvantage that the method is highly computationally expensive for large K. We concentrate therefore on *advantage learning* which is based on *contrasts* between outcomes under different decision rules, and specifically on the *regret* function approach (Moodie et al., 2007; Henderson et al., 2010; Barrett et al., 2014; Robins, 2004; Rosthoj et al., 2006).

We begin with notation. Data will be available from a cohort of patients but, unless necessary, we consider a single patient to illustrate the ideas. The patient acts as the plant in engineering terminology. We use s_k and u_k as general outputs and inputs at time k and S_k and U_k to denote the actual values observed on the patient. An overbar implies current and previous values: thus $\bar{S}_k = (S_1, S_2, \ldots, S_k)$ and $\bar{U}_k = (U_1, U_2, \ldots, U_k)$. The information available to the controller at time k is thus $(\bar{S}_k, \bar{U}_{k-1})$. We use $\underline{\mathbf{u}}_{\mathbf{k}}^{\mathbf{opt}}$ as the unknown sequence of optimum inputs from time k to K. The regret function is defined as:

$$\mu_k\left(u_k \mid \bar{S}_k, \bar{U}_{k-1}, \psi\right) =$$
$$\mathbb{E}\left[Y \mid \bar{S}_k, \bar{U}_{k-1}, \underline{\mathbf{u}}_{\mathbf{k}}^{\mathbf{opt}}\right] - \mathbb{E}\left[Y \mid \bar{S}_k, \bar{U}_{k-1}, u_k, \underline{\mathbf{u}}_{\mathbf{k+1}}^{\mathbf{opt}}\right]$$

We will take these terms in sequence.

(1) $\mu_k(.)$ is a function which can change over k. It has an argument u_k which is the potential input at time k.
(2) At that time $(\bar{S}_k, \bar{U}_{k-1})$ have been observed and will influence the choice of the next input.
(3) We will assume a parametric model for $\mu_k(.)$ which can depend on a vector parameter ψ.
(4) The first term on the right is the expected response, given the information to hand, and assuming we follow optimal rules in the future. Thus it represents the best we can expect to do.
(5) The next term on the right is similar, but assumes the input at time k is u_k, the function argument, and that subsequently the optimal rules are followed.

Since the aim is to maximise the expected Y, the right-hand side can never be negative. It measures the loss or regret caused by choosing input u_k rather than the optimal and will be zero if, indeed, optimal input is selected.

The statistical algorithm then has three steps.

(1) A parametric model is assumed for $\mu_k(.)$.
(2) A parameter estimator $\widehat{\psi}$ is constructed from the data: see e.g. Murphy (2003); Moodie et al. (2007); Henderson et al. (2010).
(3) An estimator of the optimal treatment strategy value $\widehat{\mathbf{u}_k^{opt}}\left(\bar{S}_k, \bar{U}_{k-1}\right)$ is derived by solving the equation:

$$\mu_k\left(u_k \mid \bar{S}_k, \bar{U}_{k-1}, \widehat{\psi}\right) = 0 \quad (1)$$

There is thus no attempt to describe the dynamics of the underlying biological processes: our assumptions are on the contrast given above. A disadvantage of this approach is that there is no way directly to check the modelling assumptions at the first step, though it is expected to be more robust to model misspecification than Q-learning (Schulte et al., 2014). An advantage is that the optimal input can usually be obtained very easily at the third step, without recursive procedures. Attempts have been made recently to develop robust procedures (Zhang et al., 2013; Barrett et al., 2014; Wallace and Moodie, 2015; Wallace et al., 2016) but to our knowledge there have been no attempts to date to adapt to this setting any of the robust methods developed for control engineering.

3. STATE-SPACE MODEL

Our aim is to design a treatment strategy which is robust to the presence of model misspecification and measurement noise. We restrict to the case of chronic conditions where the aim is to find the optimal treatment strategy $\mathbf{u^{opt}}$ which makes the patient state as close as possible to a sequence of targets $\overline{w}_n = (w_1^*, \dots w_n^*)$. The objective can then be formalised as a tracking problem:

$$Y(\mathbf{u}) = \sum_{k=1}^{K} -\left(S_k(\mathbf{u}) - w_k^*\right)^2 \tag{2}$$

In the statistical literature, the presence of uncertainty is usually only taken into account in parameter estimation. Yet in practice we face three sources of uncertainty which can affect the optimal strategy:

(1) Dynamic uncertainty. Stochastic or deterministic elements acting on the system dynamic may have been omitted during regret function parametrisation.
(2) Measurement uncertainty. We do not have access to S_k but only to noisy observations $\widetilde{S_k} = S_k + \varepsilon_k^m$ with ε_k^m representing measurement noise.
(3) Parametric uncertainty. Error due to parametric estimation in the assumed parameter value i.e we took $\widehat{\psi} = \psi + \triangle_\psi$, with $\triangle_\psi$ an unknown error term, instead of ψ.

In terms of control theory (Sontag, 1998; Clarke, 2013; Taylor et al., 2013; Ljung, 1999), this problem can be restated as:

Claim 1. ODT estimation methods use only the nominal plant without taking into account model uncertainty to determine the feedback control.

Our aim here is to take advantage of existing firmly-based statistical methods to estimate the parameter ψ, but to use results from control theory in order to design a treatment strategy which is robust to both model misspecification and noise measurement. In order to do so, we need to formulate our ODT problem in a framework compatible with control theory. The next proposition makes the connection between regret functions and a state-space representation required to apply control theory methods. Here we assume the system evolution between S_k and S_{k+1} can be described by the following state-space model:

$$S_{k+1} = f_k(S_k, U_k) + \varepsilon_k^d \tag{3}$$

with ε_k^d a random variable independent of (S_k, U_k) representing stochastic elements acting on the system dynamic. Then, we have the following result.

Proposition 2. Assume that at each $k \leq n-1$, there exists u_k such that: $-f_k(S_k, u_k) + w_{k+1}^* = \mathbb{E}\left[\varepsilon_k^d\right]$

Then the optimal strategy $\mathbf{u^{opt}}$ at each time step is:

$$\mathbf{u_k^{opt}}(s_k, u_{k-1}) = \arg\min_{u_k}$$
$$\mathbb{E}\left[\left(S_{k+1}\left(\mathbf{u}_k\left(\overline{S}_k, \overline{U}_{k-1}\right)\right) - w_{k+1}^*\right)^2 \mid S_k = s_k,\, U_{k-1} = u_{k-1}\right]$$

and

$$\mu_k\left(u_k \mid \overline{S}_k, \overline{U}_{k-1}\right) = \left(f_k(S_k, u_k) - w_{k+1}^* + \mathbb{E}\left[\varepsilon_k^d\right]\right)^2$$

Thus, when we assume a parametric formulation for the regret function $\mu_k\left(u_k \mid \overline{S}_k, \overline{U}_{k-1}\right) \simeq \mu_k\left(u_k \mid \overline{S}_k, \overline{U}_{k-1}, \psi\right)$, we automatically have a parametric form for $f_k(S_k, u_k) \simeq f_k\left(S_k, u_k, \psi\right)$. In particular we have access to the estimator $\widehat{f}_k(S_k, u_k) = f_k\left(S_k, u_k, \widehat{\psi}\right)$ with $\widehat{\psi}$ an estimator of ψ obtained by any of the available methods.

We now formulate a control problem to obtain the optimal treatment strategy sequence $\overline{u}_{k-1}$:

$$\begin{aligned} &\min_{\overline{u}_{k-1}} \lambda \left\|\overline{S}_n - \overline{w}_n\right\| + \left\|\overline{u}_{K-1}\right\| \\ &\text{under constraint } S_{k+1} = f_k(S_k,\, u_k,\, \widehat{\psi}) + \varepsilon_k^{glob} \end{aligned} \tag{4}$$

with access to the noisy observations: $\widetilde{S}_k = S_k + \varepsilon_k^m$.

Here ε_k^{glob} represents the whole committed misspecification error made by choosing $f_k(S_k,\, u_k,\, \widehat{\psi})$ as the model. This uncertainty term can be written $\varepsilon_k^{glob} = \varepsilon_k^d + \triangle_{det} + \triangle_\psi$, where ε_k^d is the stochastic disturbance, $\triangle_{det}$ the uncertainty term due to deterministic misspecification in $\mu_k\left(u_k \mid \overline{S}_k, \overline{U}_{k-1}, \psi\right)$ and $\triangle_\psi$ the term due to error made during ψ estimation. Solving (4) means driving S_k toward w_k^* while ensuring the dose sequence $\overline{u}_{K-1}$ takes reasonable values. The balance between these somewhat opposite objectives is made through the selection of the hyper-parameter λ. The norm $\|.\|$ in (4) is left unspecified because it depends on the method used to solve the problem. In the following, we use H^∞ (Doyle et al., 1989; Glover and Doyle, 1988; Iwasaki and Skelton, 1994; Francis, 1987), whilst we base our control framework on NMSS design, as discussed below.

3.1 Non-minimal state space representation

Problem (4) only involves the input and observed output because regret models only involve observations and treatment values. In linear control system design, such models are typically expressed in Transfer Function form:

$$s_k = \frac{B\left(z^{-1}\right)}{A\left(z^{-1}\right)} u_k \tag{5}$$

where s_k is the output and u_k the input, while $B(z^{-1})$ and $A(z^{-1})$ are polynomials in z^{-1}, the backward shift operator i.e. $z^{-1}u_k = u_{k-1}$. For example, section 4 utilises an illustrative model in which $B(z^{-1}) = \psi_2 z^{-1} + \psi_4 z^{-2}$ and $A(z^{-1}) = \psi_1 z^{-1} + \psi_3 z^{-2}$. More generally, denoting the highest power of z^{-i} in $B(z^{-1})$ and $A(z^{-1})$ as m and n respectively, the non-minimal state vector is defined:

$$\boldsymbol{x}_k = \begin{bmatrix} s_k & s_{k-1} & \cdots & s_{k-n+1} & u_{k-1} & \cdots & u_{k-m+1} \end{bmatrix}^T \tag{6}$$

and the NMSS representation of the TF model (5) is:

$$\boldsymbol{x}_k = \boldsymbol{F}\boldsymbol{x}_{k-1} + \boldsymbol{g}u_{k-1} \;\; ; \;\; s_k = \boldsymbol{h}\boldsymbol{x}_k \tag{7}$$

where $\boldsymbol{F}$, $\boldsymbol{g}$ and $\boldsymbol{h}$ are defined in e.g. Taylor et al. (2013) and the example below. Equations (7) provide ready access to standard control design techniques such as pole assignment or H^∞ design. Note that the order of this NMSS model is $n + m - 1$, rather than n as would be the case for a minimal state space model based on a standard canonical form. However, we are motivated to use the NMSS form for the following reasons: (i) state variable feedback control can be implemented directly rather than relying upon any form of state reconstruction; (ii) the non-minimal state vector (6) provides the most straightforward way of representing the engineering model (5) and yet also mirrors the regret formulation based on $\bar{S}_k = (S_1, S_2, \ldots, S_k)$ and $\bar{U}_k = (U_1, U_2, \ldots, U_k)$; and (iii) in comparison to the minimal case, the NMSS approach provides more design freedom as required here for ODT.

In particular, the basic NMSS form above can be extended to include other elements and can be constrained and used to mimic exactly other control approaches (Taylor et al., 2000, 2013). This is illustrated in section 5, when an integral–of–error state variable is appended to the state vector so that the control system tracks the target w_k^* (target tracking is initially addressed through eqn. (2) in the present work, hence the NMSS model (7) represents the 'regulator' form that would not normally be utilised in engineering applications).

4. COMBINED REGRET AND H^∞ APPROACH

We proceed to Monte-Carlo simulation to investigate the difference between the treatment policy defined by the classic regret approach and the one developed here. We compare these treatment policies in terms of robustness with respect to presence of parameter, dynamic and noise uncertainties. In each simulation run, 100 longitudinal data sequences of length $K = 15$ were generated (the training data sets) by using the state-space representation:

$f_k(s_k,\, u_k,\, \psi) = f_k^*(s_k,\, s_{k-1},\, u_k,\, u_{k-1},\, \psi) + g_k^{(j)}(s_{k-1},\, u_{k-1})$
with
$f_k^*(s_k,\, s_{k-1},\, u_k,\, u_{k-1},\, \psi) = \psi_1 s_k + \psi_2 u_k + \psi_3 s_{k-1} + \psi_4 u_{k-1}$
and

$$g_k^{(j)}(s_{k-1},\, u_{k-1}) = 0.01 \times j \times s_{k-1} u_{k-1}$$

The equivalent NMSS representation (7) is based on:

$$\boldsymbol{x_k} = [s_k \;\; s_{k-1} \;\; u_{k-1}]^T$$
$$\boldsymbol{F} = \begin{bmatrix} \psi_1 & \psi_3 & \psi_4 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad \begin{array}{c} \boldsymbol{g} = [\psi_2 \;\; 0 \;\; 1]^T \\ \boldsymbol{h} = [1 \;\; 0 \;\; 0] \end{array} \tag{8}$$

The true parameter values $(\psi_1,\, \psi_2,\, \psi_3,\, \psi_4)$ were taken to be $(0.6,\, 0.2,\, 0.15,\, 0.25)$ respectively and the targets w_k^* were all set to zero. The regret function model $\mu_k\left(u_k \mid \overline{S}_k, \overline{U}_{k-1}, \psi\right)$ was specified as in Section 3, and the parameters estimated from the training data set by using the regret regression approach developed by Henderson et al. (2010). Two treatment policies were considered: the classic one denoted $\mathbf{u}_{nom}$ based on (1); and $\mathbf{u}_\infty$ derived through H^∞ control design on (8) with ψ replaced by $\hat{\psi}$, and implemented using the Matlab Robust Control toolbox. Each strategy was tested by generating new longitudinal data sequences (the tested data sets) with the same state-space model as the training data set. We generated the different kinds of uncertainty as follows.

(1) Dynamic uncertainty. To add stochastic disturbance at each time k, we took $\varepsilon_k^d \sim N(0,\, \sigma_d^2)$. In order to imitate misspecification, we took as regret functions $\mu_k\left(u_k \mid \overline{S}_k, \overline{U}_{k-1}, \psi\right) =$
$$\left(f_k^*(S_k,\, S_{k-1},\, u_k,\, U_{k-1},\, \psi) - w_{k+1}^*\right)^2$$
i.e the interactions terms $g_k^{(j)}$ are mistakenly omitted.
(2) Measurement uncertainty. At each time k, we took $\varepsilon_k^m \sim N(0,\, \sigma_m^2)$.
(3) Parametric uncertainty. Instead of ψ, we used the estimator $\widehat{\psi}$ obtained from the training data.

To illustrate, we took each of σ_d^2 and σ_m^2 to be either 0.01 or 0.02 and chose j in $g_k^{(j)}$ to be either 1 or 2, the latter reflecting stronger missspecification. For each triplet $\left(g_k\,, \sigma_d^2,\, \sigma_m^2\right)$ we estimated the quantities: $\mathrm{ERR}\,[\mathbf{u}] = -\mathbb{E}_{g_k\,, \sigma_d^2, \sigma_m^2}\,[Y(\mathbf{u})]$, as well as $\mathrm{VERR}\,[\mathbf{u}] = \mathrm{Var}_{g_k\,, \sigma_d^2, \sigma_m^2}\,[Y(\mathbf{u})]$ for both $\mathbf{u}_{nom}$ and $\mathbf{u}_\infty$. Results are presented in Table 1, each based on 100 simulation runs. It is clear that the new approach based on $\mathbf{u}_\infty$ is substantially more robust than $\mathbf{u}_{nom}$.

$g_k^{(1)}$					
$\left(\sigma_d^2 \times 10^{-2},\, \sigma_m^2 \times 10^{-2}\right)$	(1, 1)	(1, 2)		(2, 1)	(2, 2)
ERR $\boxed{\mathbf{u}_{inf}}$	0.23	0.41		0.31	0.42
ERR $[\mathbf{u}_{nom}]$	7.95	7.89		5.43	7.25
VERR $\boxed{\mathbf{u}_{inf}} \times 10^{-1}$	0.03	0.10		0.20	0.35
VERR $[\mathbf{u}_{nom}] \times 10^{-1}$	20.30	42.82		20.89	49.68
$g_k^{(2)}$					
$\left(\sigma_d^2 \times 10^{-2},\, \sigma_m^2 \times 10^{-2}\right)$	(1, 1)	(1, 2)		(2, 1)	(2, 2)
ERR $\boxed{\mathbf{u}_{inf}}$	0.89	0.98		0.96	0.96
ERR $[\mathbf{u}_{nom}]$	8.57	7.67		4.55	5.65
VERR $\boxed{\mathbf{u}_{inf}} \times 10^{-1}$	0.18	0.31		0.75	0.76
VERR $[\mathbf{u}_{nom}] \times 10^{-1}$	46.85	64.16		36.48	91.88

Table 1. Simulation results

5. H^∞ EVALUATION OF NMSS/PIP CONTROL

Whilst the above analysis concerns the design of a regret-based treatment strategy that takes into account uncertainty, section 5 of the article addresses the robustness of the engineering-based PIP control algorithm. The robustness of the PIP control elements within a control system and a design strategy for stabilising the controller have been considered by other authors (Liu et al., 2001a,b). However, the overall closed-loop robustness of PIP control is generally evaluated empirically by Monte Carlo simulation (Taylor et al., 2013). Here, by contrast, we evaluate the robustness of the PIP control system using an appropriately defined H^∞ norm.

To obtain the PIP controller, we first introduce an integral-of-error state variable q_k to the NMSS model (7). This ensures that the control algorithm inherently tracks the target w_k^* i.e. Type 1 servomechanism performance in control engineering terms. Here, $q_k = q_{k-1} + (w_k^* - s_k)$ and the $n + m$ dimensional non-minimal state vector is:

$$\boldsymbol{x}_k = [s_k \; s_{k-1} \; \cdots \; s_{k-n+1} \; u_{k-1} \; \cdots \; u_{k-m+1} \; q_k]^T \tag{9}$$

Hence, the model (5) is equivalently represented using the following NMSS equations:

$$\boldsymbol{x}_k = \boldsymbol{F}\boldsymbol{x}_{k-1} + \boldsymbol{g}u_{k-1} + \boldsymbol{d}w_k^* \;\; ; \;\; s_k = \boldsymbol{h}\boldsymbol{x}_k \tag{10}$$

and the state variable feedback control law associated with this NMSS model takes the form $u_k = -\boldsymbol{k}\boldsymbol{x}_k$ where,

$$\boldsymbol{k} = [f_0 \;\; f_1 \;\; \cdots \;\; f_{n-1} \;\; g_1 \;\; \cdots \;\; g_{m-1} \;\; -k_I] \tag{11}$$

Fig. 1 shows a block diagram representation of resulting NMSS/PIP control algorithm, in which,

$$F(z^{-1}) = f_0 + f_1 z^{-1} \cdots + f_{n-1} z^{-n+1}$$
$$G(z^{-1}) = 1 + g_1 z^{-1} + \cdots + g_{m-1} z^{-m+1}$$

For the preliminary results in the present article, we evaluate the robustness when $w_k^* = 0$ for all k and the external input consists of a load disturbance d_k at the plant input (similar to the random variable ε_k^d described in section 3) and generalised noise n_k at the plant output (equivalent to the measurement uncertainty ε_k^m). The corresponding H^∞ norm describes the robustness of the closed-loop system to process uncertainties, hence minimising this norm is equivalent to increasing the control system robustness.

If the generalised error is $\zeta = (s, -u)^T$ and input is $\xi = (n, d)^T$, the closed loop system is denoted $\zeta = H_{FB}(P, C)\xi$, where,

$$\zeta = \begin{pmatrix} \dfrac{AG\Delta}{AG\Delta + B(F\Delta + k_I)} & \dfrac{BG\Delta}{AG\Delta + B(F\Delta + k_I)} \\ \dfrac{A(F\Delta + k_I)}{AG\Delta + B(F\Delta + k_I)} & \dfrac{B(F\Delta + k_I)}{AG\Delta + B(F\Delta + k_I)} \end{pmatrix} \xi,$$

in which $\Delta = 1 - z^{-1}$. Here A, B, F, G and B are polynomials in z^{-1} but the operator notation has been omitted for brevity, both here and in later equations. The H^∞ norm of the closed-loop system thus provides a measure of robustness of the PIP controller that is dependent upon the controller parameters and plant model. If the closed-loop control system is denoted $P(z^{-1})$, then the H^∞ norm is defined as follows (Stoorvogel A. A., 1994),

$$\|P(z^{-1})\|_\infty := \sup_{\theta \in (0,\pi)} \bar{\sigma}\left(P(e^{-i\theta})\right) \tag{12}$$

where $\bar{\sigma}(\cdot)$ denotes the largest singular value of the matrix $P(e^{-i\theta})$. This relationship can be used to calculate the H^∞ norm of the PIP control system as follows,

$$\|H_{FB}\|_\infty = \sup_{\theta \in (0,\pi)} \sqrt{\frac{Q}{\overline{(AG\Delta + B(F\Delta + k_I))}(AG\Delta + B(F\Delta + k_I))}} \tag{13}$$

where,

$$Q = \overline{AG\Delta} AG\Delta + \overline{BG\Delta} BG\Delta + \overline{A(F\Delta + k_I)} A(F\Delta + k_I) + \overline{B(F\Delta + k_I)} B(F\Delta + k_I) \tag{14}$$

and $\bar{X}$ denotes the complex conjugate of X. This equation provides a way of evaluating the H^∞ norm of a PIP control system, as well as a method for designing control parameters F, G, k_I to minimise the H^∞ norm.

To briefly illustrate, the state vector (8) for the previously developed regret-based NMSS model is expanded to now include q_k as follows,

$$\boldsymbol{x_k} = [s_k \ s_{k-1} \ u_{k-1} \ q_k]^T$$

with,

$$\boldsymbol{F} = \begin{bmatrix} \psi_1 & \psi_3 & \psi_4 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -\psi_1 & -\psi_3 & -\psi_4 & 1 \end{bmatrix}$$

$\boldsymbol{g} = [\psi_2 \ 0 \ 1 \ -\psi_2]^T$, $\boldsymbol{d} = [0 \ 0 \ 0 \ 1]^T$ and $\boldsymbol{h} = [1 \ 0 \ 0 \ 0]$. In Fig. 2 we use pole placement (fully described in Taylor et al. (2013)) to set F, G, and k_I. For the purpose of this example, the four desired closed-loop poles are constrained to the same value on the real axis of the complex z–plane. Fig.2 shows how the choice of this constrained real pole position, and so controller parameters, effects the H^∞ norm. In further research, we will compare PIP/NMSS control with the regret-based NMSS approach for both medical and engineering applications.

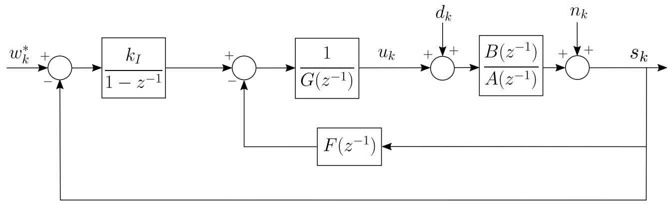

Fig. 1. Block diagram of the PIP control system.

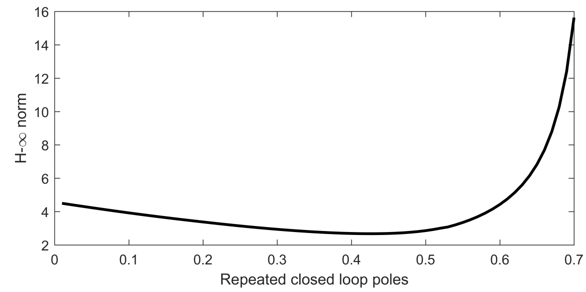

Fig. 2. H^∞ norm of NMSS/PIP control system for a range of (real) pole values.

6. CONCLUSIONS

The article has proposed a method for designing a medical treatment strategy that takes into account model and measurement uncertainty. We have shown how a particular class of ODT problem can be transformed into an optimal control problem and have used H^∞ synthesis to derive a treatment (input) policy. The state-space representation used in problem (4) only involves the input and observed output, and their past values, hence non-minimal state space models that are explicitly based on these variables are particularly apt. Using this formulation, a novel way of designing a treatment strategy that is robust to model misspecification and measurement noise is presented. This involves using statistical methods to estimate the regret parameters and, by connecting regret functions to the state-space representation, subsequently using robust H^∞ to estimate a treatment (input) policy.

Comparison with classic A-Learning (the nominal method in this article) shows that the new H^∞ approach yields a better outcome on average when misspecification is present and is less sensitive to other perturbations. In other words, the new method is more robust. Currently, this H^∞ design approach is restricted to a particular class of ODT problems that have linear time invariant state-space models and so simple quadratic regret functions. This limitation needs to be overcome by addressing how to transform other classes of ODT problems into a control problem and finding a more general way of linking the regret function and state-space formulation.

Other control methods such as NMSS/PIP design can also be used to optimise the treatment strategy. A simple PIP control system is presented in this article for comparison, and we have shown how the choice of controller parameters effects the robustness of the system. Furthermore, NMSS/PIP control includes an integral-of-error state that is not considered in conventional A-learning. Including such action in ODT type methods could improve target tracking by removing steady state errors.

To summarise, three methods of controlling a simple, linear system have been presented: i) A-learning; ii) a novel robust H^{∞} state-space-regret approach; and iii) standard NMSS/PIP control. Here, we have compared the A-learning and robust H^{∞} state-space-regret approach. However, we have not compared, or looked at combining NMSS/PIP (including the integral-of-error state) with robust state-space-regret methods, and this remains a key challenge and an area of on-going research by the authors.

REFERENCES

Barrett, J., Henderson, R., and Rosthoj, S. (2014). Doubly robust estimation of optimal dynamic treatment regimes. *Statistics in Biosciences*, 6, 244–260.

Cangar, O., Aerts, J.M., Vranken, E., and Berckmans, D. (2007). Online growth control as an advance in broiler farm management. *Poultry Science*, 86, 439–43.

Chakraborty, B. and Moodie, E. (2013). *Statistical Methods for Dynamic Treatment Regimes.* Springer, New York.

Clarke, F. (2013). *Functional Analysis, Calculus of Variations and Optimal Control.* Springer-Verlag London, London.

Doyle, J., Glover, K., Khargonekar, P., and Francis, B. (1989). State-space solutions to standard h-infinity and h2 control problems. *IEEE Transacations on Automatics Control*, 34, 831–847.

Doyle III, F.J., Bequette, B.W., Middleton, R., Ogunnaike, B., Paden, B., Parker, R.S., and Vidyasagar, M. (2011). Control in biological systems. In T. Samad and A.M. Annaswamy (eds.), *The Impact of Control Technology.* IEEE Control Systems Society, www.ieeecss.org.

Francis, B. (1987). *Lecture Notes in Control and information Sciences.* Spinger-Verlag, Berlin.

Glover, K. and Doyle, J. (1988). State-space formulae for all stabilizing controllers that satisfy an h-infinity-norm bound and relations to risk sensitivity. *Systems and Control Letters*, 167–172.

Henderson, R., Ansell, P., and Alsibani, D. (2010). Regret-regression for optimal dynamic treatment regimes. *Biometrics*, 66, 1192–1201.

Henderson, R., Ansell, P., and Alsibani, D. (2011). Optimal dynamic treatment methods. *Revstat Statistical Journal*, 9, 19–36.

Iwasaki, T. and Skelton, R.E. (1994). All controllers for the ceneral h-infinity control problem: LMI existence conditions and state space formulas. *Automatica*, 30, 1307–1317.

Liu, G.P., Dixon, R., and Daley, S. (2001a). Design of stable proportional-integral-plus controllers. *International Journal of Control*, 74, 1581–1587.

Liu, G.P., Duan, G.R., and Dixon, R. (2001b). Robust control with stable proportional-integral-plus controllers. *European Control Conference*, 3201–3206.

Ljung, L. (1999). *System Identification: Theory for the User.* Prentice Hall, New Jersey.

Moodie, E., Richardson, T., and Stephens, D. (2007). Demystifying optimal dynamic treatment regimes. *Biometrics*, 63, 447–455.

Murphy, S. (2003). Optimal dynamic treatment regimes. *J. Royal Statistical Society B*, 65, 331–366.

Mustafa, D. and Glover, K. (1990). *Minimum entropy H–infinity control.* Springer–Verlag, Berlin.

Rich, B., Moodie, E., and Stephens, D. (2016). Optimal individualized dosing strategies: a pharmacologic approach to developing dynamic treatment regimens for continuous-valued treatments. *Biometrical Journal*, 58, 512–517.

Robins, J.M. (2004). Optimal structural nested models for optimal sequential decisions. In D.Y. Lin and P. Heagerty (eds.), *Proceedings of the Second Symposium on Biostatistics*, 189–326. Springer, New York.

Rosthoj, S., Fullwood, C., Henderson, R., and Stewart, S. (2006). Estimation of optimal dynamic anticoagulation regimes from observational data: a regret-based approach. *Stat Med*, 4197-4215(25).

Rosthoj, S., Keiding, N., and Schmeiglow, K. (2012). Estimation of dynamic treatment strategies for maintenance therapy of children with acute lymphoblastic leukaemia: an application of history-adjusted marginal structural models. *Statistics in Medicine*, 31(5), 470–488.

Schulte, P., Tsiatis, A., Laber, E., and Davidian, M. (2014). Q- and A-learning methods for estimating optimal dynamic treatment regimes. *Statistical Science*, 29, 640–661.

Sontag, E. (1998). *Mathematical Control Theory: Deterministic Finite-Dimensional Systems.* Springer-Verlag, New York.

Stoorvogel A. A., Saberi A., C.B.M. (1994). The discrete-time h-∞ control problem with measurement feedback. *International Journal of Robust and Nonlinear Control*, 4, 457–479.

Taylor, C.J. and Aerts, J.M. (2014). Control of nonlinear biological systems by non–minimal state variable feedback. *Statistics in Biosciences*, 6, 290–313.

Taylor, C.J., Chotai, A., and Young, P.C. (2000). State space control system design based on non–minimal state–variable feedback: further generalisation and unification results. *International Journal of Control*, 73(14), 1329–1345.

Taylor, C.J., Young, P.C., and Chotai, A. (2013). *True Digital Control: Statistical Modelling and Non–Minimal State Space Design.* John Wiley and Sons.

Wallace, M. and Moodie, E. (2015). Doubly-robust dynamic treatment regimen estimation via weighted least squares. *Biometrics*, 71, 636–644.

Wallace, M., Moodie, E., and Stephens, D. (2016). Model assessment in dynamic treatment regimen estimation via double robustness. *Biometrics*, 72, 855–864.

Zhang, B., Tsiatis, A., Laber, E., and Davidian, M. (2013). Robust estimation of optimal dynamic treatment regimes for sequential treatment decisions. *Biometrika*, 100, 681–694.

Zhou, K., Doyle, J.C., and Glover, K. (1995). *Robust and Optimal Control.* Prentice Hall, NY.