Online Control For Adaptive Tapering Of Medications

Paula Gradu and Benjamin Recht

August 2023

Abstract

We investigate adaptive protocols for the elimination or reduction of the use of medications or addictive substances. We formalize this problem as online optimization, minimizing the cumulative dose subject to constraints on well-being. We adapt a model of addiction from the psychology literature and show how it can be described by a class of linear time-invariant systems. For such systems, the optimal policy amounts to taking the smallest dose that maintains well-being. We derive a simple protocol based on integral control that requires no system identification, only needing approximate knowledge of the instantaneous dose response. This protocol is robust to model misspecification and is able to maintain an individual's well-being during the tapering process. Numerical experiments demonstrate that the adaptive protocol outperforms non-adaptive methods in terms of both maintenance of well-being and rate of dose reduction.

1 Introduction

Tapering medications and assisting cessation of addictive substances are related challenges in health care. Relapse rates are high, and many return to drug use within weeks of entering treatment. Though no general protocols are equally valuable to all medications or all people, a personalized approach to cessation may allow caregivers flexibility to meet the diverse needs of a diverse care-seeking population. This paper investigates adaptive tapering protocols that enable individuals to self-regulate their cessation rate. In particular, we formalize tapering as an optimization problem of minimizing the cumulative dose subject to constraints on well-being.

Such a formulation requires modeling an individual's dose-response dynamics, and we propose a mathematical formulation of the *opponent process* model of addiction due to Solomon [16]. As we discuss in Section 2, an opponent process consists of two competing systemic reactions to treatment. The first system governs the "positive" effect with a rapid onset and fast subsequent decay. The second system governs "negative effects" with a long latency and slow decay. In Section 3, we show that modeling opponent processes as LTI systems captures the qualitative behaviors specified in the psychology literature and allows us to analyze tapering as an optimal control problem.

In Section 4, we first derive a greedy policy that solves the optimal control problem. The dosage should be decreased by as much as possible while maintaining constraints on well-being, but no look-ahead or planning is needed to compute the dose. We provide a controller that maintains this greedy behavior and is optimal under various models of possible exogenous disturbances.

In reality, the opponent process model is an approximation, and the particular model of an individual is unknown. A possible approach informed by control theoretic practice might involve learning a dynamics model of the individual, but such learning procedures require wide varying of inputs and would not be feasible for patients attempting drug cessation. Instead of leaning on system identification, we investigate the potential of "model-free" controllers for tapering in

Section 5. We analyze a simple integral controller, where the error signal is the deviation from a minimally acceptable level of well-being. Using methods from the online learning literature, we show that integral control robustly reduces the dosage while minimally violating constraints on well-being. Under further mild assumptions about the time between doses, we show that the integral controller monotonically reduces the dose to zero in finite time without violating the specified constraints. In the numerical experiments of Section 6, we show that our adaptive protocol outperforms non-adaptive methods in terms of both maintenance of well-being and rate of dose reduction.

2 Related Work

Tapering Protocols. Our work attempts to synthesize adaptive protocols that can be applied to both "deprescribing" medications via gradual tapering and to managing the symptoms of withdrawal from addictive substances. In both of these applications, most studies of tapering protocols have focused on non-personalized procedures. Mujika and Padilla [14] study tapering protocols for SSRI medications. They find that short tapers (2-4 weeks) have minimal benefits over abrupt discontinuation while longer tapers (multiple months) are more successful at minimizing withdrawal. More specifically they suggest a slow, hyperbolic dosage reduction. Horowitz et al. [6, 7] propose a similar method (very slow, hyperbolic) for tapering of antipsychotic medication. This work is based on case studies and notes that there is no standard guideline for tapering antipsychotic medications.

Several studies have also investigated protocols for managing withdrawal from addictive substances. In a meta-analysis of opiod tapering protocols, Berna et al. [2] found that longer tapers are typically better. The survey by Fenton et al. [3] argues that most tapering protocols which focus primarily on getting through the acute withdrawal phase may be too rapid and have negative mental health consequences. In an observational study, Agnoli et al. [1] find that tapering of opiods is significantly associated with increased risk of overdose and mental health crisis. Henry et al. [5] attempt to characterize patients' subjective experiences with opiod tapering in an effort to minimize negative tapering reactions.

Models of Drug Response and Tolerance. In the present work, we will build upon a popular model of addiction, proposed by Solomon [16], called the Opponent-Process theory of acquired motivation. In this model, the drug response is the result of an initial 'positive' effect with short lag and fast decay (A process) followed by a counter 'negative' effect with high latency and slow decay (B process). This formalizes the empirical qualitative observation that the withdrawal symptoms of a drug are characterized as opposite to its acute effects [8]. The shape of such a response is plotted in Figure 1.

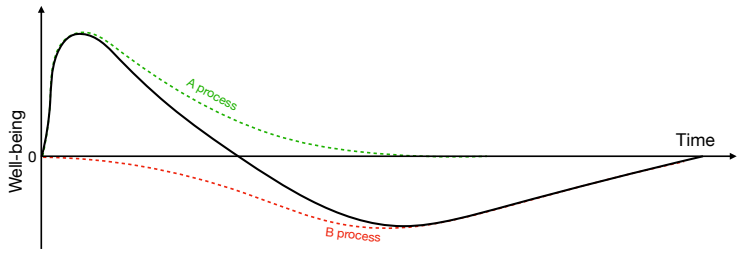

Figure 1: Shape of an Instantaneous Opponent Process response as described by Solomon [16].

Koob and Moal [10] and Koob [9] extend the opponent process framework to account for a chronic deviation of the regulatory system from baseline under repeated administration of the drug, which is referred to as allostasis and is illustrated in Figure 2. In this work, we propose a simple mathematical formulation of the opponent process that captures the salient aspects of this model.

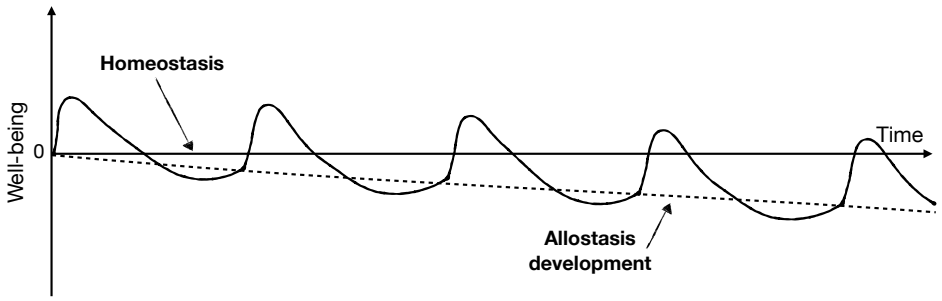

Figure 2: Allostasis development due to overlay of opponent processes over time as described by Koob and Moal [10].

3 Problem Setting and Preliminaries

To mathematically formalize the opponent process model, we will denote an individual's well-being at any time as y_t. This formulation asserts that y_t is a concrete numerical quantity, but our methods can operate with imprecise quantification of well-being. For example, for modeling opiate withdrawal, y_t could be a value on the Clinical Opiate Withdrawal Scale [17] or the Subjective/Objective Opiate Withdrawal Scales [4]. For modeling an individual's response to anti-depressants, y_t could model the number of Discontinuation-Emergent Signs and Symptoms (DESS) [15], a depression assessment like PHQ-9 [11], or a combination of both.

Let u_t denote the amount of a drug ingested at time t. We assume that y_t evolves according to an LTI system

$$y_{t+1} = \sum_{k=0}^{t} g(k) u_{t-k} + y_{t+1}^{\text{nat}} \tag{3.1}$$

where g is the impulse response. We focus our attention on linear models as they are sufficiently general to capture much of the qualitative behavior observed in the psychology literature, but we discuss potential extensions to nonlinear models in the Appendix. y^{nat} comprises the remaining dynamics that cannot be attributed to the drug under study. We will only lightly model this exogenous disturbance by asserting lower bounds on its values or bounds on its rate of change. The time $t = 0$ will be the moment an individual adopts a tapering protocol. The effect of all inputs before $t = 0$ and any measurement noise are implicitly represented in the "natural progression" signal, y^{nat}.

We formalize the goal of tapering as minimizing the total consumption of a drug subject to keeping well-being above some specified threshold. The total consumed drug is $\sum_{t=0}^{T-1} u_t$ and the constraint on well-being is $y_t \geq y_{\min}$ for all t from 1 to T.

Modeling Opponent Processes. We now turn to our formal model of an opponent process. Solomon [16] asserts that an opponent process reaction to a single dose is first an A-process where $y > 0$ and then a following B-process where $y < 0$. The following definition summarizes this property.

Definition 3.1 (Opponent Process)**.** *We say that g is an opponent process if there exists a time τ_0 such that $g(\tau) > 0$ when $\tau < \tau_0$ and $g(\tau) \leq 0$ if $\tau \geq \tau_0$.*

In this work, we focus on a subclass of opponent processes that both induce some notion of tolerance or addiction and are sufficiently well-behaved to be "treatable."

Definition 3.2. *An opponent process g is a* Linearly Progressing Opponent Process *(LPOP) if there additionally exists an $\alpha \in (0, 1)$ such that*

$$g(t+1) \leq \alpha \cdot g(t) \quad \textit{for} \quad t < \tau_0 - 1$$
$$|g(t+1)| \geq \alpha \cdot |g(t)| \quad \textit{for} \quad t \geq \tau_0 \,.$$

This definition quantifies opponent processes where the A-process decays exponentially, and the B-process does not decay as quickly as the A-process. For example, this would model a drug where the positive effects decay at some rate α, then become negative, trending towards some peak of withdrawal, and finally decay back to 0, potentially at a prolonged rate.

Examples of Opponent Processes. Consider a linear system with impulse response

$$g(t) = \sum_{\lambda} c_\lambda \lambda^t$$

where all of the $\lambda \in [0, 1)$ and the c_λ are real valued scalars. The terms where c_λ are positive correspond to effects that increase well-being. The terms where c_λ are negative decrease well-being. With this in mind, define $\Lambda_+ = \{\lambda \,:\, c_\lambda \geq 0\}$ and $\Lambda_- = \{\lambda \,:\, c_\lambda < 0\}$. Then, the system is an LPOP if the following three conditions hold:

1. $\max_{\lambda \in \Lambda_+} \lambda \leq \min_{\lambda' \in \Lambda_-} \lambda'$
2. $\sum_\lambda c_\lambda > 0$
3. There exists some τ_0 s.t. $g(\tau_0) \leq 0$

Intuitively, this holds because the part of the system corresponding to positive effects decay more rapidly than all of the terms corresponding to negative effects. In particular, the second condition implies an initial positive response, and third condition implies that after a single dose, the well-being will eventually be negative.

The formal argument of why such systems are LPOPs proceeds by bounding $g(t+1)$ in terms of $g(t)$. Let $\Lambda_{++} := \max_{\lambda \in \Lambda_+} \lambda$ and $\Lambda_{--} := \min_{\lambda' \in \Lambda_-} \lambda'$. Then we have

$$
\begin{aligned}
g(t+1) = \sum_{\lambda} c_\lambda \lambda^{t+1} &= \sum_{\lambda_+ \in \Lambda_+} c_{\lambda_+} \lambda_+^{t+1} + \sum_{\lambda_- \in \Lambda_-} c_{\lambda_-} \lambda_-^{t+1} \\
&\leq \Lambda_{++} \cdot \sum_{\lambda_+ \in \Lambda_+} c_{\lambda_+} \lambda_+^{t} + \Lambda_{--} \cdot \sum_{\lambda_- \in \Lambda_-} c_{\lambda_-} \lambda_-^{t} \\
&\leq \Lambda_{--} \left(\sum_{\lambda_+ \in \Lambda_+} c_{\lambda_+} \lambda_+^{t} + \sum_{\lambda_- \in \Lambda_-} c_{\lambda_-} \lambda_-^{t} \right) \\
&= \Lambda_{--} g(t)
\end{aligned}
$$

Setting $\alpha \doteq \Lambda_{--}$, this implies that for $t < \tau_0 - 1$, $g(t+1) \leq \alpha g(t)$ and for $t \geq \tau_0$, $|g(t+1)| \geq \alpha |g(t)|$.

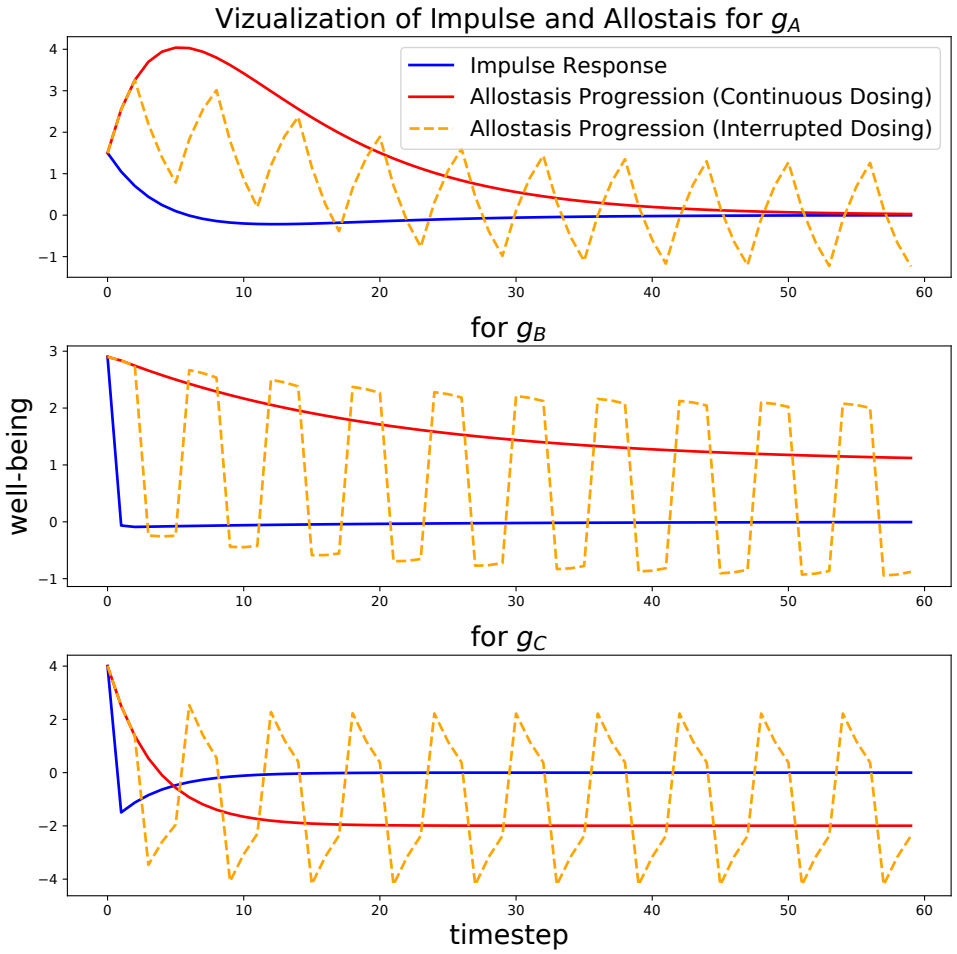

Figure 3: Impulse responses and allostasis progression over 60 timesteps for three opponent processes, each with different time constants and relative effect magnitudes. The first two panels illustrate an opponent processes with immediate benefits and slowly accumulating tolerance, characteristic of therapeutic medications. The third panel captures an opponent process with immediate transient effect and also a rapid onset of a low value of well-being, more characteristic of subsances of abuse.

4 Optimal tapering protocol with system knowledge

In this section, we derive the optimal tapering protocol when the quantities described in the previous section are known to the care provider. In particular, we will show that when the system dynamics are known, the optimal control problem is equivalent to setpoint matching with setpoint $y_{\min}$.

We begin with the formal definition of a *maximally greedy* dose, the smallest dose that does not lead to an immediate violation of the constraint on y. The main theorem of this section states that a maximally greedy protocol results in the lowest cumulative dose.

Definition 4.1 (Maximally Greedy Dose)**.** *We say that a dose u_t is* maximally greedy *at time t if it either leads to $y_{t+1} = y_{\min}$ or is equal to 0.*

Theorem 4.2. *For any LPOP g, taking the maximally greedy dose at every time maintains $y_t \geq y_{\min}$ for all t with minimal cumulative dose.*

Proof. Consider a dosing sequence $u_t^\star$ that maintains $y_t^\star \geq y_{\min}$ for all t with minimal cumulative dose. We will prove the theorem by contradiction, assuming the sequence $u_t^\star$ is not maximally greedy. Then there must exist t_0 and $\epsilon > 0$ such that replacing $u_{t_0}^\star$ with $u_{t_0}^\star - \epsilon$ would result in a new $y'_{t_0+1} < y^\star_{t_0+1}$ with $y'_{t_0+1} \geq y_{\min}$. Consider the modified dosing schedule

$$u'_t = \begin{cases} u^\star_{t_0} - \epsilon & t = t_0 \\ u^\star_{t_0+1} + \alpha\epsilon & t = t_0 + 1 \\ u^\star_t & \text{otherwise} \end{cases}$$

and let y'_t be the resulting sequence. Since $u'_t = u^\star_t$ for all $t < t_0$, we have that $y'_t = y^\star_t \geq y_{\min}$ up to t_0. Plugging in the effect of the dose modifications at t_0 and $t_0 + 1$ into the dynamics equation (3.1), we get:

$$\begin{aligned} y'_{t+1} &= y^\star_{t+1} - g(t - t_0)\epsilon + g(t - t_0 - 1)\alpha\epsilon \\ &\geq y^\star_{t+1} && (g \text{ LPOP}) \\ &\geq y_{\min} \end{aligned}$$

since this holds for all $t > t_0$, we have that y'_t satisfies the constraints for all $1 \leq t \leq T$. But we note that u'_t has a smaller cumulative dose than $u^\star_t$ because

$$\sum_{t=1}^{T} u'_t = \sum_{t=1}^{T} u^\star_t - \epsilon + \alpha\epsilon < \sum_{t=1}^{T} u^\star_t \,.$$

This is a contradiction, completing the proof. □

This theorem shows that long-term planning is unnecessary to optimally taper a medication. It always suffices to take the smallest dose necessary to maintain well being at the current time. The question remains if such myopic strategies remain possible when properties of the dynamics and dose responses are unknown.

Let us first assume that we know the impulse response g. Our first protocol considers the case when we have a lower bound on the values of y_t^{nat}.

Proposition 4.3. *For any LPOP g, assume we know a lower bound $y_t^{\text{nat,lb}}$ such that $y_t^{\text{nat,lb}} \leq y_t^{\text{nat}}$ for all t. Then for any instance of y_t^{nat} that is greater than the prescribed lower bound, the dosing sequence*

$$u_t = \max\left\{0, \frac{y_{\min} - y_{t+1}^{\text{nat,lb}} - \sum_{k=1}^{t} g(k)u_{t-k}}{g(0)}\right\} \tag{4.1}$$

maintains $y_t \geq y_{\min}$ for all t with minimal cumulative dose.

Proof. To prove this proposition, we first derive the maximally greedy dose at time t. This dose will require access to y_{t+1}^{nat} – we will replace this term with its lower bound and show the resulting protocol is optimal.

The maximally greedy dose at time t (given past doses) is

$$u_t^{\text{MG}} = \max\left\{0, \frac{y_{\min} - y_{t+1}^{\text{nat}} - \sum_{k=1}^{t} g(k)u_{t-k}}{g(0)}\right\} \tag{4.2}$$

To see why, observe that the dose is either 0 or satisfies

$$\begin{aligned} y_{t+1}^{\text{MG}} &= u_t^{\text{MG}} g(0) + \sum_{k=1}^{t} g(k)u_{t-k} + y_{t+1}^{\text{nat}} \\ &= y_{\min} - y_{t+1}^{\text{nat}} - \sum_{k=1}^{t} g(k)u_{t-k} + \sum_{k=1}^{t} g(k)u_{t-k} + y_{t+1}^{\text{nat}} \\ &= y_{\min} \end{aligned}$$

For any $y_{t+1}^{\text{nat}} \geq y_{t+1}^{\text{nat,lb}}$, the dose given by (4.1) is larger than the maximally greedy dose (i.e. $u_t \geq u_t^{\text{MG}}$) and therefore maintains $y_{t+1} \geq y_{t+1}^{\text{MG}} \geq y_{\min}$. This holds independently for all t (irrespective of past doses) and hence (4.1) maintains $y_t \geq y_{\min}$ for all t for any sequence y_t^{nat} such that $y_t^{\text{nat}} \geq y_t^{\text{nat,lb}}$ for all t.

Finally, since the sequence of inputs given in (4.2) is maximally greedy with respect to the lower bound sequence $y_y^{\text{nat,lb}}$, Theorem 4.2 implies that any dosing schedule with a strictly lower cumulative dose would have to violate one of the constraints. This means that u_t given by (4.1) results in the minimal cumulative dose that maintains feasibility across all admissible y_t^{nat} sequences (i.e. lower bounded by $y_t^{\text{nat,lb}}$). □

There are two natural choices to lower bound y_t^{nat}. First, we may assume that $y_{t+1}^{\text{nat}} \geq y_t^{\text{nat}}$ for all t. In this case, the best lower bound candidate is $y_{t+1}^{\text{nat,lb}} = y_t^{\text{nat}}$ which we can compute via

$$y_t^{\text{nat}} = y_t - \sum_{k=0}^{t-1} g(k)u_{t-k-1}\,. \tag{4.3}$$

Second, we could bound the variation of y_t^{nat}, assuming there $L_{\text{nat}} \geq 0$ such that $y_{t+1}^{\text{nat}} \geq y_t^{\text{nat}} - L_{\text{nat}}$ for all t, In this case, the best lower bound is $y_{t+1}^{\text{nat,lb}} = y_t^{\text{nat}} - L_{\text{nat}}$, where we again compute y_t^{nat} via Eq. 4.3.

Note that the proof of Theorem 4.2 is valid even if the constraints on y or the impulse response g are time-varying. Hence, even under more general settings, setpoint matching is the optimal

approach to tapering. The doses for the time-varying setting are the same as in Proposition 4.3 but with $y_{\min}^{(t)}$ and/or g_t plugged into the formula. In de-medication, time-varying constraints are relevant when the patient can withstand more severe withdrawal effects at particular times of their lives or when they can no longer tolerate an initially agreed-upon baseline.

5 Tapering by Integral Control

Computing the maximally greedy dose in the previous section required knowledge of the model of the underlying dynamics, g. In practice, we would like to avoid estimating g and instead derive an update rule requiring minimal system knowledge. Since we know that the optimal control law is setpoint matching and that integral action can match setpoints asymptotically, we analyze the performance of a simple integral controller on the tapering problem in this section. We demonstrate that integral control with appropriately chosen gains can reduce a dose while rarely violating the constraint on y_t. One need only know an approximate value of the immediate effect of a single dose, which is equal to the value $g(0)$. Moreover, if $g(t) \leq 0$ for all $t \geq 1$, then integral control results in a dose sequence monotonically decreases to zero in finite time while ensuring $y_t \geq y_{\min}$ for all t.

To proceed, let $(x)_+ := \max(x, 0)$ and $(x)_- := \min(x, 0)$ Consider the integral control law

$$u_t = \max\left\{0, u_{t-1} - K_+(y_t - y_{\min})_+ - K_-(y_t - y_{\min})_-\right\}. \tag{5.1}$$

This protocol is integral control if $K_+ = K_-$. However, it also allows for a different gain when y_t is above and below the desired set point $y_{\min}$. Using two gains allows for more conservative (i.e., higher) doses. The control action is clipped at 0 as it is impossible to take a negative dose.

The main theorem of this section bounds the *long-term constraint violation* [12, 13] of the integral controller. We aim to study how the average value of y_t compares to the constraint $y_{\min}$ over time. The bound we derive ensures the overall sequence of states satisfies the constraint more strictly as time increases. This ensures that the running average of the observations up until a time T is greater than $y_{\min}$ minus a penalty decaying at rate $\frac{1}{T}$.

Theorem 5.1. *For any underlying y^{nat}, (5.1) ensures that for any T*

$$\frac{\sum_{t=1}^{T} y_t}{T} \geq y_{\min} - \frac{y_0 - y_{\min}}{T},$$

provided $K_+ \leq g(0)^{-1}$ and $K_- \geq g(0)^{-1}$.

The proof of this theorem is broadly inspired from analyses of the finite time behavior of average constraint violation in regret minimization by Mannor and Tsitsiklis [13] and Mahdavi et al. [12]. To prove the theorem, we need the following technical lemma that lower bounds the instance-wise deviation of the y_t obtained from the dosing scheme (5.1) from $y_{\min}$ if the gains K_+/K_- upper/lower bound $g(0)^{-1}$ respectively. The proof of the lemma is given in the Appendix.

Lemma 5.2. *Adopt the convention that $u_{-1} = 0$. For any gains satisfying $K_+ \leq g(0)^{-1}$ and $K_- \geq g(0)^{-1}$, setting u_t according to (5.1) for $t \geq 0$ yields y_{t+1} satisfying*

$$y_{t+1} \geq y_{\min} + (y_{t+1}^{\text{nat}} - y_t^{\text{nat}}) - \sum_{k=1}^{t} g(k)(u_{t-k-1} - u_{t-k}).$$

With this lemma, we can now prove Theorem 5.1.

of Theorem 5.1. Using Lemma 5.2, we have that

$$\begin{aligned}\sum_{t=0}^{T} y_{t+1} &\geq (T+1)y_{\min} \\ &\quad + \sum_{t=0}^{T}\left((y_{t+1}^{\text{nat}} - y_t^{\text{nat}}) - \sum_{k=1}^{t} g(k)(u_{t-k-1} - u_{t-k})\right) \\ &= (T+1)y_{\min} - y_0^{\text{nat}} + y_{T+1}^{\text{nat}} + \sum_{t=0}^{T} u_t g(T-t) \\ &= (T+1)y_{\min} + y_{T+1} - y_0\end{aligned}$$

where the first equality follows by telescoping and the second by noting that $y_0^{\text{nat}} = y_0$. By cancelling y_{T+1} on both sides, we get

$$\sum_{t=0}^{T-1} y_{t+1} \geq T y_{\min} - (y_0 - y_{\min})$$

which can be immediately rearranged into the stated bound:

$$\frac{\sum_{t=1}^{T} y_t}{T} = \frac{\sum_{t=0}^{T-1} y_{t+1}}{T} \geq y_{\min} - \frac{y_0 - y_{\min}}{T}$$

□

Note that if it is undesirable to fluctuate around $y_{\min}$, we can run a padded integral controller adding δ to $y_{\min}$. Theorem 5.1 then implies

$$\frac{\sum_{t=1}^{T} y_t}{T} \geq y_{\min} + \delta - \frac{y_0 - y_{\min} - \delta}{T}.$$

Also note that, just as in the previous section, we can extend the integral controller (5.1) to time-varying constraints by plugging in $y_{\min}^{(t)}$. In this case, the guarantee from Theorem 5.1 would instead bound deviation from the average $y_{\min}$, i.e. from $T^{-1}\sum_{t=1}^{T} y_{\min}^{(t)}$.

We close this section by deriving stronger guarantees for a special class of LPOPs. Consider an opponent process with $\tau_0 = 1$. Requiring $\tau_0 = 1$ can be interpreted as an assumption on the discretization of the dose updating scheme (e.g., every day vs. every week). For example, we can turn a system g with $\tau_0 > 1$ into a system g' with $\tau_0' = 1$ by taking $g'(t) = \tau_0^{-1} \cdot \sum_{t'=t\cdot\tau_0}^{(t+1)\cdot\tau_0 - 1} g(t')$.

Proposition 5.3. *For any g with $\tau_0 = 1$, any initial u_0 which ensures $y_1 \geq y_{\min}$, and any natural progression sequence satisfying $y_{t+1}^{\text{nat}} \geq y_t^{\text{nat}} - g(t)u_0 + \delta/t$ for some (arbitrarily small) $\delta > 0$, the doses prescribed by (5.1): (i) are non-increasing, (ii) maintain $y_t \geq y_{\min}$ for all t, and (iii) there exists some time T_0 such that $u_t = 0$ for all $t \geq T_0$.*

Proof. To prove (i) and (ii) jointly, we will proceed by induction on t. The base case holds by the assumption that we have access to an initial u_0 for which $y_1 \geq y_{\min}$. Assume for the inductive hypothesis that $0 \leq u_{t-1} \leq \ldots \leq u_0$ and $y_t \geq y_{\min}$. First, observe that since $y_t - y_{\min} \geq 0$, we have

$$u_t \leq \max\{0, u_{t-1}\} = u_{t-1}$$

showing the inductive step for (i). Second, by Lemma 5.2,

$$\begin{aligned} y_{t+1} &\geq y_{\min} + (y_{t+1}^{\text{nat}} - y_t^{\text{nat}}) - \sum_{k=1}^{t} g(k)(u_{t-k-1} - u_{t-k}) \\ &\geq y_{\min} + (y_{t+1}^{\text{nat}} - y_t^{\text{nat}}) + g(t)u_0 \\ &\geq y_{\min} + \delta/t \end{aligned}$$

where the second-to-last inequality follows since $g(k) \leq 0$ and $u_{t-k-1} - u_{t-k} \geq 0$ for all $1 \leq k \leq t-1$. The last inequality follows by our assumption $y_{t+1}^{\text{nat}} \geq y_t^{\text{nat}} - g(t)u_0 + \delta/t$. This shows the inductive step for (ii). Therefore, by induction, (i) and (ii) are true.

Finally, by the calculation above, we know that u_t is decreasing and, further, that it can be expressed as

$$u_t = \max\left\{0, u_0 - K_+\delta \cdot \sum_{k=1}^{t} \frac{1}{t}\right\}$$

which, since $\sum_{k=1}^{t} 1/t \geq \ln t$, implies $u_t = 0$ for all $t \geq \exp\left\{\dfrac{u_0}{K_+\delta}\right\}$, proving (iii). □

6 Numerical Simulations

In this section, we present simulations demonstrating the behavior of the proposed protocols and standard (non-adaptive) tapering approaches. Due to a lack of available simulators/open source data, we rely on synthetic impulse response functions that we model based on broad properties of immediate response and allostatic adaptation.

Task definition We consider the problem of tapering in the three opponent processes from Figure 3, starting from the allostatic state reached after taking dose $u = 1$ for $T_{\text{init}} = 60$ timesteps. For each of these opponent processes, we are interested in the ability of a protocol to perform well on average for a population with different constraints. More concretely, for each setting, we consider 'treating' a population of N units over a horizon T_{taper} with $y_{\min}$ distributed uniformly between the allostatic equilibrium and a lower baseline. Concretely, we take $y_{\min}^A \sim \text{Unif}(-2, 0)$, $y_{\min}^B \sim \text{Unif}(-1, 1)$, $y_{\min}^C \sim \text{Unif}(-4, -2)$ and $T_{\text{taper}}^A = T_{\text{taper}}^B = 60$, $T_{\text{taper}}^C = 14$. For all setups, we also inject random $\text{Unif}(-0.25, 0.25)$ noise to the observations to accommodate random disturbances from the underlying dynamics.

For a given individual, we measure the average (over T_{taper}) dose suboptimality of any sequence $(u_1, \ldots, u_{T_{\text{taper}}})$ with respect to the true maximally greedy dose given by (4.2) which uses the full model and true underlying natural progression. That is, we can compare against the best possible control scheme had we known the true signal y_{t+1}^{nat} in advance. We additionally measure the average (over T_{taper}) constraint violation. We then average these two metrics over the N units to obtain a sense of the 'performance' of a protocol over a population.

Tapering protocol specification. As baselines, we consider the linear ($u_t = u_0 - \alpha t$) and the exponential ($u_t = \alpha^t u_0$) dose decay protocols and vary the decay rate α to get the trade-off curve between our two performance metrics for each opponent process when evaluated over units with

varied constraints. We also tested the cold turkey protocol ($u_t \equiv 0$), but its performance was too poor to be included.

We compare against the integral controller of Section 5 for which we can directly input the patient constraint $y_{\min}$. The main obstacle is that we do not know the exact value of $g(0)$. However, since this is the immediate response, we can assume access to a course range. For our evaluation, we take the lower and upper bounds to be 50% and 150% of the true response. This directly corresponds to conservative settings for K_+ and K_- equal to $2g(0)^{-1}/3$ and $2g(0)^{-1}$. In the Appendix, we provide an ablation over the coarseness of our range specification, observing that our method can handle wide ranges. As discussed in Section 5, our method can be instantiated with a padded $y^{\delta}_{\min} = y_{\min} + \delta$ to provide stronger constraint violation guarantees. Since this is the one hyperparameter that might affect practice, we also vary it in our experiments.

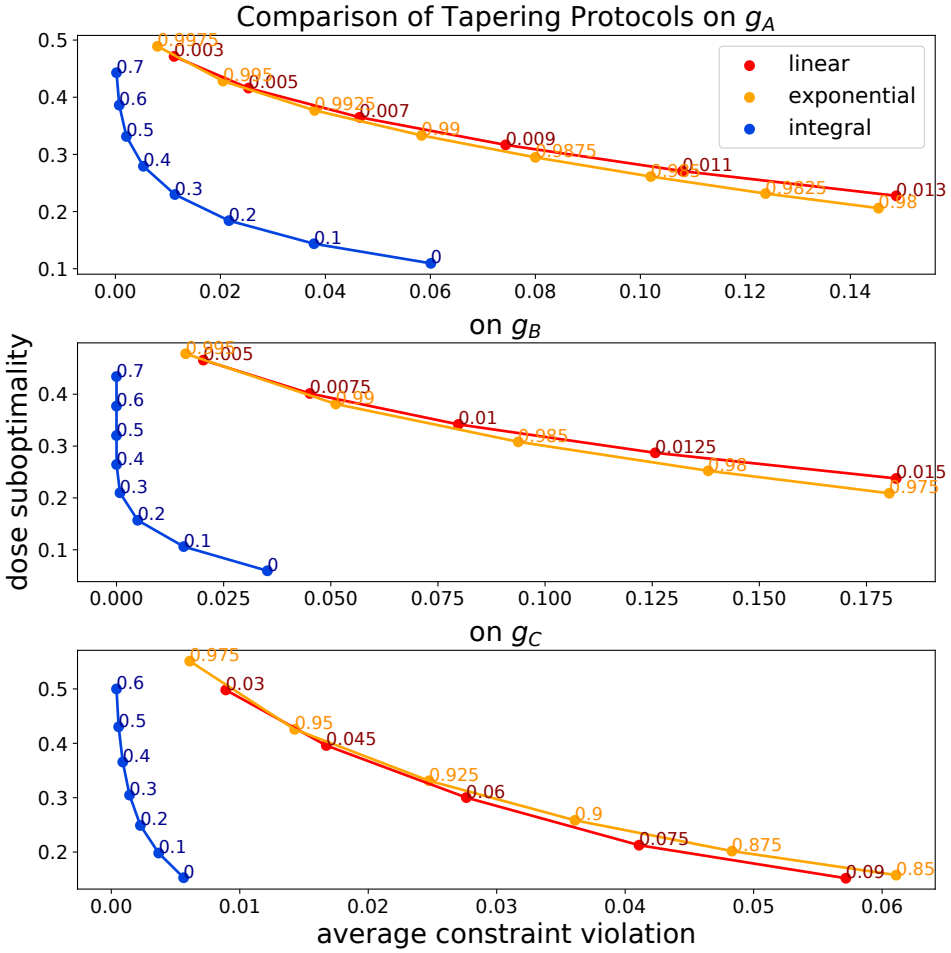

Figure 4: Average cumulative constraint violation (x-axis) against average dose suboptimality (y-axis). Each point is averaged over $N = 100$ units.

Results. In Figure 4, we see that the integral control law dominates the baselines considered, as the curves obtained by sweeping over the various parameters do not intersect. In the Appendix, we provide further experiments covering the robustness of the method to even more extreme mis-

specification of $g(0)$, as well as behavior in terms of a potential additional metric of interest: the percentage of units that successfully taper within the specified time frame.

7 Conclusion & Discussion

There remain several control-theoretic considerations in tapering protocol design. For example, the desirable guarantee of monotonic dose decrease was induced by a discretization such that the full positive effect occurs within one timestep. How can we find the appropriate time window without excessive exploration? Can we develop methods for estimating the switching point of the process and discretizing such that the induced discrete system has $\tau_0 = 1$?

Additionally, numerous practical, clinical, and ethical considerations must be considered for the downstream deployment of the proposed methods. For example, neither the optimal policy nor our default integral control policy decrease the dose monotonically. For some individuals, this could result in recommending increasing dosages, which is undesirable in most clinical settings. A translation to practice may necessarily add the constraint that a dose never exceeds some level. And this may mean that desired levels of well-being are not achievable for some patients.

To our knowledge, this work is the first formalization of tapering as an optimal control problem. This formalism leads to a concrete approach that departs from existing literature on tapering and prescribes adaptivity. Excitingly, the proposed protocols may be able to address some of the limitations of current approaches while still providing simple, explainable rules. We hope that the straightforward protocols are amenable to undergoing human subject validation in future studies.

References

[1] Alicia Agnoli, Guibo Xing, Daniel J Tancredi, Elizabeth Magnan, Anthony Jerant, and Joshua J Fenton. Association of dose tapering with overdose or mental health crisis among patients prescribed long-term opioids. *Jama*, 326(5):411–419, 2021.

[2] Chantal Berna, Ronald J Kulich, and James P Rathmell. Tapering long-term opioid therapy in chronic noncancer pain: evidence and recommendations for everyday practice. In *Mayo clinic proceedings*, volume 90,6, pages 828–842. Elsevier, 2015.

[3] Joshua J Fenton, Alicia L Agnoli, Guibo Xing, Lillian Hang, Aylin E Altan, Daniel J Tancredi, Anthony Jerant, and Elizabeth Magnan. Trends and rapidity of dose tapering among patients prescribed long-term opioid therapy, 2008-2017. *JAMA network open*, 2(11):e1916271–e1916271, 2019.

[4] Leonard Handelsman, Kenneth J Cochrane, Marvin J Aronson, Robert Ness, Karen J Rubinstein, and Philip D Kanof. Two new rating scales for opiate withdrawal. *The American journal of drug and alcohol abuse*, 13(3):293–308, 1987.

[5] Stephen G Henry, Debora A Paterniti, Bo Feng, Ana-Maria Iosif, Richard L Kravitz, Gary Weinberg, Penney Cowan, and Susan Verba. Patients' experience with opioid tapering: a conceptual model with recommendations for clinicians. *The Journal of Pain*, 20(2):181–191, 2019.

[6] Mark Abie Horowitz, Sameer Jauhar, Sridhar Natesan, Robin M Murray, and David Taylor. A method for tapering antipsychotic treatment that may minimize the risk of relapse. *Schizophrenia Bulletin*, 47(4):1116–1129, 2021.

[7] Mark Abie Horowitz, Robin M Murray, and David Taylor. Tapering antipsychotic treatment. *JAMA psychiatry*, 78(2):125–126, 2021.

[8] Harold Kalant. Biological models of alcohol tolerance and physical dependence. *Alcohol Intoxication and Withdrawal I: Experimental Studies*, pages 3–14, 1973.

[9] George F Koob. The dark side of emotion: the addiction perspective. *European journal of pharmacology*, 753:73–87, 2015.

[10] George F Koob and Michel Le Moal. Drug addiction, dysregulation of reward, and allostasis. *Neuropsychopharmacology*, 24(2):97–129, 2001.

[11] Kurt Kroenke and Robert L Spitzer. The phq-9: a new depression diagnostic and severity measure, 2002.

[12] Mehrdad Mahdavi, Rong Jin, and Tianbao Yang. Trading regret for efficiency: online convex optimization with long term constraints, 2012.

[13] Shie Mannor and John N Tsitsiklis. Online learning with constraints. *Lecture notes in computer science*, 4005:529, 2006.

[14] Iñigo Mujika and Sabino Padilla. Scientific bases for precompetition tapering strategies. *Medicine & Science in Sports & Exercise*, 35(7):1182–1187, 2003.

[15] Jerrold F Rosenbaum, Maurizio Fava, Sharon L Hoog, Richard C Ascroft, and William B Krebs. Selective serotonin reuptake inhibitor discontinuation syndrome: a randomized clinical trial. *Biological psychiatry*, 44(2):77–87, 1998.

[16] Richard L Solomon. The opponent-process theory of acquired motivation: the costs of pleasure and the benefits of pain. *American psychologist*, 35(8):691, 1980.

[17] Donald R Wesson and Walter Ling. The clinical opiate withdrawal scale (cows). *Journal of psychoactive drugs*, 35(2):253–259, 2003.

A Proof of Lemma 5.2

We first derive an alternate expression of the (possibly negative) dose $u_t^\star$ that leads to $y_{\min}$. Concretely, using the same derivation in Propositon 4.3 of the maximally greedy protocol (4.2), we have that

$$u_t^\star = \frac{y_{\min} - y_{t+1}^{\text{nat}} - \sum_{k=1}^{t} g(k)u_{t-k}}{g(0)}$$

We aim to express this in terms of the current set-point error $y_t - y_{\min}$ and u_{t-1}, so we add and subtract y_t^{nat} and $\sum_{k=1}^{t-1} u_{t-k-1}$ (which generate $y_t - g(0)u_{t-1}$) to obtain:

$$\begin{aligned} u_t^\star = g(0)^{-1}\big(&y_{\min} - y_t^{\text{nat}} + (y_t^{\text{nat}} - y_{t+1}^{\text{nat}}) \\ &- \sum_{k=1}^{t-1} g(k)u_{t-k-1} - \sum_{k=1}^{t} g(k)(u_{t-k} - u_{t-k-1})\big) \end{aligned}$$

where we take $u_{-1} = 0$. Since $y_t^{\text{nat}} + \sum_{k=1}^{t-1} g(k)u_{t-k-1} = y_t - g(0)u_{t-1}$, we can rewrite the above as

$$\begin{aligned} u_t^\star = u_{t-1} &- g(0)^{-1}(y_t - y_{\min}) - g(0)^{-1}(y_{t+1}^{\text{nat}} - y_t^{\text{nat}}) \\ &+ g(0)^{-1}\left(\sum_{k=1}^{t} g(k)(u_{t-k-1} - u_{t-k})\right) \end{aligned}$$

Since $K_+ \le g(0)^{-1}$ and $K_- \ge g(0)^{-1}$, we have that

$$-g(0)^{-1}(y_t - y_{\min}) \le -K_+(y_t - y_{\min})_+ - K_-(y_t - y_{\min})_-$$

And therefore, if u_t is given by (5.1), then

$$\begin{aligned} u_t^\star \le u_t &- g(0)^{-1}(y_{t+1}^{\text{nat}} - y_t^{\text{nat}}) \\ &+ g(0)^{-1}\left(\sum_{k=1}^{t} g(k)(u_{t-k-1} - u_{t-k})\right) \end{aligned}$$

Multiplying both sides by $g(0)$ and then adding $\sum_{k=1}^{t} g(k)u_{t-k} + y_{t+1}^{\text{nat}}$, we have

$$\begin{aligned} &y_{t+1}^{\text{nat}} + g(0)u_t^\star + \sum_{k=1}^{t} g(k)u_{t-k} \\ \le &y_{t+1}^{\text{nat}} + \sum_{k=0}^{t} g(k)u_{t-k} - (y_{t+1}^{\text{nat}} - y_t^{\text{nat}}) \\ &+ \sum_{k=1}^{t} g(k)(u_{t-k-1} - u_{t-k}) \end{aligned}$$

Since by construction the LHS is equal to exactly $y_{\min}$ and by the dynamics equation $y_{t+1}^{\text{nat}} + \sum_{k=0}^{t} g(k)u_{t-k} = y_t$, rearranging the above yields:

$$y_{\min} + (y_{t+1}^{\text{nat}} - y_t^{\text{nat}}) - \sum_{k=1}^{t} g(k)(u_{t-k-1} - u_{t-k}) \le y_t$$

which is precisely the stated inequality.

B Generalization to non-linear opponent processes

Our definition of (3.1) implies a linear contribution of the dose to the observed well-being metric. One may wonder if we can accommodate the case where the opponent process varies with u, i.e. y_t evolves according to the following generalized version of (3.1):

$$y_{t+1} = \sum_{k=0}^{t} g(k, u_{t-k}) + y_{t+1}^{\text{nat}} \tag{B.1}$$

In this subsection, we show we can, under a corresponding generalization of well-behavedness from Definition 3.2. The rate of change of each $g(t, \cdot)$ with respect to the dose u will play an important role, so we introduce the following two shorthand notations:

$$\partial_u^-(t) \doteq \inf_u \partial_u g(t, u) \qquad \partial_u^+(t) \doteq \sup_u \partial_u g(t, u)$$

Definition B.1 (Generalized Opponent Process)**.** *We say that g is a generalized opponent process if there exists a time τ_0 such that:*

1. $\inf_u g(\tau, u) \geq 0$ *and* $\partial_u^-(t) \geq 0$ *when* $\tau < \tau_0$,
2. $\sup_u g(\tau, u) \leq 0$ *and* $\partial_u^+(t) \leq 0$ *when* $\tau \geq \tau_0$.

Definition B.2 (Generalized LPOP)**.** *A generalized opponent process g is a generalized linearly progressing opponent process (G-LPOP) if there exists an $\alpha \in (0, 1)$ such that:*

1. $\partial_u^+(t+1) \leq \alpha \cdot \partial_u^-(t)$ *when* $t < \tau_0 - 1$,
2. $\partial_u^-(t+1) \geq \alpha \cdot \partial_u^+(t)$ *when* $t \geq \tau_0$.

In this case the proof of Theorem 4.2 changes a bit, relying on Mean Value Theorem. We give the statement and proof in Theorem B.3 below.

Theorem B.3. *For any G-LPOP g, taking the maximally greedy dose at every time maintains $y_t \geq y_{\min}$ for all t with minimal cumulative dose.*

Proof. Consider a dosing sequence $u_t^\star$ that maintains $y_t^\star \geq y_{\min}$ for all t with minimal cumulative dose. Assume it is not maximally greedy (otherwise we would be done). This means that there exists t_0 and $\epsilon > 0$ such that $u_{t_0}^\star - \epsilon$ would produce a new $y'_{t_0+1} \geq y_{\min}$. Consider the modified dosing schedule

$$u'_t = \begin{cases} u_{t_0}^\star - \epsilon, t = t_0 \\ u_{t_0+1}^\star + \alpha\epsilon, t = t_0 + 1 \\ u_t^\star \text{ otherwise} \end{cases}$$

and let y'_t be the resulting observations. Since $u'_t = u_t^\star$ for $t < t_0$, we have that $y'_t = y_t^\star \geq y_{\min}$ up to t_0. Plugging in the effect of the dose modifications at t_0 and $t_0 + 1$ into the generalized dynamics equation (B.1), by the Mean Value Theorem, we get:

$$\begin{aligned} y'_{t+1} &= y_{t+1}^\star - g(t - t_0, u_{t_0}^\star) + g(t - t_0, u_{t_0}^\star - \epsilon) \\ &\quad - g(t - t_0 - 1, u_{t_0+1}^\star) + g(t - t_0 - 1, u_{t_0+1}^\star + \alpha\epsilon) \\ &= y_{t+1}^\star - \epsilon \cdot \partial_u g(t - t_0, u_{t_0}^\star - \delta_0\epsilon) \\ &\quad + \alpha\epsilon \cdot \partial_u g(t - t_0 - 1, u_{t_0+1}^\star + \delta_1\epsilon) \end{aligned}$$

for some $\delta_0 \in (0,1), \delta_1 \in (0,\alpha)$. Since g is a G-LPOP, this final expression is greater than or equal to $y^\star_{t+1}$ and hence is also greater than or equal to $y_{\min}$. Since this holds for all $t > t_0$, we have that y'_t satisfies the constraints for all t. To finalize, note that y'_t is achieved with smaller cumulative dose since

$$\sum_{t=1}^{T} u'_t = \sum_{t=1}^{T} u^\star_t - \epsilon + \alpha\epsilon < \sum_{t=1}^{T} u^\star_t,$$

This is a contradiction, completing the proof. □

Instantiating this in (4.1) can be done by solving for the dose which exactly matches the setpoint:

Proposition B.4. *For any G-LPOP g, assume we know a lower bound $y_t^{\text{nat,lb}}$ such that $y_t^{\text{nat,lb}} \leq y_t^{\text{nat}}$ for all t. Then, for any instance of y_t^{nat} that is greater than the prescribed lower bound, the dosing sequence which takes u_t be the solution to*

$$g(0,u) = \max\left\{0, y_{\min} - y_{t+1}^{\text{nat,lb}} - \sum_{k=1}^{t} g(k, u_{t-k})\right\} \tag{B.2}$$

maintains $y_t \geq y_{\min}$ for all t with minimal cumulative dose.

Since $g(0,u)$ is increasing in u, we have the following guarantee:

Remark B.1. *An ϵ-accurate solution to (B.2) can be computed in $\mathcal{O}(\log \frac{1}{\epsilon})$ steps.*

The above effectively shows that optimal tapering is equivalent to setpoint matching under non-linear g as well. While it is not as clean to derive precise guarantees for the integral controller as in the main text, the result of Theorem B.3, combined with the robustness to $g(0)$ errors observed experimentally, strongly motivates the same approach in this more general setting.

C Additional Experiments

Ablation over $g(0)$ range coarseness. Figure 5 (Left) demonstrates that the integral controller can support a wide misspecification on the range of $g(0)$. In fact, the experiments suggest an overly conservative upper bound on $g(0)$ may improve performance under an optimal choice of δ.

Number of subjects fully tapered. Figure 5 (Right) plots the percentage of units the proposed protocol fully tapered within T_{taper} steps for varying values of δ. We see that our method fully tapers a portion of the units within some pre-specified timeline with minimal constraint violation. This is in contrast with fixed methods which induce a 0% vs. 100% tapering success step function given a timeframe T_{taper}. In all settings used, non-adaptive methods would have to incur significant constraint violations to fully taper within the pre-specified timeframe.

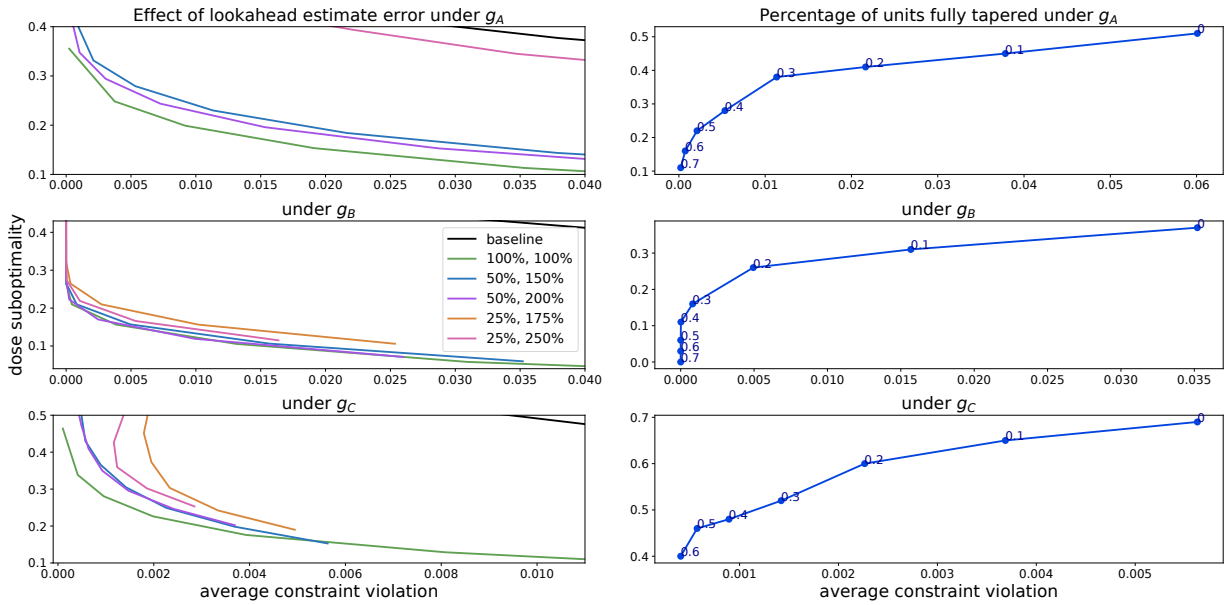

Figure 5: **Left:** Average cumulative constraint violation (x-axis) against average dose suboptimality (y-axis) for different ranges of lower and upper bounds on $g(0)$. A label of (p_1, p_2) in the legend corresponds to setting K_- to $(p_1 g(0))^{-1}$ and K_+ to $(p_2 g(0))^{-1}$. **Right:** Average cumulative constraint violation (x-axis) against the percentage of units fully tapered (y-axis).