Control Under the Wrong Model Can Be Better Than Under the Correct One

Francesco Damiani[*]
Center for Brain and Cognition,
Department of Engineering
Pompeu Fabra University
Barcelona, ES
francesco.damiani@upf.edu

Rubén Moreno-Bote[*†]
Center for Brain and Cognition,
Department of Engineering,
Serra Húnter Fellow Programme
Pompeu Fabra University
Barcelona, ES
ruben.moreno@upf.edu

Abstract

Building an optimal controller requires solving two coupled problems: (1) computing beliefs about the state by filtering observations, and (2) designing a control signal based on those beliefs to minimize a cost function. Ideally, these two problems can be solved independently: first, a filter is computed to estimate the state; then, a controller is built on top of that estimate. When exact inference is tractable, this approach is optimal and applies to the prominent Linear-Quadratic-Gaussian stochastic control problem. However, perfect inference is generally intractable, and filtering and control are closely intertwined problems. The question, then, is whether the commonly used first-filter-then-control approach remains optimal, and whether optimal control strategies should generally rely on internal representations that mirror the dynamics of the external world. We show that this does not need to be the case when internal computations are themselves subject to noise that cannot be independently observed or removed through additional filtering: even in a simple setting with linear dynamics, multiplicative noise, and internal noise, the optimal linear controller can have internal forward dynamics that differ from those of the external state. Instead, it relies on internal representations that mix estimation and control, with this mismatch becoming more pronounced as internal noise increases. Our work challenges standard approaches that prioritize external world modeling over control.

1 Introduction

Control theory has emerged as a fundamental tool to understand sensorimotor loops in neuroscience [1, 2, 3, 4] and for designing goal-directed behavior in robotics [5, 6, 7, 8, 9, 10]. In this framework, the state dynamics (world) evolve over time, and a controller (the agent) receives observations of the world and executes online control signals (actions) so as to minimize a cost function of the control and state variables [11]. When the state is only partially observable, building a good controller implies solving two problems [11, 12]: (1) a filtering problem, in which past observations and actions are used to construct an internal representation of the state, and (2) a control problem, in which actions are selected on the basis of this internal representation.

Under unlimited computational resources and full knowledge of the state dynamics, the filtering and control problems can be formulated as a two-step procedure. First, beliefs over states are computed from past observations and control signals. Then, a controller that executes actions based on those

[*]These authors contributed equally.
[†]Corresponding author.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

beliefs is optimized to minimize a cost function. This *first-filter-then-control* approach has strongly influenced control theory ever since the idea that a "good regulator" is a "good model" of the world was proposed [13, 14], and it is widely applied in representation learning for control [15, 16, 17]. This view is supported, at different levels, by two classical results. First, the theory of partially observable Markov decision processes (POMDPs) shows that, under full knowledge of the transition and observation models, the optimal policy can be written as a function of the belief state alone [12]. Second, in the Linear-Quadratic-Gaussian (LQG) problem [18] – characterized by linear state dynamics, quadratic costs, and additive Gaussian process and observation noise – this belief-based formulation takes the stronger form of the classical *separation principle* [18], whereby the optimal filter and controller can be derived independently. In particular, in the LQG problem, linear state dynamics and additive Gaussian noise allow exact inference through a state estimator that is a linear filter of past observations and control signals. Thus, the filtering problem can be solved exactly. Because the state estimate provides the sufficient statistics in this setting, the optimal controller can be written as a function of this estimate alone. For quadratic costs, this controller is linear. Despite its simplicity, LQG control has been successfully and extensively applied in neuroscience [19, 20, 21, 22] and has inspired approximate control methods in robotics [7, 6, 5, 23].

Although under limited resources exact inference becomes intractable in all but the simplest problems, the success and widespread use of LQG control [19, 6, 7, 22] have encouraged a common approach for solving more general control problems: first build an approximate model of the world through approximate inference, and then construct a controller that acts on the resulting *approximate* state estimates [1, 24, 25, 15, 16, 17, 26]. For instance, in a generalization of the LQG problem that incorporates multiplicative observation and control noise [24], exact inference over the state variables becomes intractable, yielding non-Gaussian posterior beliefs even when the underlying noise variables are Gaussian. Following the *first-filter-then-control* approach, this problem is generally addressed by assuming that the state-estimate dynamics are linear, as in the world-state dynamics, and, more importantly, that the forward (prediction) dynamics of the state estimate match the forward dynamics of the world state [24, 25]. In other words, the internal dynamics are assumed to represent and match the world dynamics. Emphasizing the problem of building a model over the control problem is also common in some prominent theories of brain function [27], where the priority is to construct a good representation of the world, while control plays a secondary role.

An alternative approach to address stochastic optimal control under intractable inference is to recognize that filtering and control generally become deeply intertwined components of the same computational process, which we refer to as *control-and-filter-at-once*. Moreover, there is no general reason to assume that building an accurate model of the world should take precedence over the control problem. In addition, this perspective risks underestimating the role of internal representations, which are central to many motor control studies [1, 28, 2, 3, 29, 30]. Beyond these classical formulations, a broader neuroscience literature suggests that internal models need not faithfully mirror the external dynamics. Frameworks such as optimal feedback control and forward-model learning emphasize that internal dynamics may be simplified, biased, or task-dependent [28, 31, 32, 33]. Empirical work further suggests that brain activity can reflect internally generated dynamics optimized for control or prediction rather than a veridical copy of the physical plant [34, 35]. These ideas align with the conceptual motivation behind a *control-and-filter-at-once* approach, where the internal model is optimized jointly with the control policy rather than constrained to follow the true system dynamics. Allowing internal models to differ from the laws governing the external world extends the flexibility of the stochastic optimal control framework, opening the door to a richer class of biologically plausible computations. In addition, this flexibility may lead to improved solutions in terms of cost minimization, particularly when internal representations are themselves affected by noise [36, 37, 38].

In this paper, we ask whether the *first-filter-then-control* approach or the *control-and-filter-at-once* approach provides a better formulation for stochastic optimal control. We consider the stochastic optimal control problem first introduced in [24], and later extensively applied to the study of sensorimotor control [39, 22, 40, 41, 42, 43, 44]. This formulation includes multiplicative observation and control noise, as well as internal noise in the agent's internal process. We refer to this setting as the Linear-Quadratic-Multiplicative-Internal (LQMI) problem and extend it to allow for arbitrary forward internal dynamics. The LQMI problem is well suited for our purposes because it is complex enough to make exact inference intractable, due to multiplicative observation and control noise, yet structured enough to admit an analytically derived, convergent solution method under standard

linearity assumptions, as we show in this paper. From a theoretical standpoint, the LQMI problem is also important because the separation principle no longer applies [24, 25]: filtering and control become intertwined problems that must be co-optimized. However, whether further assumptions in the standard model, such as constraining the forward internal dynamics to match the forward world dynamics, are optimal remains unknown.

Going beyond additive noise models is essential for incorporating empirically observed multiplicative noise in both motor execution [45, 46, 47, 24] and sensory feedback [48, 49, 24]. These sources of variability are key ingredients for explaining human behavior from a rational control perspective [24, 50, 47, 51, 46]; in robotics, analogous noise formulations are used to account for uncertainty in sensors and actuators [7, 6]. Critically, the LQMI problem also includes internal noise in the agent's internal processes, capturing possible inaccuracies in internal inference. This inclusion is important because it allows us to modulate the level of approximate inference performed by the internal model. Empirical evidence for such limitations includes persistent uncertainty despite repeated sensory observations, increasing behavioral variability during memory delays, neural fluctuations predictive of subsequent behavioral errors, substantial contributions of internal noise to perceptual decision-making, and endogenous fluctuations in neural activity [24, 52, 53, 54, 55, 38, 56]. Rather than reproducing the precise biophysical mechanisms underlying neural variability, the LQMI framework captures these limitations through an effective internal-noise term, representing computational fluctuations that are not independently observable or removable by an additional noiseless filtering mechanism. We also extend the framework to account for more biologically motivated forms of internal variability, such as multiplicative internal noise.

In this paper, we first define the LQMI stochastic optimal control problem with free forward internal dynamics, in which the internal dynamics are no longer constrained to match the forward dynamics of the world state (Sec. 2). We then derive a convergent algorithm to solve this problem (Sec. 3). Notably, no approximations are introduced beyond the standard assumption of linear internal dynamics and linear control [24, 25]. Our main result is that, under internal noise, optimizing the forward internal dynamics without constraining them to match the forward world dynamics leads to lower costs than enforcing matched dynamics; in this sense, *the wrong linear model can be the best control model* (Sec. 4). This advantage becomes stronger as the level of internal noise increases. Because the exact belief dynamics are intractable in this setting, by a *correct* model we refer to the standard inference-oriented approximation in which the forward internal dynamics are constrained to match the world dynamics. Our results have direct implications for approaches that prioritize inference over control [27]; by contrast, our theory is more aligned with theories of brain function that posit control as a central computational component [57, 58, 59].

2 From LQG to LQMI with Free Internal Dynamics

In this section, we first review the classical LQG formulation and the LQMI model, before extending LQMI to a more general setting with freely optimized internal dynamics and generalized additive and multiplicative noise sources.

The LQG Model In the standard LQG formulation (Fig. 1, left panel), given a stochastic linear system that evolves according to

$$x_{t+1} = Ax_t + Bu_t + \xi_t \ , \tag{1}$$

an optimal agent produces a causal control policy $u_t(y_{t-1}, \dots, y_0)$ based on the noisy observations received along the trial,

$$y_t = Hx_t + \omega_t \ , \tag{2}$$

with the goal of minimizing the expected cumulative quadratic cost

$$C = \sum_{t=0}^{T} \mathbb{E}\left[x_t^\top Q_t x_t + u_t^\top R_t u_t\right] \ . \tag{3}$$

Here, $x_t \in \mathbb{R}^m$, $y_t \in \mathbb{R}^k$, and $u_t \in \mathbb{R}^p$, for $t = 0, 1, \dots, T$. The matrix $H \in \mathbb{R}^{k\times m}$ maps the state into the observation space, while $A \in \mathbb{R}^{m\times m}$ and $B \in \mathbb{R}^{m\times p}$ define the linear evolution of the state and the effect of control inputs, respectively. The state dynamics are affected by zero-mean additive Gaussian noise $\xi_t \in \mathbb{R}^m$ with covariance Σ_ξ, while sensory feedback is corrupted by zero-mean

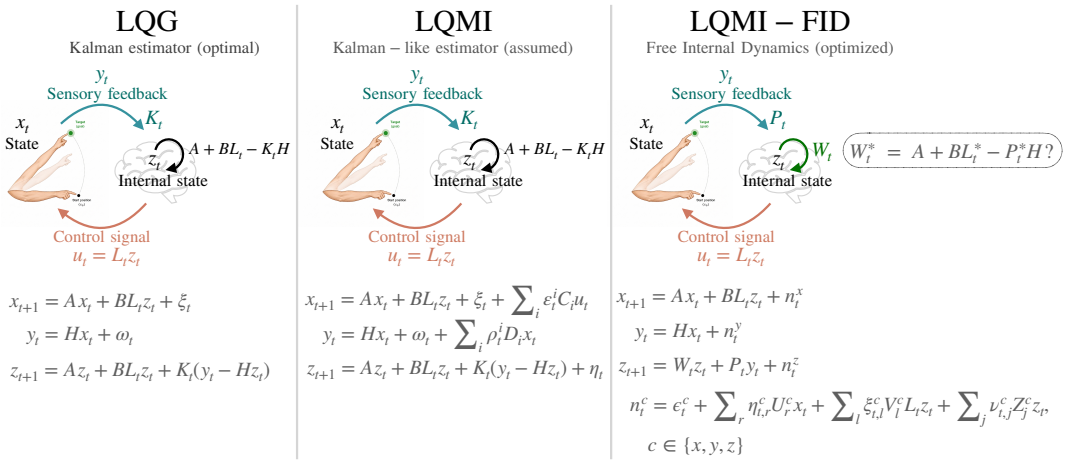


Figure 1: *From LQG to LQMI-FID.* Comparison of the three stochastic control formulations, all defined by the same quadratic cost function in Eq. 3. LQG features additive process and sensory noise with an optimal Kalman-like estimate z_t. LQMI assumes the same recursive estimator structure but adds multiplicative motor and sensory noise, together with internal noise. LQMI-FID relaxes this estimator constraint by optimizing W_t, P_t, and L_t, allowing z_t to become a task-optimized internal control variable; it also generalizes the noise model by allowing additive, state-dependent, and control-dependent multiplicative noise in the external, observation, and internal dynamics, as well as internal state-dependent multiplicative noise in the internal dynamics. A central question is whether the optimal W_t, denoted W_t^*, matches the world forward dynamics and retains the control and Kalman-like dependencies imposed by LQMI.

additive Gaussian noise $\omega_t \in \mathbb{R}^k$ with covariance Σ_ω. The cost combines two contributions: an effort term, weighted by the symmetric positive semidefinite matrix $R_t \in \mathbb{R}^{p\times p}$, $R_t \geq 0$, which discourages large control inputs, and a state-dependent term, weighted by $Q_t \in \mathbb{R}^{m\times m}$, $Q_t \geq 0$, which penalizes departures from task-relevant states such as targets or prescribed trajectories. The expectation in Eq. 3 is taken over the stochastic initial condition and all noise realizations. For the LQG problem, the optimal solution can be obtained in closed form [18]. It consists of two components: a Kalman filter, which recursively constructs the optimal estimate $\hat{x}_t \equiv z_t$ of the state x_t,

$$z_{t+1} = Az_t + Bu_t + K_t(y_t - Hz_t)\,, \tag{4}$$

where $K_t \in \mathbb{R}^{m\times k}$, and a linear feedback law, $u_t = L_t z_t$, with $L_t \in \mathbb{R}^{p\times m}$, that maps this estimate into the control input. A key property of this setting is that the estimation and control problems decouple: the filter (K_t) and control (L_t) gains can be derived separately, a result known as the separation principle [18].

The LQMI Model To account for key noise features missing in the original LQG formulation, [24] introduced multiplicative control- and sensory-dependent noise, together with internal noise affecting the internal inference process (Fig. 1, center). These extensions are empirically motivated by variability observed in motor commands [45, 46, 47], sensory feedback [48, 49], and internal neural processing [55, 38, 56, 3]. Then, the system and observation dynamics in Eqs. 1,2 become

$$x_{t+1} = Ax_t + Bu_t + \xi_t + \sum\nolimits_i \varepsilon_t^i C_i u_t \tag{5}$$

$$y_t = Hx_t + \omega_t + \sum\nolimits_i \rho_t^i D_i x_t\,, \tag{6}$$

where $C_i \in \mathbb{R}^{m\times p}$ and $D_i \in \mathbb{R}^{k\times m}$ specify the directions and amplitudes through which each multiplicative noise source enters the motor and sensory dynamics, respectively. The random variables $\varepsilon_t \in \mathbb{R}^c$ and $\rho_t \in \mathbb{R}^d$ are zero-mean vectors normalized to have identity covariance, so that $\Sigma_\varepsilon = \mathbb{I}_{c\times c}$ and $\Sigma_\rho = \mathbb{I}_{d\times d}$. Under the dynamics in Eqs.5,6, optimizing the causal control law $u_t(y_{t-1}, \ldots, y_0)$ with respect to the cost in Eq. 3 becomes a highly nontrivial problem, for which no general solution is known. In particular, no sufficient statistic is known that would allow for a Kalman filter-like recursion. Therefore, exact inference about the state is generally intractable,

as is typically the case in nonlinear or non-Gaussian filtering problems [60]. In the spirit of the *first-filter-then-control* approach, the control signal u_t is then assumed [24, 25] to be the linear readout of a state estimate z_t, namely $u_t = L_t z_t$, with $L_t \in \mathbb{R}^{p\times m}$. Importantly, the estimate is assumed to follow a linear Kalman-filter-like recursion with *matched* forward dynamics (i.e., same matrices A and B as in Eq. 5),

$$z_{t+1} = Az_t + Bu_t + K_t(y_t - Hz_t) + \eta_t \ , \quad u_t = L_t z_t \ . \tag{7}$$

Here, we use the notation of Eq. 4, with the extra term $\eta_t \in \mathbb{R}^m$ representing additive internal noise of zero mean and covariance Σ_η.

This formulation defines the Linear-Quadratic-Multiplicative-Internal (LQMI) problem: multiplicative and internal noise are added to the sensorimotor loop, while estimation and control are restricted to a Kalman-filter-like state estimate and a linear feedback policy. The objective is to find the sequence of control $L_{0,\dots,T}$ and filter $K_{0,\dots,T}$ gains that minimize the quadratic cost in Eq. 3.

The LQMI-FID Model We now consider a more general control problem, where the internal state is no longer constrained to obey a Kalman filter-like recursion. Rather, the internal dynamics are treated as optimization variables and are allowed to differ from the true forward dynamics of the external state. Thus, we define the *Free Internal Dynamics* (FID) framework by extending the LQMI problem introduced above, and refer to the resulting formulation as LQMI-FID. This framework allows the internal forward dynamics to be freely optimized and generalizes the multiplicative noise structure, so that noise in both the external and internal dynamics may depend on either the state variable or the internal variable (Fig. 1, right). The resulting control problem is therefore defined by the following dynamics

$$\begin{aligned}
x_{t+1} &= \underbrace{Ax_t}_{\text{world dynamics}} + \underbrace{BL_t z_t}_{\text{control input}} + \underbrace{n_t^x}_{\text{process noise}} \ , \quad y_t = \underbrace{Hx_t}_{\text{observation model}} + \underbrace{n_t^y}_{\text{sensory noise}} \ , \\
z_{t+1} &= \underbrace{W_t z_t}_{\text{internal dynamics}} + \underbrace{P_t y_t}_{\text{sensory feedback}} + \underbrace{n_t^z}_{\text{internal noise}} \\
\underbrace{n_t^c}_{\text{noise model}} &= \underbrace{\epsilon_t^c}_{\text{additive}} + \underbrace{\sum\nolimits_r \eta_{t,r}^c U_r^c x_t}_{\text{state-dependent}} + \underbrace{\sum\nolimits_l \xi_{t,l}^c V_l^c L_t z_t}_{\text{control-dependent}} + \underbrace{\sum\nolimits_j \nu_{t,j}^c Z_j^c z_t}_{\text{internal state-dependent}} \ , \quad c \in \{x, y, z\} \ ,
\end{aligned} \tag{8}$$

where notation follows Eqs. 5-7, with appropriate matrix dimensions and noises with covariances $\mathbb{E}[\epsilon_t^c(\epsilon_{t'}^{c'})^\top] = \Sigma_{\epsilon^c}\delta_{cc'}\delta_{tt'}$, and i.i.d., zero-mean, scalar noises $\eta_{t,r}^c$, $\xi_{t,l}^c$ and $\nu_{t,j}^z$ with unit variance. We introduce additive and multiplicative noises n_t^c in the dynamics, observation and internal dynamics z_t. Sums over r, l and j can be c-dependent. We adopt the notation $\epsilon_t^x = \xi_t$, $\epsilon_t^y = \omega_t$, and $\epsilon_t^z = \eta_t$, where ξ_t, ω_t, and η_t denote the additive noise terms introduced in Eqs. 5-7. We restrict internal state-dependent multiplicative noise to the internal dynamics by setting $Z_j^x = Z_j^y = 0$, while allowing Z_j^z to be nonzero. The matrices $U_r^c \in \mathbb{R}^{d_c\times m}$, $V_l^c \in \mathbb{R}^{d_c\times p}$ and $Z_j^c \in \mathbb{R}^{d_c\times n}$, with $d_x = m$, $d_y = k$ and $d_z = n$, determine the directions and amplitudes of the state-dependent, control-dependent and internal state-dependent multiplicative noise, respectively. $P_t \in \mathbb{R}^{n\times k}$ is a pseudo-filter matrix that maps the observation y_t into the dynamics of the internal variable $z_t \in \mathbb{R}^n$. The internal variable then evolves according to a linear system with time-dependent forward dynamics $W_t \in \mathbb{R}^{n\times n}$. The control signal $u_t = L_t z_t$ is p-dimensional, with $L_t \in \mathbb{R}^{p\times n}$.

The key novelty is that the internal dynamics matrix W_t is allowed to differ freely from the matrices A and B that define the dynamics of the external world state. Thus, in the LQMI-FID framework, the internal variable z_t can combine estimation- and control-related computations. This differs from standard LQMI, where z_t is constrained to act as a state estimate (see Fig. 1 and Eq. 7). Since W_t is not required to match the external dynamics, z_t can evolve according to its own optimized dynamics and encode internal variables tailored to control rather than to state estimation alone.

Internal-Activity Regularization Biological and engineered controllers have finite internal dynamic ranges, and large internal signals may be costly, much as large control signals are penalized in Eq. 3. In biological systems, these constraints are consistent with the saturating nature of neuronal activity and the metabolic costs of neural signaling [61, 62]. Because LQMI-FID allows the internal dynamics W_t to be freely optimized, its internal representations are less constrained than in standard LQMI, including in their overall scale. In particular, compensating changes in the sensory-input gain P_t and control readout L_t can amplify internal activity while reducing the effective influence of additive

internal noise on control. Although similar activity constraints could also be introduced in standard LQMI, they are particularly relevant in the more flexible LQMI-FID framework. To account for this additional flexibility and the associated costs of large internal activity, we also consider the extended cost function

$$C_\lambda = C + \lambda \sum_{t=0}^{T} \mathrm{Tr}(S_t^{zz}), \qquad S_t^{zz} = \mathbb{E}[z_t z_t^\top], \quad \lambda \geq 0, \tag{9}$$

where C is the original task cost of Eq. 3. Since S_t^{zz} is a non-centered second moment, this term penalizes both mean internal activity and its fluctuations, discouraging unbounded amplification; $\lambda = 0$ recovers the original objective. Another possible mechanism limiting the performance benefits of arbitrarily large internal representations is multiplicative noise acting directly on the internal state, which is already included in the generalized noise model of Eq. 8 when $Z_j^z \neq 0$. Unlike additive internal noise, this multiplicative noise cannot be attenuated by compensatory rescaling of the internal representation and control readout. This mechanism is also biologically motivated by evidence that neuronal variability depends on activity levels, including multiplicative fluctuations in neuronal gain [55, 63].

Overall, the LQMI-FID problem consists in jointly optimizing the time-dependent forward dynamics $W_{0,\dots,T}$, pseudo-filter matrices $P_{0,\dots,T}$, and control matrices $L_{0,\dots,T}$ to minimize the original cost in Eq. 3, or its regularized extension in Eq. 9, given the initial distribution $p_0(x, z)$. To our knowledge, no previous algorithm has addressed the joint optimization of these three quantities. We next derive a coordinate-descent algorithm that applies to both the original and regularized formulations.

3 Solving the LQMI-FID Problem

Solving the full LQMI-FID problem defined above requires going beyond existing approaches. Gradient-based methods such as [25] become inefficient in this setting, because the number of optimized parameters grows substantially once the internal dynamics W_t are also treated as free parameters. In addition, the algorithm of [24] is built around a Kalman-like recursive formulation of the internal state and relies on an unbiased-estimator assumption, $\mathbb{E}[x_t \mid z_t] = z_t$, which we explicitly relax. We therefore derive a coordinate-descent algorithm that directly optimizes the control gains L_t, pseudo-filter matrices P_t, and internal dynamics W_t under the LQMI-FID formulation.

The full derivation is provided in Appendix A.1. In summary, starting from an initial set of parameters $W_{0,\dots,T}$, $P_{0,\dots,T}$, and $L_{0,\dots,T}$, we can update a single matrix W_t so as to obtain a non-increasing value of the cost, while keeping all other matrices W_τ with $\tau \neq t$, as well as all $P_{0,\dots,T}$ and $L_{0,\dots,T}$, fixed. Sweeping this update over all time steps $t = 0, \dots, T$ defines a full W-pass. Analogously, we define P- and L-passes by sequentially optimizing the pseudo-filter matrices P_t and the control gains L_t, respectively. By alternating W-, P-, and L-passes from arbitrary initial values $W^{(0)}$, $P^{(0)}$, and $L^{(0)}$, we obtain a sequence of parameter updates for which the cost in Eq. 9

$$\begin{aligned} &C_\lambda(W^{(0)}, P^{(0)}, L^{(0)}) \geq C_\lambda(W^{(1)}, P^{(0)}, L^{(0)}) \geq C_\lambda(W^{(1)}, P^{(1)}, L^{(0)}) \geq \\ &\geq C_\lambda(W^{(1)}, P^{(1)}, L^{(1)}) \geq \dots \geq C_\lambda(W^{(n+1)}, P^{(n)}, L^{(n)}) \geq C_\lambda(W^{(n+1)}, P^{(n+1)}, L^{(n)}) \geq \\ &\geq C_\lambda(W^{(n+1)}, P^{(n+1)}, L^{(n+1)}) \geq \dots \geq C_{\lambda,\min} \geq 0 \end{aligned} \tag{10}$$

is monotonically non-increasing. Since the sequence is non-negative, it converges to a limiting cost that is no larger than the initial one. We denote the corresponding limiting parameters by $W^* = W^{(\infty)}$, $P^* = P^{(\infty)}$, and $L^* = L^{(\infty)}$. The algorithm pseudocode is presented in Appendix A.1.2 (Algorithm 1), while in Appendix A.1 we prove the following

Theorem 1. *Starting with arbitrary $W^{(0)}$, $P^{(0)}$ and $L^{(0)}$ and distribution of initial conditions $p_0(x, z)$, the coordinate descent algorithm defined by iterating in alternation W-, P- and L-passes (Algorithm 1) converges to an improved triplet of forward dynamics, pseudo-filter and control matrices W^*, P^* and L^*. The improved triplet corresponds to a critical point of the cost function in Eq. 9.*

The algorithm can be straightforwardly adapted to cases in which some of the matrices W_t, P_t, or L_t are fixed: one simply omits the corresponding updates from the coordinate-descent passes, while preserving the same monotonicity and convergence guarantees. The classical LQMI model can be recovered as a constrained subproblem of this framework. In that case, the internal dynamics are not free, but are tied to the external dynamics through $W_t = A + BL_t - K_tH$, with $P_t = K_t$.

Thus, the LQMI problem can be solved by applying an analogous coordinate-descent scheme over the constrained variables L_t and K_t, rather than over the free matrices W_t, P_t, and L_t (see Appendix A.1.1, Algorithm 2). Consistent with this theoretical relationship, we numerically verify that, in the absence of internal noise, the solution to the LQMI-FID problem achieves the same expected task cost as the solution to the LQMI problem (see Appendix A.4.2).

4 Experiments

Hereafter, we refer to the optimal control approach derived from the classical LQMI formulation, in which the internal dynamics are constrained to match the world dynamics, as *model-matched* control, and we refer to the optimal control derived from LQMI-FID, in which the internal dynamics are freely optimized, as *model-free* control. Here, *model-free* refers specifically to the absence of constraints matching the internal forward dynamics to those of the external world, rather than to the absence of an internal dynamical model. We apply these two approaches to two motor-control tasks, comparing both their performance and the computational strategies they implement.

Model-Free Control Improves Performance by Departing from External-World Dynamics We first consider a three-dimensional reaching task in which the hand is represented as a point mass moving in 3D space, a simplified multidimensional analogue of the reaching paradigms previously used to study stochastic optimal control under sensorimotor noise [24, 25]. The system state $x_t \in \mathbb{R}^6$ contains the hand position and velocity. The goal is to drive the hand from an initial displaced position to the target by the end of the movement, while minimizing control effort. The task provides a first nontrivial benchmark for testing how different control architectures cope with additive, multiplicative, and internal noise. In this setting, state, observation, control, and internal state dimensions are matched, $m = n = p = k = 6$ (see details in Appendix A.3.1). We set $\lambda = 10^{-4}$ for model-free control and $\lambda = 0$ for model-matched control. This conservative comparison tests whether the flexibility of the model-free internal dynamics improves task performance even when the model-matched controller is optimized solely to minimize the original task cost, without any internal-activity penalty.

We first verify that the algorithm for optimizing the model-free controller converges reliably across internal-noise levels (Fig. 2a). Despite the additional regularization, model-free control consistently achieves a lower task cost than model-matched control in the presence of internal noise (Fig. 2b). Importantly, this improvement does not rely on amplifying internal state activity. For example, at $\sigma_\eta = 0.13$, the time-averaged activity, measured by $\langle \mathrm{Tr}(S_t^{zz}) \rangle_t$ after excluding initial and final transients, is approximately 0.2 for model-free compared with 24 for model-matched control (i.e., 2 orders of magnitude lower in model-free), despite the lower task cost achieved by the former.

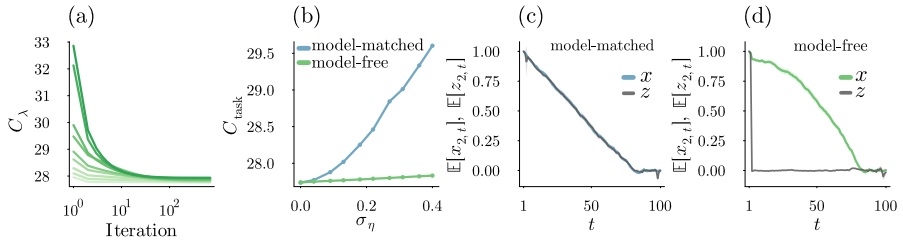


Figure 2: *Model-Free Control Outperforms Model Matching Despite Internal-Activity Regularization.* **(a)** Optimization traces for regularized model-free control across internal-noise levels σ_η ranging from 0 to 0.4, showing the convergence of the regularized objective C_λ. Different shades indicate increasing internal-noise levels, from lighter to darker green. **(b)** Expected cumulative task cost C_{task}, given by Eq. 3 and thus excluding the internal-activity penalty, as a function of σ_η for model-matched and model-free control. The model-free controller is optimized with an internal-activity penalty $\lambda = 10^{-4}$, while the model-matched controller is optimized without regularization ($\lambda = 0$). **(c,d)** Trial-averaged trajectories of the second physical-state component, $x_{2,t}$, and the corresponding internal variable, $z_{2,t}$, for the model-matched (c) and regularized model-free (d) approaches at the representative noise level $\sigma_\eta = 0.13$. Shaded bands denote SEM over 10000 Monte Carlo trials.

This improved performance can be understood by examining how model-free control optimizes its internal representations. Unlike in model-matched control, the internal recurrent dynamics W_t are not constrained to follow the Kalman-like recursion associated with the external physical system. Indeed, while the internal variable z_t in model-matched control approximately tracks the corresponding external-state component (Fig. 2c), model-free control develops internal representations that need not encode the external state coordinate by coordinate (Fig. 2d). To quantify this departure, we compare the optimized internal dynamics W_t with the model-matched reference $\widetilde{W}_t = A + BL_t - P_tH$, where P_t plays the role of the filter gain K_t in Eq. 7. If $W_t = \widetilde{W}_t$, the model-free internal dynamics reduce to the classical model-matched recursion, in which z_t follows a Kalman-like state-estimation structure. Deviations from $\widetilde{W}_t$ therefore quantify the extent to which model-free control departs from this constraint. We find that the time-averaged deviation $\langle |W_t - \widetilde{W}_t| \rangle_t$, normalized by the time-averaged norm of $\widetilde{W}_t$, exceeds 97% across all tested positive internal-noise levels. Thus, in the presence of internal noise, model-free control achieves more effective control while allowing internal representations to depart from the forward dynamics of the external physical system.

We further investigate these findings in the more structured redundant-arm control task presented next. In a subsequent analysis of both tasks, we show that the advantage of model-free control cannot be explained by a simple change of coordinates, but involves a more substantial reorganization of internal representations and control strategies. Finally, in Appendix A.4.1, we examine the unregularized solutions, showing how model-free control can exploit internal state rescaling to attenuate additive internal noise. The present results demonstrate that its performance advantage persists even when this scaling mechanism is penalized.

Beyond State Estimation: Task-Optimized Internal Representations To investigate whether these findings extend to more complex control settings, we next consider a redundant three-joint arm stabilizing around a reference posture. As in the reaching task, the state is six-dimensional, now comprising joint angles and angular velocities. However, control is mediated by nine muscle-like channels that are mapped linearly onto the three joint torques (see Appendix A.3.2 for details). This overactuation allows different muscle-level control patterns to produce equivalent joint-level commands. Such redundancy, ubiquitous in biological musculoskeletal systems [64], makes the task a useful benchmark for investigating how different control architectures organize high-dimensional motor commands in the presence of noise. Sensory feedback is three-dimensional, with only joint angles directly observed. Thus, the state and internal state dimensions are $m = n = 6$, the control dimension is $p = 9$, and the observation dimension is $k = 3$. The task cost primarily penalizes deviations of the joint angles from the reference posture, while angular velocities are only weakly penalized.

Our results above showed that internal representations not explicitly constrained to estimate the external state can support better task performance in the presence of internal noise (Fig. 2). This challenges the classical stochastic optimal control framework, in which control policies are typically based on optimal state estimates [18, 19, 24]. We now ask whether, even within model-matched control, improving state estimation necessarily leads to better control. To test this, we introduce a variant of model-matched control, termed Separated Estimation–Control (SEC), in which filter and control gains are optimized with respect to different objectives, while the internal dynamics remain constrained to follow the Kalman-like recursion in Eq. 7. Control gains are optimized to minimize the original task cost in Eq. 3, whereas filter gains are optimized to minimize the mean-squared estimation error $C_{\text{est}} = \sum_{t=0}^{T} \mathbb{E}\left[(x_t - z_t)^\top (x_t - z_t)\right]$. This defines an iterative optimization procedure in which control and filter gains are updated according to two distinct objectives (see details in Appendix A.2). We refer to this approach as SEC because the internal representation is explicitly optimized as a world state estimator before being used for control.

In the presence of internal noise, we find that better state estimation does not necessarily imply better control. In this task, SEC produces internal states with lower mean-squared estimation error than those of model-matched control, with the improvement increasing as σ_η becomes larger (Fig. 3a). However, despite this improvement in estimation accuracy, SEC yields higher task costs than model-matched control (Fig. 3b). Although modest, this difference provides an example in which improving state-estimation accuracy does not translate into better task performance, even when internal representations are constrained to follow a Kalman-like recursion.

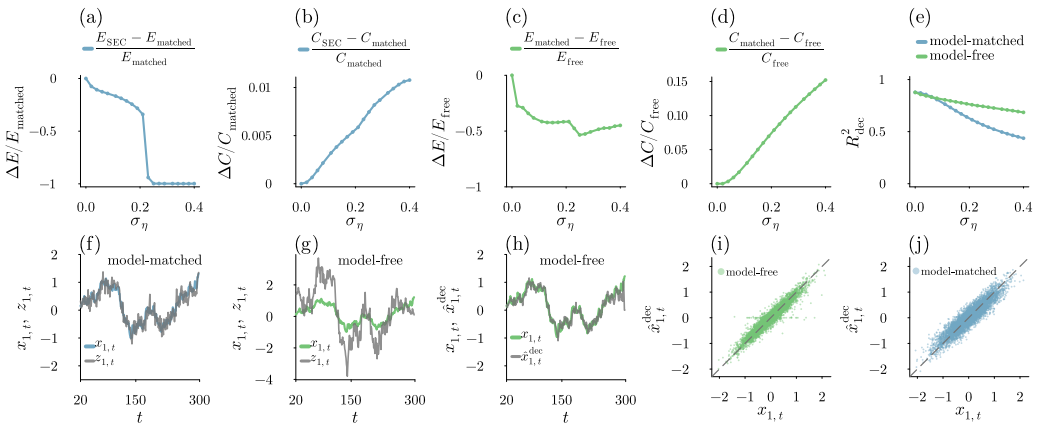


Figure 3: *Model-Free Control Improves Performance While Preserving Decodable State Information.* **(a,b)** Comparison of Separated Estimation–Control (SEC) and model-matched control in the redundant-arm task, without internal-activity regularization ($\lambda = 0$). Relative differences in (a) full-state estimation error E and (b) task cost C as functions of internal-noise magnitude σ_η. Estimation error is computed as $E = \sqrt{C_{\text{est}}}/T$, where $C_{\text{est}} = \sum_{t=0}^{T} \mathbb{E}[(x_t - z_t)^\top (x_t - z_t)]$ over the analysis window. **(c,d)** Same analysis comparing model-matched and model-free control, both optimized with internal-activity regularization ($\lambda = 10^{-3}$). Model-free control achieves lower task cost despite larger estimation error. Performance is evaluated using the original task cost, excluding the internal-activity penalty. **(e)** Linear decodability of the full physical state x_t from the internal state z_t, quantified by R^2_{dec}. **(f,g)** Representative single-trial trajectories of the first physical-state component and the corresponding internal state component for model-matched and model-free control, respectively ($\sigma_\eta = 0.13$). **(h)** Model-free trajectory of the same physical-state component and its linear reconstruction $\hat{x}^{\text{dec}}_{1,t}$ from the full internal state. **(i,j)** Decoded versus true values of the first state component for model-free and model-matched control, respectively ($\sigma_\eta = 0.20$), sampled across trials and time. Decoding uses a time-dependent affine map in all three panels (see Appendix A.4.4). Dashed lines indicate identity.

This distinction becomes even more evident in model-free control. When both model-matched and model-free controllers are optimized with the same internal-activity penalty ($\lambda = 10^{-3}$), the latter achieves lower task costs (Fig. 3d), consistent with the reaching-task results (Fig. 2b), despite exhibiting larger mean-squared estimation errors (Fig. 3c). However, this does not mean that the model-free internal representations abandon relevant world information. Such a strategy would be inconsistent with the basic requirements of closed-loop control. Indeed, the internal state z_t maintains high linear decodability of the full physical state x_t across internal-noise levels, even exceeding that of model-matched control (Fig. 3e). Thus, although z_t in model-free control does not directly track x_t coordinate by coordinate, as illustrated by the representative trajectories (Figs. 3f–g), it still contains information that allows accurate linear reconstruction of the external state (Figs. 3h–j). Together, these findings show that robust control under internal noise does not require more accurate explicit estimates of the physical state, but can instead rely on task-optimized internal representations that preserve decodable state information.

Further analyses of state decodability and robustness mechanisms are provided in Appendices A.4.5 and A.4.6. The model-free advantage persists in generalization tests across task objectives, reward magnitudes, target locations, and unexpected internal-noise changes (Appendix A.4.11). We also present examples of distinguishable behavioral signatures (Appendices A.4.8 and A.4.9).

Optimized Internal Dynamics Enable More Effective Control To further investigate whether the advantage of model-free control reflects a functional reorganization of internal computations rather than a simple change of coordinates, we performed two constrained optimizations at $\sigma_\eta = 0.4$, without internal-activity regularization ($\lambda = 0$). In the first, we optimized the recurrent dynamics W_t and control gain L_t while keeping the sensory gain P_t fixed at its model-matched value. In the second, more restrictive case, we optimized W_t alone, keeping both P_t and L_t fixed at their model-matched values. In the redundant-arm task, optimizing W_t and L_t reduced task cost by 13.5% relative to the model-matched solution, while optimizing W_t alone still achieved a 7.9% improvement. Similar

results were obtained in the three-dimensional reaching task, with improvements of 2.8% and 1%, respectively. Importantly, the improvements obtained by optimizing W_t alone show that relaxing the model-matched recurrence can improve control even without modifying the sensory-input or control gains. Together with the regularized results (Figs. 2 and 3) and the additional analyses in Appendix A.4.6, these findings further support the role of task-optimized internal dynamics in achieving robust control under internal noise. This advantage also persists in a purely linear-Gaussian setting with internal noise (Appendix A.4.3). Model-free control also outperforms previous optimization methods in an additional benchmark (Appendix A.4.7). Finally, we tested the robustness of model-free control to multiplicative internal noise, a biologically plausible form of variability that cannot be attenuated through internal state amplification (Appendix A.4.10). Model-free control continued to outperform model-matched control, with relative task-cost improvements of 2.1%–16.6% in the three-dimensional reaching task and approximately 1%–9.8% in the redundant-arm task.

A Broader Class of Problems for Stochastic Optimal Control By allowing internal dynamics to be optimized independently of external dynamics, our framework extends classical stochastic control to settings such as the control of one recurrent neural population by another, where A and W describe the connectivity of two distinct networks (Appendix A.4.12). Such settings are not directly accommodated by classical model-matched formulations [19, 24].

5 Conclusions

A central approach in stochastic optimal control theory, and more generally in control problems, has long taken the view that, in order to act on the world, an agent must first build accurate beliefs about the external state to be controlled and then construct a control signal based on those beliefs [13, 19]. However, this *first-filter-then-control* approach is feasible only for problems with simplified noise models, where exact inference is tractable, such as the classical Linear-Quadratic-Gaussian (LQG) problem [18]. It is therefore unclear whether, in general, an optimal agent should prioritize accurate world modeling over control, or whether more flexible internal representations can improve control performance. In addition, a vast literature in neuroscience has long emphasized the role of internal models in sensorimotor control, going beyond inference alone [28, 31].

In this work, we have shown that under more complex noise models, the *first-filter-then-control* approach fails to yield optimal behavior. That is, representations built to favor accurate world modeling do not necessarily lead to optimal behavior. We have shown this by extending stochastic optimal control theory to include flexible internal representations. More specifically, we extended the noise model of [24], here referred to as LQMI, to allow for free internal dynamics that are no longer constrained to follow a Kalman-like recursion, but are instead optimized jointly with the control and filter gains. In this way, we defined a novel control model, which we refer to as LQMI-FID, and solved it through an analytically derived and convergent algorithm.

Our results demonstrate that, in multidimensional motor-control tasks with internal noise, having free internal representations outperforms classical formulations in which internal states are explicitly constrained to act as state estimates. This reveals novel computational strategies in sensorimotor control. Moreover, by allowing for flexible internal dynamics, our approach makes stochastic optimal control applicable to a wider range of optimal-control problems, including the modeling of closed-loop interactions between two linear neural populations. Overall, our work shows, from a stochastic optimal control perspective, that prioritizing accurate world modeling over control does not, in general, lead to optimal control behavior, challenging a central approach in classical control theories.

Limitations and Future Work We chose the LQMI framework because it is sufficiently complex to render exact inference intractable while remaining structured enough to admit an analytically derived, convergent solution method. Our finding that model-free control can outperform model-matched control, even with an incorrect internal model in this relatively simple setting, suggests that the performance gap could be larger in more complex problems, motivating further extensions of LQMI. Following previous work [24, 25], we assumed linear dynamics, linear control laws, and quadratic costs. One possible extension is to embed our approach within iterative linearization schemes, such as iLQG and DDP [6, 8, 65, 66], to address nonlinear systems. Another direction is to allow the internal-state dimensionality to differ freely from that of the external state, a promising but unexplored possibility that could enable nonlinear control strategies through high-dimensional linear representations [67, 68].

Acknowledgments and Disclosure of Funding

This work was supported by the grant MICIU/AEI/10.13039/501100011033 funded by the Spanish Ministry of Science, Innovation and Universities, by the grants "FEDER A way of making Europe" (ref: PID2023-146524NB), PRE2021-097778 and "ESF+", and by ICREA ACADÈMIA (2022) funded by the Catalan Institution for Research and Advanced Studies. The authors declare no competing interests.

References

[1] Daniel M Wolpert, Zoubin Ghahramani, and Michael I Jordan. An internal model for sensorimotor integration. *Science*, 269(5232):1880–1882, 1995.

[2] Reza Shadmehr and John W Krakauer. A computational neuroanatomy for motor control. *Experimental brain research*, 185:359–381, 2008.

[3] David W Franklin and Daniel M Wolpert. Computational mechanisms of sensorimotor control. *Neuron*, 72(3):425–442, 2011.

[4] Emanuel Todorov. Optimality principles in sensorimotor control. *Nature neuroscience*, 7(9):907–915, 2004.

[5] Vu Anh Huynh and Nicholas Roy. iclqg: combining local and global optimization for control in information space. In *2009 IEEE International Conference on Robotics and Automation*, pages 2851–2858. IEEE, 2009.

[6] Weiwei Li and Emanuel Todorov. Iterative linearization methods for approximately optimal control and estimation of non-linear stochastic system. *International Journal of Control*, 80(9):1439–1453, 2007.

[7] Yuval Tassa, Tom Erez, and Emanuel Todorov. Synthesis and stabilization of complex behaviors through online trajectory optimization. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 4906–4913. IEEE, 2012.

[8] Yuval Tassa, Nicolas Mansard, and Emo Todorov. Control-limited differential dynamic programming. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1168–1175. IEEE, 2014.

[9] Sergey Levine, Chelsea Finn, Trevor Darrell, and Pieter Abbeel. End-to-end training of deep visuomotor policies. *Journal of Machine Learning Research*, 17(39):1–40, 2016.

[10] Stefan Schaal and Nicolas Schweighofer. Computational motor control in humans and robots. *Current opinion in neurobiology*, 15(6):675–682, 2005.

[11] Donald E Kirk. *Optimal control theory: an introduction*. Courier Corporation, 2004.

[12] Leslie Pack Kaelbling, Michael L Littman, and Anthony R Cassandra. Planning and acting in partially observable stochastic domains. *Artificial intelligence*, 101(1-2):99–134, 1998.

[13] Roger C Conant and W Ross Ashby. Every good regulator of a system must be a model of that system. *International journal of systems science*, 1(2):89–97, 1970.

[14] Nathaniel Virgo, Martin Biehl, Manuel Baltieri, and Matteo Capucci. A "good regulator theorem" for embodied agents. In *Artificial Life Conference Proceedings 37*, volume 2025, page 46. MIT Press, 2025.

[15] Denis Yarats, Rob Fergus, Alessandro Lazaric, and Lerrel Pinto. Reinforcement learning with prototypical representations. In *International Conference on Machine Learning*, pages 11920–11931. PMLR, 2021.

[16] Ramanan Sekar, Oleh Rybkin, Kostas Daniilidis, Pieter Abbeel, Danijar Hafner, and Deepak Pathak. Planning to explore via self-supervised world models. In *International conference on machine learning*, pages 8583–8592. PMLR, 2020.

[17] Jürgen Schmidhuber. Formal theory of creativity, fun, and intrinsic motivation (1990–2010). *IEEE transactions on autonomous mental development*, 2(3):230–247, 2010.

[18] Mark Davis. *Stochastic modelling and control*. Springer Science & Business Media, 2013.

[19] Emanuel Todorov and Michael I Jordan. Optimal feedback control as a theory of motor coordination. *Nature neuroscience*, 5(11):1226–1235, 2002.

[20] Gerald E Loeb, WS Levine, and Jiping He. Understanding sensorimotor feedback through optimal control. In *Cold Spring Harbor symposia on quantitative biology*, volume 55, pages 791–803. Cold Spring Harbor Laboratory Press, 1990.

[21] Arthur D Kuo. An optimal control model for analyzing human postural balance. *IEEE transactions on biomedical engineering*, 42(1):87–101, 1995.

[22] Dominik Straub and Constantin A Rothkopf. Putting perception into action with inverse optimal control for continuous psychophysics. *Elife*, 11:e76635, 2022.

[23] Jur Van Den Berg, Sachin Patil, and Ron Alterovitz. Motion planning under uncertainty using iterative local optimization in belief space. *The International Journal of Robotics Research*, 31(11):1263–1278, 2012.

[24] Emanuel Todorov. Stochastic optimal control and estimation methods adapted to the noise characteristics of the sensorimotor system. *Neural computation*, 17(5):1084–1108, 2005.

[25] Francesco Damiani, Akiyuki Anzai, Jan Drugowitsch, Gregory DeAngelis, and Rubén Moreno-Bote. Stochastic optimal control and estimation with multiplicative and internal noise. *Advances in Neural Information Processing Systems*, 37:123291–123327, 2024.

[26] Yilin Wu, Qian Zhang, and Zhiping Shen. Kalman filtering with multiplicative and additive noises. In *2016 12th World Congress on Intelligent Control and Automation (WCICA)*, pages 483–487. IEEE, 2016.

[27] Thomas Parr and Karl J Friston. Generalised free energy and active inference. *Biological cybernetics*, 113(5):495–513, 2019.

[28] Mitsuo Kawato. Internal models for motor control and trajectory planning. *Current opinion in neurobiology*, 9(6):718–727, 1999.

[29] Matthew Golub, Steven Chase, and Byron Yu. Learning an internal dynamics model from control demonstration. In *International Conference on Machine Learning*, pages 606–614. PMLR, 2013.

[30] Matthew D Golub, Byron M Yu, and Steven M Chase. Internal models for interpreting neural population activity during sensorimotor control. *Elife*, 4:e10015, 2015.

[31] Daniel M Wolpert and Zoubin Ghahramani. Computational principles of movement neuroscience. *Nature neuroscience*, 3(11):1212–1217, 2000.

[32] Reza Shadmehr and Henry H Holcomb. Neural correlates of motor memory consolidation. *Science*, 277(5327):821–825, 1997.

[33] Stephen H Scott. Optimal feedback control and the neural basis of volitional motor control. *Nature Reviews Neuroscience*, 5(7):532–545, 2004.

[34] Mark M Churchland, John P Cunningham, Matthew T Kaufman, Justin D Foster, Paul Nuyujukian, Stephen I Ryu, and Krishna V Shenoy. Neural population dynamics during reaching. *Nature*, 487(7405):51–56, 2012.

[35] Juan A Gallego, Matthew G Perich, Lee E Miller, and Sara A Solla. Neural manifolds for the control of movement. *Neuron*, 94(5):978–984, 2017.

[36] Omer Hazon, Victor H Minces, David P Tomàs, Surya Ganguli, Mark J Schnitzer, and Pablo E Jercog. Noise correlations in neural ensemble activity limit the accuracy of hippocampal spatial representations. *Nature communications*, 13(1):4276, 2022.

[37] Stefano Panzeri, Monica Moroni, Houman Safaai, and Christopher D Harvey. The structures and functions of correlations in neural population codes. *Nature Reviews Neuroscience*, 23(9):551–567, 2022.

[38] Rubén Moreno-Bote, Jeffrey Beck, Ingmar Kanitscheider, Xaq Pitkow, Peter Latham, and Alexandre Pouget. Information-limiting correlations. *Nature neuroscience*, 17(10):1410–1417, 2014.

[39] Matthias Schultheis, Dominik Straub, and Constantin A Rothkopf. Inverse optimal control adapted to the noise characteristics of the human sensorimotor system. *Advances in Neural Information Processing Systems*, 34:9429–9442, 2021.

[40] Jonathon W Sensinger and Strahinja Dosen. A review of sensory feedback in upper-limb prostheses from the perspective of human motor control. *Frontiers in neuroscience*, 14:345, 2020.

[41] Dan Liu and Emanuel Todorov. Evidence for the flexible sensorimotor strategies predicted by optimal feedback control. *Journal of Neuroscience*, 27(35):9354–9368, 2007.

[42] Jun Izawa, Tushar Rane, Opher Donchin, and Reza Shadmehr. Motor adaptation as a process of reoptimization. *Journal of Neuroscience*, 28(11):2883–2891, 2008.

[43] Tomohiko Takei, Stephen G Lomber, Douglas J Cook, and Stephen H Scott. Transient deactivation of dorsal premotor cortex or parietal area 5 impairs feedback control of the limb in macaques. *Current Biology*, 31(7):1476–1487, 2021.

[44] Maryam M Shanechi, Ziv M Williams, Gregory W Wornell, Rollin C Hu, Marissa Powers, and Emery N Brown. A real-time brain-machine interface combining motor target and trajectory intent using an optimal feedback control design. *PloS one*, 8(4):e59049, 2013.

[45] GG Sutton and K Sykes. The variation of hand tremor with force in healthy subjects. *The Journal of physiology*, 191(3):699–711, 1967.

[46] Richard A Schmidt, Howard Zelaznik, Brian Hawkins, James S Frank, and John T Quinn Jr. Motor-output variability: a theory for the accuracy of rapid motor acts. *Psychological review*, 86(5):415, 1979.

[47] Christopher M Harris and Daniel M Wolpert. Signal-dependent noise determines motor planning. *Nature*, 394(6695):780–784, 1998.

[48] Chistina A Burbeck and Yen Lee Yap. Two mechanisms for localization? evidence for separation-dependent and separation-independent processing of position information. *Vision research*, 30(5):739–750, 1990.

[49] David Whitaker and Keziah Latham. Disentangling the role of spatial scale, separation and eccentricity in weber's law for position. *Vision research*, 37(5):515–524, 1997.

[50] Tamar Flash and Neville Hogan. The coordination of arm movements: an experimentally confirmed mathematical model. *Journal of neuroscience*, 5(7):1688–1703, 1985.

[51] Emanuel Todorov. Cosine tuning minimizes motor errors. *Neural computation*, 14(6):1233–1260, 2002.

[52] Martin Lemay and Luc Proteau. A distance effect in a manual aiming task to remembered targets: a test of three hypotheses. *Experimental brain research*, 140(3):357–368, 2001.

[53] Klaus Wimmer, Duane Q Nykamp, Christos Constantinidis, and Albert Compte. Bump attractor dynamics in prefrontal cortex explains behavioral precision in spatial working memory. *Nature neuroscience*, 17(3):431–439, 2014.

[54] Jan Drugowitsch, Valentin Wyart, Anne-Dominique Devauchelle, and Etienne Koechlin. Computational precision of mental inference as critical source of human choice suboptimality. *Neuron*, 92(6):1398–1411, 2016.

[55] A Aldo Faisal, Luc PJ Selen, and Daniel M Wolpert. Noise in the nervous system. *Nature reviews neuroscience*, 9(4):292–303, 2008.

[56] Mark M Churchland, Afsheen Afshar, and Krishna V Shenoy. A central source of movement variability. *Neuron*, 52(6):1085–1096, 2006.

[57] Rubén Moreno-Bote and Jorge Ramirez-Ruiz. Empowerment, free energy principle and maximum occupancy principle compared. In *NeurIPS 2023 workshop: Information-Theoretic Principles in Cognitive Systems*, 2023.

[58] Jorge Ramirez-Ruiz, Dmytro Grytskyy, Chiara Mastrogiuseppe, Yamen Habib, and Ruben Moreno-Bote. Complex behavior from intrinsic motivation to occupy future action-state path space. *Nature Communications*, 15(1):6368, 2024.

[59] Alexander S Klyubin, Daniel Polani, and Chrystopher L Nehaniv. Empowerment: A universal agent-centric measure of control. In *2005 ieee congress on evolutionary computation*, volume 1, pages 128–135. IEEE, 2005.

[60] Christopher M Bishop and Nasser M Nasrabadi. *Pattern recognition and machine learning*, volume 4. Springer, 2006.

[61] Hugh R Wilson and Jack D Cowan. Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical journal*, 12(1):1–24, 1972.

[62] David Attwell and Simon B Laughlin. An energy budget for signaling in the grey matter of the brain. *Journal of Cerebral Blood Flow & Metabolism*, 21(10):1133–1145, 2001.

[63] Robbe LT Goris, J Anthony Movshon, and Eero P Simoncelli. Partitioning neuronal variability. *Nature neuroscience*, 17(6):858–865, 2014.

[64] Kenji Tahara, Suguru Arimoto, Masahiro Sekimoto, and Zhi-Wei Luo. On control of reaching movements for musculo-skeletal redundant arm model. *Applied Bionics and Biomechanics*, 6(1):11–26, 2009.

[65] Jur Van Den Berg, Sachin Patil, and Ron Alterovitz. Motion planning under uncertainty using differential dynamic programming in belief space. In *Robotics research: The 15th international symposium ISRR*, pages 473–490. Springer, 2016.

[66] L-Z Liao and Christine A Shoemaker. Convergence in unconstrained discrete-time differential dynamic programming. *IEEE Transactions on Automatic Control*, 36(6):692–706, 1991.

[67] Milan Korda and Igor Mezić. Linear predictors for nonlinear dynamical systems: Koopman operator meets model predictive control. *Automatica*, 93:149–160, 2018.

[68] Steven L Brunton, Bingni W Brunton, Joshua L Proctor, and J Nathan Kutz. Koopman invariant subspaces and finite linear representations of nonlinear dynamical systems for control. *PloS one*, 11(2):e0150171, 2016.

[69] Steven M Kay. Statistical signal processing: estimation theory. *Prentice Hall*, 1:Chapter–3, 1993.

[70] Friedrich Schuessler, Francesca Mastrogiuseppe, Srdjan Ostojic, and Omri Barak. Aligned and oblique dynamics in recurrent neural networks. *Elife*, 13:RP93060, 2024.

[71] Noam Razin, Yotam Alexander, Edo Cohen-Karlik, Raja Giryes, Amir Globerson, and Nadav Cohen. Implicit bias of policy gradient in linear quadratic control: Extrapolation to unseen initial states. *arXiv preprint arXiv:2402.07875*, 2024.

[72] Johannes Friedrich, Siavash Golkar, Shiva Farashahi, Alexander Genkin, Anirvan Sengupta, and Dmitri Chklovskii. Neural optimal feedback control with local learning rules. *Advances in Neural Information Processing Systems*, 34:16358–16370, 2021.

[73] Daniel J O'Shea, Lea Duncker, Werapong Goo, Xulu Sun, Saurabh Vyas, Eric M Trautmann, Ilka Diester, Charu Ramakrishnan, Karl Deisseroth, Maneesh Sahani, et al. Direct neural perturbations reveal a dynamical mechanism for robust computation. *bioRxiv*, pages 2022–12, 2022.

[74] Herbert Jaeger and Harald Haas. Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *science*, 304(5667):78–80, 2004.

[75] Wolfgang Maass, Thomas Natschläger, and Henry Markram. Real-time computing without stable states: A new framework for neural computation based on perturbations. *Neural computation*, 14(11):2531–2560, 2002.

[76] David Sussillo and Larry F Abbott. Generating coherent patterns of activity from chaotic neural networks. *Neuron*, 63(4):544–557, 2009.

[77] Ta-Chu Kao, Mahdieh S Sadabadi, and Guillaume Hennequin. Optimal anticipatory control as a theory of motor preparation: A thalamo-cortical circuit model. *Neuron*, 109(9):1567–1581, 2021.

[78] Laureline Logiaco, LF Abbott, and Sean Escola. Thalamic control of cortical dynamics in a model of flexible motor sequencing. *Cell reports*, 35(9), 2021.

[79] Haim Sompolinsky, Andrea Crisanti, and Hans-Jurgen Sommers. Chaos in random neural networks. *Physical review letters*, 61(3):259, 1988.

A Appendix

A.1 Solution to the LQMI with Free Forward Internal Dynamics

The LQMI-FID problem is defined by the equations

$$\begin{aligned} x_{t+1} &= Ax_t + BL_t z_t + n_t^x \ , \quad y_t = Hx_t + n_t^y \ , \quad z_{t+1} = W_t z_t + P_t y_t + n_t^z \\ n_t^c &= \epsilon_t^c + \sum\nolimits_r \eta_{t,r}^c U_r^c x_t + \sum\nolimits_l \xi_{t,l}^c V_l^c L_t z_t + \sum\nolimits_j \nu_{t,j}^c Z_j^c z_t \ , \quad c \in \{x, y, z\} \ , \end{aligned} \tag{11}$$

identical to Eqs. 8. We consider $Z_j^x = Z_j^y = 0$, while allowing only Z_j^z to be nonzero. The goal is to optimize the forward dynamics $W_t \in \mathbb{R}^{n\times n}$, pseudo-filter $P_t \in \mathbb{R}^{n\times k}$ and control $L_t \in \mathbb{R}^{p\times n}$ – where p is the dimensionality of the control signal $u_t = L_t z_t$ – matrices so as to minimize the expected cumulative quadratic cost

$$\begin{aligned} C_\lambda &= \sum_{t=0}^{T} \mathbb{E}\left[x_t^\top Q_t x_t + z_t^\top L_t^\top R_t L_t z_t\right] + \lambda \sum_{t=0}^{T} \mathrm{Tr}(S_t^{zz}) = \\ &= \sum_{t=0}^{T} \left(\mathrm{Tr}(Q_t S_t^{xx}) + \mathrm{Tr}(L_t^\top R_t L_t S_t^{zz}) + \lambda \,\mathrm{Tr}(S_t^{zz})\right) \ , \end{aligned} \tag{12}$$

with $S_t^{zz} = \mathbb{E}[z_t z_t^\top]$, $\lambda \geq 0$ and initial condition $p_0(x, z)$. The original objective in Eq.3 is recovered for $\lambda = 0$. Since the initial distribution $p_0(x, z)$ is fixed, the contribution $\lambda \,\mathrm{Tr}(S_0^{zz})$ is constant under optimization. The additional penalty preserves the quadratic structure of the problem and does not modify the dynamics of the second-order moments.

Eqs. 11 can be put in a more compact form as

$$\begin{aligned} x_{t+1} &= Ax_t + BL_t z_t + n_t^x \\ z_{t+1} &= W_t z_t + P_t H x_t + P_t n_t^y + n_t^z \\ n_t^c &= \epsilon_t^c + \sum\nolimits_r \eta_{t,r}^c U_r^c x_t + \sum\nolimits_l \xi_{t,l}^c V_l^c L_t z_t + \sum\nolimits_j \nu_{t,j}^c Z_j^c z_t \ , \quad c \in \{x, y, z\} \ , \end{aligned} \tag{13}$$

from which it is clearer that the system consists of two coupled linear dynamical systems with free parameters W_t, P_t and L_t chosen so as to minimize the cost. The sums $\sum_r$, $\sum_l$ and $\sum_j$ can run over different limits depending on the source c, but here we use the same symbol to avoid cluttered notation.

Note that the model-free framework is strictly more general than the model-matched one, because one is always free to choose $P_t = K_t$ and $W_t = A + BL_t - K_t H$, recovering the corresponding model-matched formulation under the same noise model. The original LQMI formulation in Eqs. 5-7 is recovered when $Z_j^z = 0$. The reverse mapping, from model-free control to the model-matched one, is in general not possible.

The 2nd-order moments, appearing in the cost 12, obey

$$\begin{aligned} S_{t+1}^{xx} &= AS_t^{xx}A^\top + BL_t S_t^{zz} L_t^\top B^\top + AS_t^{xz} L_t^\top B^\top + BL_t (S_t^{xz})^\top A^\top + \Sigma_t^x \\ S_{t+1}^{zz} &= P_t H S_t^{xx} H^\top P_t^\top + W_t S_t^{zz} W_t^\top + P_t H S_t^{xz} W_t^\top + W_t (S_t^{xz})^\top H^\top P_t^\top + P_t \Sigma_t^y P_t^\top + \Sigma_t^z \\ S_{t+1}^{xz} &= AS_t^{xx} H^\top P_t^\top + BL_t S_t^{zz} W_t^\top + AS_t^{xz} W_t^\top + BL_t (S_t^{xz})^\top H^\top P_t^\top \ , \end{aligned} \tag{14}$$

with $\Sigma_t^c = \Sigma_{\epsilon^c} + \sum_r U_r^c S_t^{xx} (U_r^c)^\top + \sum_l V_l^c L_t S_t^{zz} L_t^\top (V_l^c)^\top + \sum_j Z_j^c S_t^{zz} (Z_j^c)^\top$, $c \in \{x, y, z\}$. Even though the model-free formulation is more general than the model-matched one, defined in Eqs. 5,6,7, it is already apparent that the equations for the second moments are simpler, more compact and transparent. This will be a recurrent theme in all next derivations and equations, so we will not repeat this below.

Total Cost and Cost-to-Go Let us define the cost-to-go at time t starting from x and z as

$$\begin{aligned} C_{\lambda,t}(x, z) &= x^\top Q_t x + z^\top L_t^\top R_t L_t z + \lambda z^\top z \\ &\quad + \mathbb{E}\left[\sum_{\tau=t+1}^{T} \left(x_\tau^\top Q_\tau x_\tau + z_\tau^\top L_\tau^\top R_\tau L_\tau z_\tau + \lambda z_\tau^\top z_\tau\right) \middle| \, x_t = x, \ z_t = z\right] \end{aligned} \tag{15}$$

where the expectation is over the noises with initial conditions fixed at x and z at time t, and for specific P, L and W from time t onward. The cost-to-go obeys the Bellman equation

$$\begin{aligned} C_{\lambda,t}(x,z) = \; & \mathrm{Tr}(Q_t x x^\top) + \mathrm{Tr}\left[(L_t^\top R_t L_t + \lambda I_n) z z^\top\right] \\ & + \int dx' dz' \, C_{\lambda,t+1}(x',z') \, p_{x,t+1}(x'|x,z) p_{z,t+1}(z'|x,z). \end{aligned} \tag{16}$$

where I_n is the n-dimensional identity matrix and the transition probability densities $p_{x,t+1}(x'|x,z)$ and $p_{z,t+1}(z'|x,z)$ are the transition probability functions over x' and z' at time $t+1$ when starting from x and z at time t, as defined by Eqs. 11. Using backward induction, it can be shown that the cost-to-go can be written for all t ($t = 0, ..., T$) as

$$C_{\lambda,t}(x,z) = \mathrm{tr}(\Lambda_t x x^\top) + \mathrm{tr}(\Omega_t z z^\top) + \mathrm{tr}(\Gamma_t x z^\top) + \gamma_t \; , \tag{17}$$

with matrices $\Lambda_t \in \mathbb{R}^{m\times m}$, $\Omega_t \in \mathbb{R}^{n\times n}$, and $\Gamma_t \in \mathbb{R}^{n\times m}$ and scalar γ_t obeying equations

$$\begin{aligned} \Lambda_t = \; & Q_t + A^\top \Lambda_{t+1} A + H^\top P_t^\top \Omega_{t+1} P_t H + H^\top P_t^\top \Gamma_{t+1} A \\ & + \sum\nolimits_r (U_r^x)^\top \Lambda_{t+1} U_r^x + \sum\nolimits_r (U_r^y)^\top P_t^\top \Omega_{t+1} P_t U_r^y + \sum\nolimits_r (U_r^z)^\top \Omega_{t+1} U_r^z \; , \\ \Omega_t = \; & \lambda I_n + L_t^\top R_t L_t + L_t^\top B^\top \Lambda_{t+1} B L_t + W_t^\top \Omega_{t+1} W_t + W_t^\top \Gamma_{t+1} B L_t \\ & + \sum\nolimits_r L_t^\top (V_r^x)^\top \Lambda_{t+1} V_r^x L_t + \sum\nolimits_r L_t^\top (V_r^y)^\top P_t^\top \Omega_{t+1} P_t V_r^y L_t \\ & + \sum\nolimits_r L_t^\top (V_r^z)^\top \Omega_{t+1} V_r^z L_t + \sum_j (Z_j^z)^\top \Omega_{t+1} Z_j^z \; , \\ \Gamma_t = \; & L_t^\top B^\top (\Lambda_{t+1} + \Lambda_{t+1}^\top) A + W_t^\top (\Omega_{t+1} + \Omega_{t+1}^\top) P_t H + W_t^\top \Gamma_{t+1} A + L_t^\top B^\top \Gamma_{t+1}^\top P_t H \; , \\ \gamma_t = \; & \mathrm{Tr}(\Lambda_{t+1} \Sigma_{\epsilon^x}) + \mathrm{Tr}(P_t^\top \Omega_{t+1} P_t \Sigma_{\epsilon^y}) + \mathrm{Tr}(\Omega_{t+1} \Sigma_{\epsilon^z}) + \gamma_{t+1} \; , \end{aligned} \tag{18}$$

with boundary conditions $\Lambda_T = Q_T$, $\Omega_T = \lambda I_n + L_T^\top R_T L_T$, $\Gamma_T = 0$ and $\gamma_T = 0$ (in this way the boundary condition that $C_T(x,z) = \mathrm{Tr}(Q_T x x^\top) + \mathrm{Tr}(L_T^\top R_T L_T z z^\top) + \lambda z^\top z$ is satisfied). The internal-activity penalty contributes the additional term λI_n to the backward recursion for Ω_t, while leaving the remaining recursions unchanged in form. Consequently, the same analytical framework applies to both the original and regularized objectives.

We now define the averaged cost-to-go at time t as

$$C_{\lambda,t} \equiv \int dx dz p_t(x,z) C_{\lambda,t}(x,z) = \mathrm{tr}(\Lambda_t S_t^{xx}) + \mathrm{tr}(\Omega_t S_t^{zz}) + \mathrm{tr}(\Gamma_t S_t^{xz}) + \gamma_t \; , \tag{19}$$

where $p_t(x,z)$ is the joint probability density over x and z given initial condition $p_0(x,z)$ and W_τ, L_τ, and P_τ for $\tau < t$. We note that the total cost C_λ in Eq. 12 can be written as

$$C_\lambda = C_{\lambda,0} \equiv \int dx dz p_0(x,z) C_{\lambda,0}(x,z) = \mathrm{tr}(\Lambda_0 S_0^{xx}) + \mathrm{tr}(\Omega_0 S_0^{zz}) + \mathrm{tr}(\Gamma_0 S_0^{xz}) + \gamma_0 \; , \tag{20}$$

which can also be expressed as

$$C_\lambda = C_{\lambda,<t} + C_{\lambda,t} \tag{21}$$

with $C_{\lambda,<t} = \sum_{\tau=0}^{t-1} \left[\mathrm{Tr}(Q_\tau S_\tau^{xx}) + \mathrm{Tr}(L_\tau^\top R_\tau L_\tau S_\tau^{zz}) + \mathrm{Tr}(\lambda I_n S_\tau^{zz}) \right]$. It is important to note that Eq. 21 is valid for all t.

Algorithm We note from Eqs. 18 that the coefficients Λ_t, Ω_t, Γ_t and γ_t depend on W_τ, P_τ and L_τ only for $\tau \geq t$, while S_t^{ab}, $ab \in \{xx, zz, xz\}$, only depend on those matrices for $\tau < t$, as it can be seen from Eqs. 14. Therefore, choosing an arbitrary t, in Eq. 21 only the term $C_{\lambda,t}$ depends on W_t, and in that term, Eq. 19, only the coefficients Λ_t, Ω_t, Γ_t and γ_t can depend on W_t. In conclusion, starting with a set of $W_{0,...,T}$, $P_{0,...,T}$ and $L_{0,...,T}$, we can improve the value of W_t as

$$W_t^* = \arg\min_{W_t} C_\lambda = \arg\min_{W_t} C_{\lambda,t} \; , \tag{22}$$

while keeping the W_τ for $\tau \neq t$ and all $P_{0,...,T}$ and $L_{0,...,T}$ fixed. A global minimum exists because $C_{\lambda,t}$ is always non-negative. Using elementary matrix operations, we find that

$$W_t^* = -P_t H S_t^{xz} (S_t^{zz})^{-1} - (\Omega_{t+1} + \Omega_{t+1}^\top)^{-1} \Gamma_{t+1} \left(B L_t + A S_t^{xz} (S_t^{zz})^{-1} \right) \; . \tag{23}$$

After the optimization, we must have

$$C^*_\lambda = C_\lambda(W_0, ..., W^*_t, ..., W_T) \leq C_\lambda(W_0, ..., W_t, ..., W_T) \ , \tag{24}$$

so that the total cost is non-increasing. After optimizing W_t, using the new W^*_t, the cost can be written as

$$\begin{aligned} C^*_\lambda =& C_{\lambda,<t+1} + C^*_{\lambda,t+1} = \\ =& C_{\lambda,<t+1} + \mathrm{Tr}(\Lambda_{t+1} S^{xx,*}_{t+1}) + \mathrm{Tr}(\Omega_{t+1} S^{zz,*}_{t+1}) + \mathrm{Tr}(\Gamma_{t+1} S^{xz,*}_{t+1}) + \gamma_{t+1} \end{aligned} \tag{25}$$

where the coefficients at time $t+1$ do not need to be updated (as they do not depend on W^*_t), but where the $S^{ab,*}_{t+1}$ need to be updated using Eqs. 14 with the new W^*_t.

Redefining W^*_t as W_t and the $S^{ab,*}_{t+1}$ as S^{ab}_{t+1}, we can now proceed to optimize W_{t+1} using the same procedure as above (changing t to $t+1$) to minimize again the total cost $C_\lambda(W_0, ..., W_t, W^*_{t+1}, ..., W_T) \leq C_\lambda(W_0, ..., W_t, W_{t+1}, ..., W_T)$ fixing $P_{0,...,T}$, $L_{0,...,T}$ and all W_τ except for $\tau = t+1$. This procedure can be repeated consecutively from $t=0$ up to T.

After this forward pass, we would like to repeat the process for P_t and L_t instead of W_t. But before doing this, the values of the coefficients in Eqs. 18 have to be recomputed so that Eq. 20 is true again. The process of forward updating the W_t from $t=0$ up to time T and, after this, recomputing the coefficients using a backwards pass is called W-pass. Note that in this process, the moments have been already recomputed. Starting from $W^{(n)} = W^{(n)}_{0,...,T}$, $P^{(n)} = P^{(n)}_{0,...,T}$ and $L^{(n)} = L^{(n)}_{0,...,T}$, the W-pass leads to a new set of forward dynamics matrices $W^{(n+1)}$ such that the cost is non-increasing, $C_\lambda(W^{(n+1)}, P^{(n)}, L^{(n)}) \leq C_\lambda(W^{(n)}, P^{(n)}, L^{(n)})$. We define a P-pass as that consisting in exactly repeating the same procedure for the $P_{0,...,T}$ instead of the $W_{0,...,T}$ while keeping fixed $W_{0,...,T}$ and $L_{0,...,T}$, and using the expression (obtained after some calculations)

$$P^*_t = -\left[W_t (S^{xz}_t)^\top + (\Omega_{t+1} + \Omega^\top_{t+1})^{-1} \Gamma_{t+1} \left(A S^{xx}_t + B L_t (S^{xz}_t)^\top\right)\right] H^\top E^{-1}_t \ , \tag{26}$$

with $E_t = H S^{xx}_t H^\top + \sum_l U^y_l S^{xx}_t (U^y_l)^\top + \sum_r V^y_r L_t S^{zz}_t L^\top_t (V^y_r)^\top + \Sigma_{\epsilon^y}$. Starting from $W^{(n+1)} = W^{(n+1)}_{0,...,T}$, $P^{(n)} = P^{(n)}_{0,...,T}$ and $L^{(n)} = L^{(n)}_{0,...,T}$, the P-pass leads to a new set of pseudo-filter matrices $P^{(n+1)}$ such that the cost is non-increasing, $C_\lambda(W^{(n+1)}, P^{(n+1)}, L^{(n)}) \leq C_\lambda(W^{(n+1)}, P^{(n)}, L^{(n)})$. Finally, we define an L-pass as that consisting in following similar steps to the previous ones to sequentially update the $L_{0,...,T}$ while keeping fixed $W_{0,...,T}$ and $P_{0,...,T}$, and using the expression (after some calculations)

$$L^*_t = -F^{-1}_t B^\top \left\{\tilde{\Lambda}_{t+1} A S^{xz}_t (S^{zz}_t)^{-1} + \Gamma^\top_{t+1} \left[P_t H S^{xz}_t (S^{zz}_t)^{-1} + W_t\right]\right\} \ , \tag{27}$$

with $F_t = 2R_t + B^\top \tilde{\Lambda}_{t+1} B + \sum_l (V^x_l)^\top \tilde{\Lambda}_{t+1} V^x_l + \sum_l (V^y_l)^\top P^\top_t \tilde{\Omega}_{t+1} P_t V^y_l + \sum_l (V^z_l)^\top \tilde{\Omega}_{t+1} V^z_l$, where we have defined $\tilde{\Lambda}_t = \Lambda_t + \Lambda^\top_t$ and $\tilde{\Omega}_t = \Omega_t + \Omega^\top_t$. Starting from $W^{(n+1)} = W^{(n+1)}_{0,...,T}$, $P^{(n+1)} = P^{(n+1)}_{0,...,T}$ and $L^{(n)} = L^{(n)}_{0,...,T}$, the L-pass leads to a new set of control matrices $L^{(n+1)}$ such that the cost is non-increasing, $C_\lambda(W^{(n+1)}, P^{(n+1)}, L^{(n+1)}) \leq C_\lambda(W^{(n+1)}, P^{(n+1)}, L^{(n)})$.

Importantly, the activity penalty does not explicitly modify the stationarity equations for W_t, P_t, and L_t. Indeed, the term $\lambda \,\mathrm{Tr}(S^{zz}_t)$ is independent of the gain matrices at time t when the moments at that time are held fixed. The penalty therefore affects the optimal gains indirectly through the modified backward multipliers, without changing the algebraic form of the gain-update equations.

Now, alternating W-, P- and L-passes from some initial arbitrary values $W^{(0)}, P^{(0)}, L^{(0)}$ we find

$$\begin{aligned} &C_\lambda(W^{(0)}, P^{(0)}, L^{(0)}) \geq C_\lambda(W^{(1)}, P^{(0)}, L^{(0)}) \geq C_\lambda(W^{(1)}, P^{(1)}, L^{(0)}) \geq \\ &\geq C_\lambda(W^{(1)}, P^{(1)}, L^{(1)}) \geq ... \geq C_\lambda(W^{(n+1)}, P^{(n)}, L^{(n)}) \geq C_\lambda(W^{(n+1)}, P^{(n+1)}, L^{(n)}) \geq \\ &\geq C_\lambda(W^{(n+1)}, P^{(n+1)}, L^{(n+1)}) \geq ... \geq C_{\lambda,\min} \geq 0 \end{aligned} \tag{28}$$

Since the series is non-negative, it converges to a total cost (not larger than the initial one) with optimal forward dynamics $W^* = W^{(\infty)}$, pseudo-filter $P^* = P^{(\infty)}$ and control $L^* = L^{(\infty)}$ matrices. We have thus proven the first part of the following

Theorem 1. *Starting with arbitrary $W^{(0)}$, $P^{(0)}$ and $L^{(0)}$ and distribution of initial conditions $p_0(x, z)$, the coordinate-descent algorithm defined by iterating in alternation W-, P- and L-passes converges to an improved triplet of forward dynamics, pseudo-filter and control matrices W^*, P^* and L^*. The improved triplet corresponds to a critical point of the cost function in Eq. 12.*

We remark that it is straightforward to extend our algorithm to the case where any of the matrices W_t, P_t and L_t are fixed simply by not updating the corresponding matrices using the above passes, still enjoying convergence properties.

Lagrangian, Fixed-Point Equations, and Critical Points To complete the last part of the theorem, that is, that after convergence the triplet W^*, P^* and L^* is a critical point of the cost function 12, we must show that they solve all fixed-point equations of the Lagrangian,

$$\begin{aligned} C_{\mathcal{L},\lambda} = &\sum_{t=0}^{T} \left[\mathrm{Tr}(Q_t S_t^{xx}) + \mathrm{Tr}(L_t^\top R_t L_t S_t^{zz}) + \lambda \,\mathrm{Tr}(S_t^{zz})\right] \\ &- \sum_{t=1}^{T} \left[\mathrm{Tr}(\Lambda_t G_t^{xx}) + \mathrm{Tr}(\Omega_t G_t^{zz}) + \mathrm{Tr}(\Gamma_t G_t^{xz})\right]. \end{aligned} \tag{29}$$

where Λ_t, Ω_t and Γ_t are matrices of Lagrange multipliers.

The internal-activity penalty contributes an additional term to the stationarity condition with respect to S_t^{zz}, given by

$$\frac{\partial}{\partial S_t^{zz}} \left[\lambda \,\mathrm{Tr}(S_t^{zz})\right] = \lambda I_n. \tag{30}$$

Consequently, the penalty introduces the additional term λI_n in the backward recursion for Ω_t, while leaving the recursions for Λ_t and Γ_t unchanged in form. Moreover, the stationarity equations for W_t, P_t, and L_t retain their original algebraic forms, since the penalty does not depend explicitly on these matrices when the second-order moments are treated as independent variables.

The constraints $G_t^{xx} = G_t^{zz} = G_t^{xz} = 0$ enforce the second-order moment dynamics: each constraint is the difference between S_{t+1}^{xx}, S_{t+1}^{zz}, or S_{t+1}^{xz} and the corresponding update predicted from the moments at time t by Eqs. 14. Indeed, the fixed point equations of the Lagrangian $\partial C_{\mathcal{L}}/\partial W_t = 0$, $\partial C_{\mathcal{L}}/\partial P_t = 0$ and $\partial C_{\mathcal{L}}/\partial L_t = 0$ are identical to Eqs. 23,26,27, respectively, which must be satisfied after convergence by the improved triplet W^*, P^* and L^*. A straightforward calculation shows that the Lagrange equations $\partial C_{\mathcal{L}}/\partial S_t^{xx} = 0$, $\partial C_{\mathcal{L}}/\partial S_t^{zz} = 0$, and $\partial C_{\mathcal{L}}/\partial S_t^{xz} = 0$ lead exactly to the coefficient Eqs. 18, which are again satisfied by the improved triplet. Finally, the derivatives of the Lagrangian with respect to the multipliers reduce to the second-order moment Eqs. 14, which are satisfied by the improved triplet. Thus, the improved triplet is a fixed-point solution of the Lagrangian 29 and therefore a critical point of the cost function 12. The pseudocode is provided in Sec. A.1.2.

Local Robustness of Regularized Model-Free Solutions To assess the robustness of the regularized solutions, we perturbed the optimized W_t, P_t, and L_t, both separately and jointly, in the 3D reaching ($\lambda = 10^{-4}$) and redundant-arm ($\lambda = 10^{-3}$) tasks at $\sigma_\eta = 0.4$. For each configuration, we generated 50 random perturbation directions with independent Gaussian entries, normalized to the Frobenius norm of the corresponding optimized matrix sequence, and tested both signs at relative magnitudes of 1% and 5%, while preserving fixed parameters. All 1600 perturbations increased the regularized objective C_λ, providing numerical evidence consistent with local optimality of the obtained solutions.

A.1.1 Solution to the LQMI with Model-Matched Control

The solution to the LQMI problem with model-matched control can be obtained through an analogous derivation, including the possibility of penalizing internal state activity as in Eq. 9 and incorporating internal state-dependent multiplicative noise, as in the LQMI-FID formulation. This additional noise source was not included in the original LQMI formulation [24]. In this case, the internal dynamics are constrained to follow a Kalman-like recursion by imposing

$$P_t = K_t, \qquad W_t = M_t \equiv A + BL_t - K_t H. \tag{31}$$

The optimization therefore seeks critical points of the original or regularized objective only with respect to the filter and control gains, K_t and L_t. The derivation follows the same cost-to-go and Lagrangian argument used in Sec. A.1; we therefore omit the intermediate algebra and report only the resulting fixed-point equations and algorithmic updates. The resulting equations for the Lagrange multipliers, corresponding to Eq. 18 in the model-free case, are

$$\begin{aligned}
\Lambda_t &= Q_t + A^\top \Lambda_{t+1} A + H^\top K_t^\top \Omega_{t+1} K_t H + H^\top K_t^\top \Gamma_{t+1} A + \sum\nolimits_i D_i^\top K_t^\top \Omega_{t+1} K_t D_i \\
\Omega_t &= \lambda I_n + L_t^\top R_t L_t + L_t^\top B^\top \Lambda_{t+1} B L_t + M_t^\top \Omega_{t+1} M_t + M_t^\top \Gamma_{t+1} B L_t \\
&\quad + \sum\nolimits_i L_t^\top C_i^\top \Lambda_{t+1} C_i L_t + \sum_j (Z_j^z)^\top \Omega_{t+1} Z_j^z \\
\Gamma_t &= L_t^\top B^\top \tilde{\Lambda}_{t+1} A + M_t^\top \tilde{\Omega}_{t+1} K_t H + M_t^\top \Gamma_{t+1} A + L_t^\top B^\top \Gamma_{t+1}^\top K_t H \,, \qquad (32)
\end{aligned}$$

while the ones for the fixed-point equations for K_t and L_t, corresponding to Eqs. 26 and 27, are

$$L_t = E_t^{-1} \left(F_t S_t^{xz} (S_t^{zz})^{-1} + J_t \right) \qquad (33)$$

$$K_t = \left(S_{AH} + \tilde{\Omega}_{t+1}^{-1} \Gamma_{t+1} S_{LH} \right) S_{HH}^{-1} \,, \qquad (34)$$

where we use the symmetric matrices $\tilde{\Lambda}_t = \Lambda_t + \Lambda_t^\top$ and $\tilde{\Omega}_t = \Omega_t + \Omega_t^\top$, and define

$$E_t = 2R_t + B^\top (\tilde{\Lambda}_{t+1} + \tilde{\Omega}_{t+1} + \Gamma_{t+1} + \Gamma_{t+1}^\top) B + \sum\nolimits_i C_i^\top \tilde{\Lambda}_{t+1} C_i \,, \qquad (35)$$

$$F_t = -B^\top (\tilde{\Lambda}_{t+1} A + \tilde{\Omega}_{t+1} K_t H + \Gamma_{t+1} A + \Gamma_{t+1}^\top K_t H) \,, \qquad (36)$$

$$J_t = -B^\top (\tilde{\Omega}_{t+1} + \Gamma_{t+1}^\top)(A - K_t H) \,, \qquad (37)$$

and

$$S_{AH} = (A + BL_t)(S_t^{zz} - (S_t^{xz})^\top) H^\top \,, \qquad (38)$$

$$S_{LH} = \left(-A(S_t^{xx} - S_t^{xz}) + B L_t (S_t^{zz} - (S_t^{xz})^\top) \right) H^\top \,, \qquad (39)$$

$$S_{HH} = H(S_t^{xx} + S_t^{zz} - S_t^{xz} - (S_t^{xz})^\top) H^\top + \Sigma_\omega + \sum\nolimits_i D_i S_t^{xx} D_i^\top \,. \qquad (40)$$

As in the model-free formulation, the internal-activity penalty contributes only the additional term λI_n to the backward recursion for Ω_t. The fixed-point equations for K_t and L_t therefore retain their original algebraic forms, with the effect of regularization entering indirectly through the modified Lagrange multipliers. Setting $\lambda = 0$ recovers the unregularized objective.

The second-order moment updates are given by

$$\begin{aligned}
S_{t+1}^{xx} &= A S_t^{xx} A^\top + A S_t^{xz} L_t^\top B^\top + B L_t (S_t^{xz})^\top A^\top + B L_t S_t^{zz} L_t^\top B^\top + \Sigma_t^{xx} \\
S_{t+1}^{zz} &= K_t H S_t^{xx} H^\top K_t^\top + K_t H S_t^{xz} M_t^\top + M_t (S_t^{xz})^\top H^\top K_t^\top + M_t S_t^{zz} M_t^\top + \Sigma_t^{zz} \\
S_{t+1}^{xz} &= A S_t^{xx} H^\top K_t^\top + B L_t S_t^{zz} M_t^\top + A S_t^{xz} M_t^\top + B L_t (S_t^{xz})^\top H^\top K_t^\top \,, \qquad (41)
\end{aligned}$$

with $\Sigma_t^{xx} = \Sigma_\xi + \sum_i C_i L_t S_t^{zz} L_t^\top C_i^\top$, and $\Sigma_t^{zz} = K_t \left(\Sigma_\omega + \sum_i D_i S_t^{xx} D_i^\top \right) K_t^\top + \Sigma_\eta + \sum_j Z_j^z S_t^{zz} (Z_j^z)^\top$. As in the model-free case, these equations define an iterative coordinate-descent algorithm for which the corresponding objective is monotonically non-increasing. The original LQMI formulation considered by [24, 25] is recovered by setting $\lambda = 0$ and $Z_j^z = 0$. The pseudocode, analogous to Algorithm 1, is provided in Sec. A.1.3, Algorithm 2.

A.1.2 Pseudocode – Model-Free Control

The pseudocode below shows one possible ordering of the coordinate-descent updates; other orderings are equally valid, as long as each step minimizes the cost with respect to the selected variable while keeping the others fixed.

Algorithm 1 Model-Free Control

Input: $S_0^{xx}, S_0^{xz}, S_0^{zz}$; initial guesses $L^{(0)}_{0,\dots,T}, P^{(0)}_{0,\dots,T}, W^{(0)}_{0,\dots,T}$; system and cost parameters.
Output: Optimal matrices $L^*_{0,\dots,T}, P^*_{0,\dots,T}, W^*_{0,\dots,T}$.
Steps:
for each iteration $k = 1, \dots,$ optimization steps **do**
  $\Lambda_{1,\dots,T}, \Omega_{1,\dots,T}, \Gamma_{1,\dots,T} \leftarrow$ Eqs. 18 using $P^{(k-1)}_{1,\dots,T}$, $W^{(k-1)}_{1,\dots,T}$ and $L^{(k-1)}_{1,\dots,T}$ (backward equations)
  for each iteration $t = 0, \dots, T-1$ **do**
    $P_t^{(k)} \leftarrow$ Eq. 26,
    $S^{xx}_{t+1}, S^{xz}_{t+1}, S^{zz}_{t+1} \leftarrow$ Eqs. 14 using $P_t^{(k)}$, $W_t^{(k-1)}$ and $L_t^{(k-1)}$
  end for
  $\Lambda_{1,\dots,T}, \Omega_{1,\dots,T}, \Gamma_{1,\dots,T} \leftarrow$ Eqs. 18 using $P^{(k)}_{1,\dots,T}$, $W^{(k-1)}_{1,\dots,T}$ and $L^{(k-1)}_{1,\dots,T}$ (backward equations)
  for each iteration $t = 0, \dots, T-1$ **do**
    $W_t^{(k)} \leftarrow$ Eq. 23,
    $S^{xx}_{t+1}, S^{xz}_{t+1}, S^{zz}_{t+1} \leftarrow$ Eqs. 14 using $P_t^{(k)}$, $W_t^{(k)}$ and $L_t^{(k-1)}$
  end for
  $\Lambda_{1,\dots,T}, \Omega_{1,\dots,T}, \Gamma_{1,\dots,T} \leftarrow$ Eqs. 18 using $P^{(k)}_{1,\dots,T}$, $W^{(k)}_{1,\dots,T}$ and $L^{(k-1)}_{1,\dots,T}$ (backward equations)
  for each iteration $t = 0, \dots, T-1$ **do**
    $L_t^{(k)} \leftarrow$ Eq. 27,
    $S^{xx}_{t+1}, S^{xz}_{t+1}, S^{zz}_{t+1} \leftarrow$ Eqs. 14 using $P_t^{(k)}$, $W_t^{(k)}$ and $L_t^{(k)}$
  end for
end for
$P^*_{0,\dots,T} \leftarrow P^{(k)}_{0,\dots,T}$; $W^*_{0,\dots,T} \leftarrow W^{(k)}_{0,\dots,T}$; $L^*_{0,\dots,T} \leftarrow L^{(k)}_{0,\dots,T}$

A.1.3 Pseudocode – Model-Matched Control

As in the model-free case, the pseudocode below shows one possible ordering of the coordinate-descent updates; other orderings are equally valid provided that each step minimizes the cost with respect to the selected variable while keeping the others fixed.

Algorithm 2 Model-Matched Control

Input: $S_0^{xx}, S_0^{xz}, S_0^{zz}$; initial guesses $L^{(0)}_{0,\dots,T}, K^{(0)}_{0,\dots,T}$; system parameters.
Output: Optimal gains $L^*_{0,\dots,T}, K^*_{0,\dots,T}$.
Steps:
for each iteration $k = 1, \dots,$ optimization steps **do**
  $\Lambda_{1,\dots,T}, \Omega_{1,\dots,T}, \Gamma_{1,\dots,T} \leftarrow$ Eqs. 32 using $L^{(k-1)}_{0,\dots,T}$ and $K^{(k-1)}_{0,\dots,T}$ (backward equations)
  for each iteration $t = 0, \dots, T-1$ **do**
    $L_t^{(k)} \leftarrow$ Eq. 33,
    $S^{xx}_{t+1}, S^{xz}_{t+1}, S^{zz}_{t+1} \leftarrow$ Eqs. 41 using $L_t^{(k)}$ and $K_t^{(k-1)}$
  end for
  $\Lambda_{1,\dots,T}, \Omega_{1,\dots,T}, \Gamma_{1,\dots,T} \leftarrow$ Eqs. 32 using $L^{(k)}_{0,\dots,T}$ and $K^{(k-1)}_{0,\dots,T}$ (backward equations)
  for each iteration $t = 0, \dots, T-1$ **do**
    $K_t^{(k)} \leftarrow$ Eq. 34,
    $S^{xx}_{t+1}, S^{xz}_{t+1}, S^{zz}_{t+1} \leftarrow$ Eqs. 41 using $L_t^{(k)}$ and $K_t^{(k)}$
  end for
end for
$L^*_{0,\dots,T} \leftarrow L^{(k)}_{0,\dots,T}$; $K^*_{0,\dots,T} \leftarrow K^{(k)}_{0,\dots,T}$

A.2 Separated Estimation-Control Objectives in the LQMI Problem

We introduce a third model in which estimation and control are optimized according to two distinct objectives. In this model, the internal variable is explicitly constrained to act as an estimate of the external state: the filter gains are optimized to minimize the estimation error between z_t and x_t,

and the control gains are then optimized on top of this estimate to minimize the original task cost. Here, we restrict the analysis to the case with $\lambda = 0$. The extension to nonzero internal-activity regularization would follow the same procedure described above. For simplicity, we do not consider internal state-dependent multiplicative noise, setting $Z_j^z = 0$, as this additional noise source is not central to the comparison between separated and jointly optimized estimation and control.

This construction enforces, by design, the logic that instead emerges naturally in the classical LQG setting. In that case, the separation principle holds: estimation and control can be solved as two mathematically independent problems, with the optimal controller given by the Linear Quadratic Regulator (LQR) applied to the state estimate produced by an optimal Kalman filter. Equivalently, in the absence of multiplicative and internal noise, minimizing the quadratic control cost leads to an optimal linear estimator that minimizes mean-squared estimation error, followed by a controller built on this estimate.

As discussed above, this separation no longer holds in the full LQMI setting, where multiplicative and internal noise couple estimation and control. Here, we therefore test whether an architecture that preserves the logic of separation – *first-filter-then-control* – can still achieve optimal performance, while allowing iterative dependence between filter and controller. Specifically, we alternate between optimizing the filter gains K_t with respect to an estimation objective, keeping the controller fixed, and optimizing the control gains L_t with respect to the original task cost, keeping the filter fixed. Thus, estimation and control are not mathematically independent, but they are still driven by different objectives: the filter is optimized to estimate the state, whereas the controller is optimized to solve the task. We expect this separated-objective procedure to perform worse than the jointly optimized model-matched solution, particularly in the presence of internal noise, where the orthogonality principle breaks down and the optimal internal variable need not coincide with the best state estimate.

Analogously to Eq. 29, we introduce the Lagrangian objective for the filter-gain optimization,

$$\begin{aligned} C_{\mathcal{L},\text{est}} = &\sum_{t=0}^{T} \left[\operatorname{tr}(S_t^{xx}) + \operatorname{tr}(S_t^{zz}) - 2\operatorname{tr}(S_t^{xz})\right] \\ &- \sum_{t=1}^{T} \left[\operatorname{tr}\left(\Lambda_t^{\text{est}} G_t^{xx}\right) + \operatorname{tr}\left(\Omega_t^{\text{est}} G_t^{zz}\right) + \operatorname{tr}\left(\Gamma_t^{\text{est}} G_t^{xz}\right)\right], \end{aligned} \tag{42}$$

which is obtained from the mean-squared estimation error

$$C_{\text{est}} = \sum_{t=0}^{T} \mathbb{E}\left[(x_t - z_t)^\top (x_t - z_t)\right] = \sum_{t=0}^{T} \left[\operatorname{tr}(S_t^{xx}) + \operatorname{tr}(S_t^{zz}) - 2\operatorname{tr}(S_t^{xz})\right], \tag{43}$$

where again $S_t^{xx} = \mathbb{E}[x_t x_t^\top]$, $S_t^{zz} = \mathbb{E}[z_t z_t^\top]$, and $S_t^{xz} = \mathbb{E}[x_t z_t^\top]$. The constraints $G_t^{xx} = 0$, $G_t^{zz} = 0$, and $G_t^{xz} = 0$ are the same moment-propagation constraints as in the task-cost optimization, while the Lagrange multipliers are now associated with the estimation objective. Notably, we restrict this model to matched internal dynamics, with $W_t = M_t = A + BL_t - K_t H$. Thus, the internal state follows the same forward model as the external dynamics, up to the correction induced by sensory feedback. This enforces the estimation-first, control-second logic underlying the separated-objective baseline.

Following the same derivation as in the original Lagrangian formulation, the estimation cost-to-go can be written as a quadratic form,

$$C_t^{\text{est}}(x, z) = \operatorname{tr}\left(\Lambda_t^{\text{est}} x x^\top + \Omega_t^{\text{est}} z z^\top + \Gamma_t^{\text{est}} x z^\top\right) + \gamma_t^{\text{est}} \,. \tag{44}$$

with

$$\begin{aligned} \Lambda_t^{\text{est}} &= I + A^\top \Lambda_{t+1}^{\text{est}} A + H^\top K_t^\top \Omega_{t+1}^{\text{est}} K_t H + H^\top K_t^\top \Gamma_{t+1}^{\text{est}} A + \sum\nolimits_i D_i^\top K_t^\top \Omega_{t+1}^{\text{est}} K_t D_i \\ \Omega_t^{\text{est}} &= I + L_t^\top B^\top \Lambda_{t+1}^{\text{est}} B L_t + M_t^\top \Omega_{t+1}^{\text{est}} M_t + M_t^\top \Gamma_{t+1}^{\text{est}} B L_t + \sum\nolimits_i L_t^\top C_i^\top \Lambda_{t+1}^{\text{est}} C_i L_t \\ \Gamma_t^{\text{est}} &= -2I + L_t^\top B^\top \tilde{\Lambda}_{t+1}^{\text{est}} A + M_t^\top \tilde{\Omega}_{t+1}^{\text{est}} K_t H + M_t^\top \Gamma_{t+1}^{\text{est}} A + L_t^\top B^\top {\Gamma_{t+1}^{\text{est}}}^\top K_t H \\ \gamma_t^{\text{est}} &= \operatorname{tr}(\Lambda_{t+1}^{\text{est}} \Sigma_\xi + \Omega_{t+1}^{\text{est}} K_t \Sigma_\omega K_t^\top + \Omega_{t+1}^{\text{est}} \Sigma_\eta) + \gamma_{t+1}^{\text{est}} \,, \end{aligned} \tag{45}$$

where I denotes the $m \times m$ identity matrix. The terminal conditions for the backward recursions in Eqs. 45 are therefore

$$\Lambda_T^{\text{est}} = I, \qquad \Omega_T^{\text{est}} = I, \qquad \Gamma_T^{\text{est}} = -2I, \qquad \gamma_T^{\text{est}} = 0. \tag{46}$$

As in the original Lagrangian formulation, Eqs. 45 coincide with the backward updates for the Lagrange multipliers in Eq. 42. These updates are obtained by taking derivatives of $C_{\mathcal{L},\mathrm{est}}$ with respect to the second-order moments.

We can then optimize the filter gains using the same optimization scheme as before. Specifically, we impose the stationarity condition $\partial C_{\mathcal{L},\mathrm{est}}/\partial K_t = 0$ which gives the fixed-point equation

$$K_t^{\mathrm{est}} = \left(S_{AH} + \left(\tilde{\Omega}_{t+1}^{\mathrm{est}}\right)^{-1} \Gamma_{t+1}^{\mathrm{est}} S_{LH} \right) S_{HH}^{-1}. \tag{47}$$

Here, as before, we use the symmetrized matrices

$$\tilde{\Lambda}_t^{\mathrm{est}} = \Lambda_t^{\mathrm{est}} + \left(\Lambda_t^{\mathrm{est}}\right)^\top, \qquad \tilde{\Omega}_t^{\mathrm{est}} = \Omega_t^{\mathrm{est}} + \left(\Omega_t^{\mathrm{est}}\right)^\top,$$

and use the same definition as the original Lagrangian derivation,

$$S_{AH} = (A + BL_t)\left(S_t^{zz} - (S_t^{xz})^\top\right) H^\top, \tag{48}$$

$$S_{LH} = \left[-A\left(S_t^{xx} - S_t^{xz}\right) + BL_t\left(S_t^{zz} - (S_t^{xz})^\top\right)\right] H^\top, \tag{49}$$

$$S_{HH} = H\left(S_t^{xx} + S_t^{zz} - S_t^{xz} - (S_t^{xz})^\top\right) H^\top + \Sigma_\omega + \sum_i D_i S_t^{xx} D_i^\top. \tag{50}$$

Notably, replacing the task cost with the estimation cost in the optimization of the filter gains K_t only modifies the backward equations for the Lagrange multipliers, while leaving the rest of the derivation unchanged. Indeed, the estimation cost does not introduce additional terms that depend explicitly on K_t, nor does it remove any K_t-dependent terms from the cost-to-go. Instead, it only changes the derivatives with respect to the second-order moments, thereby modifying the corresponding Lagrange multipliers.

For the optimization of the control gains L_t, we use the original task objective,

$$C = \sum_{t=0}^{T} \mathbb{E}\left[x_t^\top Q_t x_t + u_t^\top R_t u_t\right] \ . \tag{51}$$

The optimization is performed again under the matched internal-dynamics assumption. This yields the following backward equations for the Lagrange multipliers

$$\begin{aligned}
\Lambda_t &= Q_t + A^\top \Lambda_{t+1} A + H^\top K_t^\top \Omega_{t+1} K_t H + H^\top K_t^\top \Gamma_{t+1} A + \textstyle\sum_i D_i^\top K_t^\top \Omega_{t+1} K_t D_i \\
\Omega_t &= L_t^\top R_t L_t + L_t^\top B^\top \Lambda_{t+1} B L_t + M_t^\top \Omega_{t+1} M_t + M_t^\top \Gamma_{t+1} B L_t + \textstyle\sum_i L_t^\top C_i^\top \Lambda_{t+1} C_i L_t \\
\Gamma_t &= L_t^\top B^\top \tilde{\Lambda}_{t+1} A + M_t^\top \tilde{\Omega}_{t+1} K_t H + M_t^\top \Gamma_{t+1} A + L_t^\top B^\top \Gamma_{t+1}^\top K_t H \ ,
\end{aligned} \tag{52}$$

together with the following fixed-point equation for L_t

$$L_t = E_t^{-1}\left(F_t S_t^{xz} (S_t^{zz})^{-1} + J_t\right) \tag{53}$$

where again $\tilde{\Lambda}_t = \Lambda_t + \Lambda_t^\top$ and $\tilde{\Omega}_t = \Omega_t + \Omega_t^\top$, and

$$E_t = 2R_t + B^\top(\tilde{\Lambda}_{t+1} + \tilde{\Omega}_{t+1} + \Gamma_{t+1} + \Gamma_{t+1}^\top) B + \textstyle\sum_i C_i^\top \tilde{\Lambda}_{t+1} C_i \ , \tag{54}$$

$$F_t = -B^\top(\tilde{\Lambda}_{t+1} A + \tilde{\Omega}_{t+1} K_t H + \Gamma_{t+1} A + \Gamma_{t+1}^\top K_t H) \ , \tag{55}$$

$$J_t = -B^\top(\tilde{\Omega}_{t+1} + \Gamma_{t+1}^\top)(A - K_t H) \ . \tag{56}$$

We can then construct an iterative procedure, as we did in the original formulation, to obtain the separated estimation-control solution, in which the filter gains define an optimal linear estimator minimizing the mean-squared estimation error, while the control gains define an optimal linear regulator built on top of this estimate. The procedure is summarized in the following pseudocode.

Algorithm 3 Separated Estimation-Control (SEC)

Input: $S_0^{xx}, S_0^{xz}, S_0^{zz}$; initial guesses $L^{(0)}_{0,\dots,T}, K^{(0)}_{0,\dots,T}$; system parameters.
Output: Optimal gains $L^*_{0,\dots,T}, K^*_{0,\dots,T}$.
Steps:
for each iteration $k = 1, \dots,$ optimization steps **do**
  $\Lambda_{1,\dots,T}, \Omega_{1,\dots,T}, \Gamma_{1,\dots,T} \leftarrow$ Eqs. 52 using $L^{(k-1)}_{0,\dots,T}$ and $K^{(k-1)}_{0,\dots,T}$ (backward equations)
  for each iteration $t = 0, \dots, T-1$ **do**
    $L_t^{(k)} \leftarrow$ Eq. 53,
    $S^{xx}_{t+1}, S^{xz}_{t+1}, S^{zz}_{t+1} \leftarrow$ Eqs. 41 using $L_t^{(k)}$ and $K_t^{(k-1)}$
  end for
  $\Lambda_{1,\dots,T}, \Omega_{1,\dots,T}, \Gamma_{1,\dots,T} \leftarrow$ Eqs. 45 using $L^{(k)}_{0,\dots,T}$ and $K^{(k-1)}_{0,\dots,T}$ (backward equations)
  for each iteration $t = 0, \dots, T-1$ **do**
    $K_t^{(k)} \leftarrow$ Eq. 47,
    $S^{xx}_{t+1}, S^{xz}_{t+1}, S^{zz}_{t+1} \leftarrow$ Eqs. 41 using $L_t^{(k)}$ and $K_t^{(k)}$
  end for
end for
$L^*_{0,\dots,T} \leftarrow L^{(k)}_{0,\dots,T}$; $K^*_{0,\dots,T} \leftarrow K^{(k)}_{0,\dots,T}$

A.3 Task Details

Unless otherwise specified, all internal state-dependent multiplicative noise matrices Z_j^z are set to zero in the numerical experiments.

A.3.1 3D Reaching Task

The 3D reaching task is an idealized state-space benchmark inspired by point-mass hand dynamics in three-dimensional space. The state contains the hand position and velocity,

$$x_t = \begin{pmatrix} p_t \\ \dot{p}_t \end{pmatrix} \in \mathbb{R}^6, \qquad p_t, \dot{p}_t \in \mathbb{R}^3.$$

The internal state has the same dimensionality, $n = 6$. In this task, the control signal, observation, and state are all six-dimensional, so that $m = n = p = k = 6$. The observation is full-state, meaning that both position and velocity are directly observed through noisy sensory feedback.

The deterministic part of the dynamics is a discrete-time linear kinematic model. With time step $\Delta t = 0.010$, the state dynamics are

$$x_{t+1} = Ax_t + Bu_t + \xi_t + \varepsilon_t C u_t,$$

with

$$A = \begin{pmatrix} 1 & 0 & 0 & \Delta t & 0 & 0 \\ 0 & 1 & 0 & 0 & \Delta t & 0 \\ 0 & 0 & 1 & 0 & 0 & \Delta t \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \qquad B = I_6,$$

where I_6 denotes the 6×6 identity matrix. The choice $B = I_6$ makes the task fully actuated in state space: the control signal has the same dimensionality as the state and can influence both position and velocity components. Thus, the control input should be interpreted as an abstract state-space correction signal, not as a biomechanically realistic force. The passive dynamics encoded by A still preserve the kinematic relation between position and velocity, while full actuation avoids degenerate control directions and improves numerical stability, providing a clean benchmark for comparing model-matched and model-free internal dynamics.

Control-dependent noise is scaled by

$$C = \sigma_\varepsilon B.$$

Since the full state is observed, the observation model is

$$y_t = Hx_t + \omega_t + \rho_t D x_t, \qquad H = I_6, \qquad D = \sigma_\rho H.$$

The task is defined in coordinates centered at the target, so the controller must drive the state toward the origin by the end of the movement. We use a terminal state cost,

$$Q_t = 0_{6\times 6}, \qquad t = 0, \dots, T-1,$$

and

$$Q_T = \mathrm{diag}\,(10, 10, 10, 1, 1, 1)\,.$$

Thus, final position errors are penalized more strongly than final velocity errors. The control cost is

$$R_t = rI_6, \qquad r = 10^{-4}, \qquad t = 0, \dots, T-1,$$

and $R_T = 0$.

The additive process, sensory, and internal noise covariances are

$$\Sigma_\xi = \sigma_\xi^2 I_6, \qquad \Sigma_\omega = \sigma_\omega^2 I_6, \qquad \Sigma_\eta = \sigma_\eta^2 I_6.$$

We set

$$\sigma_\xi = 0.5, \qquad \sigma_\omega = 0.5, \qquad \sigma_\varepsilon = 0.5, \qquad \sigma_\rho = 0.5,$$

and sweep the internal-noise magnitude over

$$\sigma_\eta \in [0, 0.4],$$

using 10 evenly spaced values. The horizon is $T = 100$ time steps.

The initial state is deterministic and displaced from the target:

$$\mathbb{E}[x_0] = \begin{pmatrix} 1.5 & 1.0 & 2.5 & 10^{-5} & 10^{-5} & 10^{-5} \end{pmatrix}^\top, \qquad \mathbb{E}[z_0] = \mathbb{E}[x_0],$$

with zero initial covariance,

$$\Sigma_{x_0} = 0_{6\times 6}, \qquad \Sigma_{z_0} = 0_{6\times 6}.$$

To keep the initialization consistent with the model-matched convention, under which z_0 is an unbiased estimate of x_0 at the first time step for the initial conditions considered here, we fix the first-step model-free matrices L_0, P_0, and W_0 to corresponding solutions of model-matched control. The model-free optimization is therefore performed from $t = 1$ onward.

A.3.2 Redundant-Arm Control Task

The redundant-arm task models a three-joint arm controlled by redundant muscle-like activation channels and stabilizing around a reference posture. This setup can be thought of as a simplified linear approximation to a more complex musculo-skeletal plant: the linear dynamics describe local deviations around the stable posture, while the redundant actuation structure is motivated by biologically inspired arm models with multiple muscle-like actuators [64]. The state contains the deviations of the three joint angles from the reference posture and their corresponding angular velocities,

$$x_t = \begin{pmatrix} \theta_t \\ \dot\theta_t \end{pmatrix} \in \mathbb{R}^6, \qquad \theta_t, \dot\theta_t \in \mathbb{R}^3.$$

The internal state has the same dimensionality, $n = 6$, whereas the control input is nine-dimensional, $u_t \in \mathbb{R}^9$, corresponding to muscle-like actuation channels. Sensory observations are limited to joint angles, giving a three-dimensional observation vector, $y_t \in \mathbb{R}^3$. Thus, in this task $m = n = 6$, $p = 9$, and $k = 3$.

The continuous arm dynamics are approximated by linear damped joint dynamics around the reference posture. With time step $\Delta t = 0.010$, the discrete-time equations are

$$\begin{aligned} \theta_{t+1} &= \theta_t + \Delta t\, \dot\theta_t, \\ \dot\theta_{t+1} &= \left(I_3 - \Delta t\, M_{\mathrm{joint}}^{-1} D_{\mathrm{joint}}\right) \dot\theta_t + \Delta t\, M_{\mathrm{joint}}^{-1} S u_t, \end{aligned}$$

where M_{joint} is the diagonal joint-inertia matrix, D_{joint} is the joint-damping matrix, $S \in \mathbb{R}^{3\times 9}$ maps muscle-like activations to joint torques, and I_3 denotes the 3×3 identity matrix. We use

$$M_{\mathrm{joint}} = \mathrm{diag}(1.2, 0.8, 0.5), \qquad D_{\mathrm{joint}} = 2.0\, I_3.$$

The moment-arm matrix is

$$S = \begin{pmatrix} 1.2 & -1.0 & 0.0 & 0.8 & -0.6 & 0.0 & 0.5 & 0.0 & 0.0 \\ 0.0 & 0.0 & 1.0 & -0.4 & 0.6 & -0.5 & 0.0 & 0.5 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 & -0.3 & 0.6 \end{pmatrix}.$$

This matrix defines a structured redundant actuation map: several muscle-like channels affect multiple joints, with signs corresponding to flexor- or extensor-like actions. The numerical values are not intended as a detailed biomechanical fit, but they preserve the key property needed for the task, namely that a nine-dimensional control space is projected onto three joint torques [64]. As a result, different muscle-level commands can generate equivalent joint-level actions.

In state-space form, the dynamics are written as

$$x_{t+1} = Ax_t + Bu_t + \xi_t + \varepsilon_t C u_t,$$

with

$$A = \begin{pmatrix} I_3 & \Delta t\, I_3 \\ 0_{3\times 3} & I_3 - \Delta t\, M_{\text{joint}}^{-1} D_{\text{joint}} \end{pmatrix}, \qquad B = \begin{pmatrix} 0_{3\times 9} \\ \Delta t\, M_{\text{joint}}^{-1} S \end{pmatrix}.$$

Control-dependent noise is scaled by

$$C = \sigma_\varepsilon B.$$

Only the joint angles are observed, so the observation matrix is

$$H = (I_3 \quad 0_{3\times 3}), \qquad D = \sigma_\rho H,$$

where D defines the state-dependent sensory-noise scaling.

The cost penalizes deviations from the reference posture and control effort. For all time steps, the state-cost matrix is

$$Q_t = \operatorname{diag}\left(q_\theta, q_\theta, q_\theta, q_{\dot\theta}, q_{\dot\theta}, q_{\dot\theta}\right),$$

with $q_\theta = 1.0$ and $q_{\dot\theta} = 10^{-3}$. Thus, joint-angle deviations carry the dominant task penalty, while angular velocities are weakly penalized. The control cost is

$$R_t = rI_9, \qquad r = 10^{-2}, \qquad t = 0, \ldots, T-1,$$

and $R_T = 0$.

The additive process, sensory, and internal noise covariances are

$$\Sigma_\xi = \sigma_\xi^2 I_6, \qquad \Sigma_\omega = \sigma_\omega^2 I_3, \qquad \Sigma_\eta = \sigma_\eta^2 I_6.$$

We set

$$\sigma_\xi = 0.1, \qquad \sigma_\omega = 0.1, \qquad \sigma_\varepsilon = 0.1, \qquad \sigma_\rho = 0.1,$$

and sweep the internal-noise magnitude over

$$\sigma_\eta \in [0, 0.4],$$

using 20 evenly spaced values. The horizon is $T = 300$ time steps.

The initial conditions are deterministic and centered at the reference posture:

$$\mathbb{E}[x_0] = 0, \qquad \mathbb{E}[z_0] = 0,$$
$$\Sigma_{x_0} = 0_{6\times 6}, \qquad \Sigma_{z_0} = 0_{6\times 6}.$$

A.4 Additional Results and Methods

A.4.1 Unregularized Model-Free Control in the Three-Dimensional Reaching Task

Here, we consider the three-dimensional reaching task of Sec. 4 without any penalty on internal state activity ($\lambda = 0$ for both approaches). These results complement the regularized analysis presented in the main text and provide further insight into the strategies adopted by the model-free controller when internal activity is left unconstrained.

The algorithm for optimizing the model-free controller exhibits monotonic convergence across internal-noise levels (Fig. 4a). As in the regularized case, model-free control outperforms classical model-matched control, with the performance gap increasing as the internal-noise magnitude σ_η

becomes larger (Fig. 4b). Moreover, the optimized recurrent dynamics W_t increasingly depart from the model-matched reference $\widetilde{W}_t = A + BL_t - P_tH$ (Fig. 4c). The internal state z_t also no longer tracks the corresponding external-state components, as illustrated by the representative trajectories in Figs. 4d–e.

The optimized gains reveal how the two approaches respond to increasing internal noise. Both approaches increase the weight assigned to sensory inputs (Figs. 4f–g; see K and P) while decreasing the control gains (see L). However, in model-matched control, changes in the control and filter gains necessarily reshape the recurrent dynamics of the internal state z_t, whose evolution is given by $z_{t+1} = M_t z_t + K_t y_t + \eta_t$, with $M_t = A + BL_t - K_tH$. Thus, the recurrent dynamics M_t are entirely determined by the requirement of preserving the Kalman-like recursion (Fig. 4f). By contrast, in model-free control, P_t, L_t, and W_t can be optimized independently (Fig. 4g). Interestingly, the effective dimensionality and determinant of the recurrent dynamics remain relatively stable across internal-noise levels (Fig. 4h).

Importantly, in the absence of an internal-activity penalty, model-free control can exploit an internal state scaling mechanism. By increasing the sensory-input gain P_t while decreasing the control gain L_t, the controller can reduce the effective impact of additive internal noise on motor output, at the expense of amplifying internal state activity. The opposing trends of P_t and L_t observed in Fig. 4g are consistent with this strategy. For example, at $\sigma_\eta = 0.13$, the time-averaged internal state activity, measured by $\langle \mathrm{Tr}(S_t^{zz}) \rangle_t$ after excluding the first and last 20 time steps (to remove initial and final transients), is approximately 677 for model-free control, compared with 24 for model-matched control, corresponding to a 28-fold increase. This amplification is consistent with internal state rescaling as a mechanism for reducing sensitivity to additive internal noise, without necessarily requiring an increase in the effective dimensionality or determinant of the recurrent dynamics W_t.

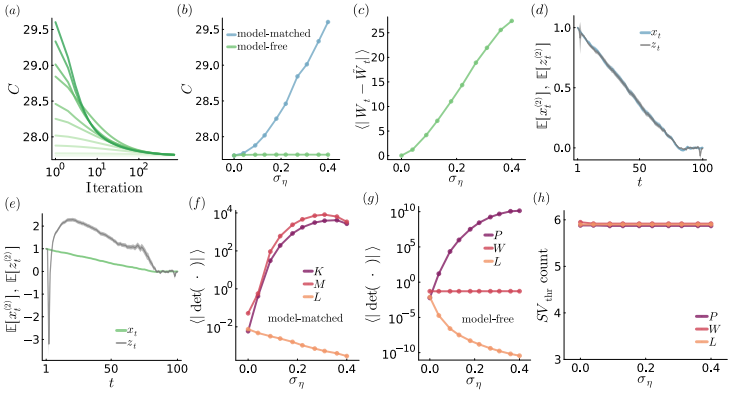


Figure 4: *Model-Free Control Without Internal-Activity Regularization in the Three-Dimensional Reaching Task.* **(a)** Optimization traces for unregularized model-free control ($\lambda = 0$) across internal-noise levels σ_η, showing the expected task cost C over iterations. **(b)** Expected cumulative task cost C as a function of σ_η for model-matched and unregularized model-free control. **(c)** Mean deviation between the optimized internal dynamics W_t of the model-free controller and the model-matched reference $\widetilde{W}_t = A + BL_t - P_tH$, averaged over time. **(d,e)** Trial-averaged trajectories of the second physical-state component, $x_t^{(2)}$, and the corresponding internal variable, $z_t^{(2)}$, for the model-matched (d) and model-free (e) approaches at $\sigma_\eta = 0.13$; shaded bands denote SEM over Monte Carlo trials. **(f,g)** Time-averaged absolute determinants of the optimized gain and internal-dynamics matrices for model-matched and model-free control, respectively, as a function of σ_η. **(h)** Effective dimensionality of the gain and internal-dynamics matrices of the model-free controller, quantified by the number of singular values exceeding 1% of the largest singular value (SV_{thr} count), averaged over time.

Thus, part of the performance advantage in the unregularized setting may arise from internal state amplification rather than exclusively from a reorganization of internal computations. Crucially, the regularized results presented in the main text demonstrate that model-free control maintains its performance advantage even when this scaling mechanism is penalized. For example, at $\sigma_\eta = 0.40$, regularized model-free control achieves a 5.99% reduction in task cost relative to model-matched control, compared with 6.26% without regularization. This suggests that the substantial amplification

of internal state activity observed without regularization is not essential for achieving improved task performance.

A.4.2 Recovery of the Classical LQMI Solution

We verified the agreement between LQMI-FID and classical LQMI in the three-dimensional reaching task at zero internal noise ($\sigma_\eta = 0$), without activity regularization ($\lambda = 0$). Initializing LQMI-FID from the optimized LQMI solution and running 700 optimization cycles, both approaches achieved the same task cost ($C_{\text{task}} = 27.74$), with variations below 10^{-14}. The relative mean Frobenius difference between the optimized internal dynamics W_t and the LQMI recurrence $A + BL_t - K_tH$ was only 5.0×10^{-11}, excluding initial and terminal transients. Thus, freeing the internal dynamics provided no additional performance benefit under these conditions.

A.4.3 Internal Noise in the Linear-Gaussian Setting

To verify that the advantage of model-free control does not arise from the non-Gaussianity induced by multiplicative noise, we repeated the redundant-arm task with both multiplicative noise sources set to zero, retaining only Gaussian additive process, sensory, and internal noise. Under these conditions, the system is linear-Gaussian, and the posterior over the external state remains Gaussian. Nevertheless, model-free control continued to outperform model-matched control, with the relative task-cost improvement increasing monotonically from 0% at $\sigma_\eta = 0$ to 16.34% at $\sigma_\eta = 0.4$. The advantage also persisted with internal-activity regularization ($\lambda = 10^{-3}$) in model-free control: relative to unregularized model-matched control, task costs were reduced by 2.5%, 6.9%, 10.3%, and 12.9% at $\sigma_\eta = 0.10, 0.20, 0.30, 0.40$, respectively. Thus, the advantage does not require non-Gaussian posterior distributions, but arises from allowing the internal dynamics to adapt to noise affecting the internal computations themselves.

A.4.4 Methods: Linear Decoding Analysis for the Redundant-Arm Task

To test whether the internal state z_t contains information about the physical state x_t, independently of whether it represents x_t in the same coordinate system, we compute the optimal affine linear decoder from z_t to x_t. At each time step, we consider decoders of the form

$$\hat{x}_t^{\text{dec}} = a_t + \Phi_t z_t, \tag{57}$$

where Φ_t is a linear readout matrix, $\Phi_t \in \mathbb{R}^{m\times m}$ and a_t is a vector, $a_t \in \mathbb{R}^m$. The decoder is chosen to minimize the mean-squared decoding error,

$$(a_t^\star, \Phi_t^\star) = \arg\min_{a_t,\Phi_t} \mathbb{E}\left[|x_t - a_t - \Phi_t z_t|^2\right]. \tag{58}$$

Let $\mu_t^x = \mathbb{E}[x_t]$ and $\mu_t^z = \mathbb{E}[z_t]$ denote the means of the physical and internal states, and let

$$\Sigma_t^{xz} = \mathbb{E}\left[(x_t - \mu_t^x)(z_t - \mu_t^z)^\top\right], \qquad \Sigma_t^{zz} = \mathbb{E}\left[(z_t - \mu_t^z)(z_t - \mu_t^z)^\top\right] \tag{59}$$

be the cross-covariance between x_t and z_t and the covariance of z_t, respectively. The optimal linear decoder, achieving the minimum mean square error estimator, is then [69]

$$\Phi_t^\star = \Sigma_t^{xz}\left(\Sigma_t^{zz}\right)^{-1}, \qquad a_t^\star = \mu_t^x - \Phi_t^\star \mu_t^z. \tag{60}$$

We quantify decodability by the fraction of state variance explained by this optimal decoder. The residual covariance is

$$\Sigma_t^{\text{res}} = \Sigma_t^{xx} - \Sigma_t^{xz}\left(\Sigma_t^{zz}\right)^{-1}\Sigma_t^{zx}, \tag{61}$$

where Σ_t^{xx} is the covariance of x_t and $\Sigma_t^{zx} = (\Sigma_t^{xz})^\top$. We then define

$$R^2_{\text{dec}} = 1 - \frac{\sum_{t\in\mathcal{T}} \text{Tr}\left(\Sigma_t^{\text{res}}\right)}{\sum_{t\in\mathcal{T}} \text{Tr}\left(\Sigma_t^{xx}\right)}, \tag{62}$$

where $\mathcal{T}$ denotes the temporal analysis window. For example, $\mathcal{T} = \{1, \ldots, T\}$ when the whole trajectory is considered. This quantity measures the proportion of variance of the physical state that is linearly decodable from the internal state. Finally, decoded trajectories are obtained in each simulated trial by applying the time-dependent decoder,

$$\hat{x}_t^{\text{dec}} = a_t^\star + \Phi_t^\star z_t. \tag{63}$$

To further distinguish which components of the physical state are represented in the internal dynamics, we also compute decodability separately for different subsets of x_t. In the redundant arm task, the state is decomposed as $x_t = (\theta_t, \dot{\theta}_t)$, where $\theta_t \in \mathbb{R}^3$ denotes joint angles and $\dot{\theta}_t$ denotes joint velocities. Since the task cost directly penalizes deviations in joint angles through the matrices Q_t, which are also the only directly observed components of the state, we refer to θ_t as the primary task-relevant components and to $\dot{\theta}_t$ as the task-irrelevant components. For a subset of state indices $\mathcal{I} \subset \{1, \dots, m\}$, we define $x_t^{\mathcal{I}}$ as the corresponding subvector of x_t and compute the optimal affine decoder

$$\hat{x}_t^{\mathrm{dec},\mathcal{I}} = a_t^{\mathcal{I}} + \Phi_t^{\mathcal{I}} z_t, \tag{64}$$

where

$$\Phi_t^{\mathcal{I}} = \Sigma_{t,\mathcal{I}}^{xz} \left(\Sigma_t^{zz}\right)^{-1}, \qquad a_t^{\mathcal{I}} = \mu_t^{x,\mathcal{I}} - \Phi_t^{\mathcal{I}} \mu_t^z. \tag{65}$$

Here $\mu_t^{x,\mathcal{I}}$ denotes the selected components of μ_t^x, and $\Sigma_{t,\mathcal{I}}^{xz}$ denotes the rows of Σ_t^{xz} corresponding to the selected state components. We then compute

$$R^2_{\mathrm{dec}}(\mathcal{I}) = 1 - \frac{\sum_{t\in\mathcal{T}} \mathrm{Tr}\left(\Sigma_{t,\mathcal{I}}^{xx} - \Sigma_{t,\mathcal{I}}^{xz}\left(\Sigma_t^{zz}\right)^{-1}\Sigma_{t,\mathcal{I}}^{zx}\right)}{\sum_{t\in\mathcal{T}} \mathrm{Tr}\left(\Sigma_{t,\mathcal{I}}^{xx}\right)}, \tag{66}$$

where $\Sigma_{t,\mathcal{I}}^{xx}$ is the covariance matrix of $x_t^{\mathcal{I}}$. We use $\mathcal{I} = \{1, 2, 3\}$ for task-relevant joint-angle components and $\mathcal{I} = \{4, 5, 6\}$ for task-irrelevant velocity components.

A.4.5 Decoding Task-Relevant and Task-Irrelevant State Components

In Fig. 3e, Sec. 4, we show that the internal states in model-free control remain highly informative about the external state x_t across internal-noise levels, even though the optimized internal dynamics depart from the world-matched dynamics (Figs. 4c,e). Thus, model-free control does not discard information about the physical state; rather, it represents this information in transformed internal coordinates (Figs. 3g,h), supporting robust task performance under internal noise (Fig. 4b).

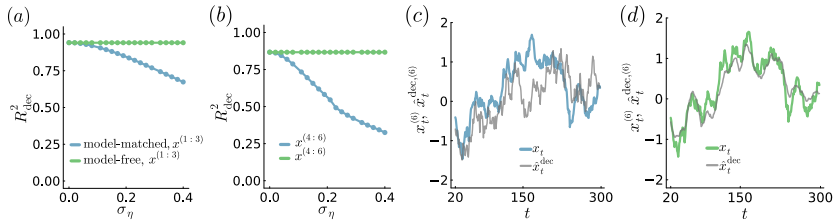


Figure 5: *Model-Free Control Preserves Decodability of Task-Relevant and Task-Irrelevant State Components Under Internal Noise*. **(a–b)** Linear decodability of the physical state x_t from the internal state z_t as a function of internal noise σ_η, for model-matched control (blue) and model-free control (green). At each time step, we compute the optimal affine decoder $\hat{x}_t^{\mathrm{dec}} = a_t + \Phi_t z_t$ and quantify decoding performance using R^2_{dec}, pooled over the temporal window $t = 20, \dots, T-2$ to exclude initial and final transients. **(a)** Decodability of the task-relevant components $x_t^{(1:3)}$, corresponding to the joint angles directly penalized by the task cost. **(b)** Decodability of the task-irrelevant components $x_t^{(4:6)}$, corresponding to joint velocities, which are only weakly penalized by the cost. While model-matched control progressively loses decodability as internal noise increases, model-free control maintains high decodability for both task-relevant and task-irrelevant components. **(c–d)** Representative single-trial trajectories of one task-irrelevant component, $x_t^{(6)}$, and its linear reconstruction $\hat{x}_t^{\mathrm{dec}}$ from z_t, for $\sigma_\eta = 0.13$. **(c)** In model-matched control, the decoded trajectory only partially tracks the true state component. **(d)** In model-free control, the same component remains accurately decodable from the internal state, showing that flexible internal dynamics preserve behaviorally useful information even when they are not constrained to match the external dynamics.

Here, we further decompose this decodability into task-relevant and task-irrelevant components of the state. In our example, task-relevant components correspond to the joint angles, which are strongly penalized by the task cost and directly observed. Task-irrelevant components correspond to angular

velocities, which are only weakly penalized. Model-free control preserves high decodability for both subspaces across internal-noise levels (Figs. 5a,b). Decodability is slightly lower for task-irrelevant components than for task-relevant ones, but remains stable and comparable across the tested noise range. By contrast, internal states in model-matched control show a stronger degradation of decodability as internal noise increases, especially for task-irrelevant velocity components (Figs. 5a,b). This difference is also visible in representative single-trial trajectories: the decoded model-matched trajectory only partially tracks the true task-irrelevant component, whereas the model-free decoder more accurately reconstructs the same state component from z_t (Figs. 5c,d). Such flexible encoding supports robust behavior and reflects the more complex computational strategies implemented by model-free control, as further discussed in Sec. A.4.6.

A.4.6 Mechanisms Underlying Robustness in Model-Free Control

Both model-matched and model-free control exploit a common strategy: they reshape internal dynamics so that internally generated fluctuations are increasingly confined to directions with weak behavioral impact. To quantify this effect, we decomposed the internal dynamics relative to the readout-visible subspace of the latent state,

$$\mathcal{R}_t = \mathrm{Im}(L_t^\top),$$

and its orthogonal complement, which we refer to as the readout-null subspace. Let $P_{\mathcal{R}}(t)$ denote the orthogonal projector onto $\mathcal{R}_t$ and $P_{\mathcal{N}}(t) = I - P_{\mathcal{R}}(t)$. Because W_t propagates the internal state from time t to time $t+1$, each block should be defined using the readout subspace before the transition and the corresponding readout subspace after the transition, obtaining

$$W_t = P_{\mathcal{R}}(t+1)W_tP_{\mathcal{R}}(t) + P_{\mathcal{R}}(t+1)W_tP_{\mathcal{N}}(t) + P_{\mathcal{N}}(t+1)W_tP_{\mathcal{R}}(t) + P_{\mathcal{N}}(t+1)W_tP_{\mathcal{N}}(t)\ .$$

We then quantified the relative strength of each block by

$$s_{\alpha\leftarrow\beta}(t) = \frac{|P_\alpha(t+1)W_tP_\beta(t)|_F^2}{|W_t|_F^2}, \qquad \alpha,\beta \in \{\mathcal{R},\mathcal{N}\}. \tag{67}$$

This time-dependent decomposition measures how activity that is readout-visible or readout-null at time t is routed by the recurrent dynamics into readout-visible or readout-null directions at the next time step.

As internal noise increases, both approaches show stronger routing from the readout-visible subspace to the readout-null subspace, suggesting that the readout-null subspace acts as a buffer for internal fluctuations (Figs. 6a,b). Consistently, the fraction of internal variability contained in the readout-visible subspace decreases with internal noise, whereas covariance in the readout-null subspace increases (Figs. 6c,d). Thus, large internal variability is preferentially expressed in directions that have little behavioral impact. In line with this interpretation, variability along the readout-visible subspace is smaller than variability along random subspaces of the same dimension, indicating readout-specific noise compression (Fig. 6e). This mechanism is reminiscent of recent analyses of recurrent neural networks showing that internal dynamics can be geometrically decoupled from the output directions, allowing large internal activity to coexist with suppressed readout variability [70].

This routing mechanism is complemented by a reduction in effective readout strength, which further filters internal fluctuations before they affect behavior. In both approaches, increasing internal noise reduces the volume of the readout map, but this reduction is substantially stronger in model-free control (Fig. 6f). At the same time, the sensory-to-internal-state input channel is increasingly reweighted, and this modulation is also stronger in model-free control (Fig. 6g). This is possible because, in model-free control, input (P_t), recurrent (W_t), and output (L_t) channels can be optimized independently: changes in P_t and L_t do not directly impose corresponding changes on W_t. As a result, sensory update gains and output gains can be modulated independently of the recurrent dynamics. This allows the system to maintain useful internal computations while controlling how sensory information enters the latent state and how internal variability is expressed in behavior.

The two approaches differ also in the geometry of the readout-visible subspace. In model-matched control, the internal variable is constrained to behave as a state estimate in the physical coordinate system of the external world. As internal noise increases, the readout-visible subspace therefore becomes increasingly aligned with the task-relevant physical coordinates, namely the joint-angle subspace (angular velocities are weakly penalized by the cost Q_t and not directly observable).

In model-free control, the internal dynamics are freely optimized and need not reproduce the external forward dynamics. The internal state z_t is therefore not constrained to represent x_t in physical coordinates, but can instead encode task-relevant information in a transformed basis optimized jointly for control and robustness. Accordingly, the readout-visible subspace in model-free control does not collapse onto the physical joint-angle subspace (Fig. 6h). Rather, it is closer to the subspace defined by the optimal linear decoder of joint angles from z_t, indicating that task-relevant information is represented in non-physical latent coordinates (Fig. 6i).

The decoder-defined angle subspace still does not fully account for the readout-visible subspace in model-free control: their distance remains bounded away from zero (Fig. 6i). Thus, the model-free control readout is not simply an estimation readout expressed in a rotated basis. Instead, it appears to define a control-oriented subspace, shaped jointly by task information, recurrent dynamics, and the need to shield behavior from internal fluctuations.

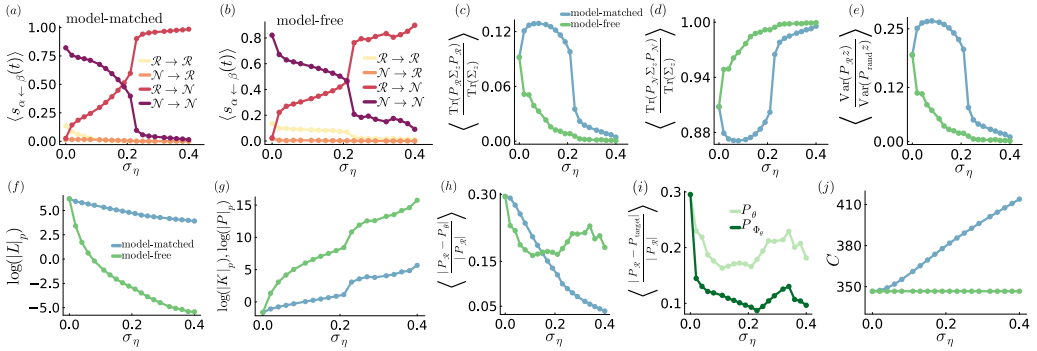


Figure 6: *Readout Geometry and Internal-Noise Routing in the Redundant Arm Task.* **(a,b)** Routing of the internal dynamics between the readout-visible subspace $\mathcal{R}$ and its null subspace $\mathcal{N}$ for model-matched control and model-free control, respectively. Each curve shows the time-averaged fraction $s_{\alpha\leftarrow\beta}(t)$ of the internal update mapping activity from subspace β at time t to subspace α at time $t+1$, normalized by the total squared magnitude of the internal dynamics (Eq. 67). **(c,d)** Fraction of internal state variance lying in the readout-visible subspace, $\mathrm{Tr}(P_{\mathcal{R}}\Sigma_z P_{\mathcal{R}})/\mathrm{Tr}(\Sigma_z)$, and in the null subspace, $\mathrm{Tr}(P_{\mathcal{N}}\Sigma_z P_{\mathcal{N}})/\mathrm{Tr}(\Sigma_z)$, averaged over time. **(e)** Readout-visible variance relative to random subspaces of the same dimensionality, $\mathrm{Var}(P_{\mathcal{R}}z)/\mathrm{Var}(P_{\mathrm{rand}}z)$, averaged over time and random projections. **(f,g)** Time-averaged log-pseudodeterminant of the control readout L_t and sensory/input gain K_t (for model-matched control) and P_t (for model-free control), computed from singular values above numerical threshold. **(h)** Normalized distance between the readout-visible projector $P_{\mathcal{R}}$ and the task-relevant physical-angle projector P_θ (see Methods below for the definition of $P_{\mathcal{R}}$ and P_θ). **(i)** For model-free control, comparison between the distance of $P_{\mathcal{R}}$ from P_θ and from P_{Φ_θ}, the projector onto the row space of the optimal linear decoder from z_t to the task-relevant state components (see Methods below for the definition of P_{Φ_θ}). **(j)** Expected cost C for model-matched and model-free control as a function of internal noise σ_η. All geometric quantities are averaged over the analysis window $t = 20, \ldots, T-2$.

Overall, these results suggest that model-free control leads to more robust and effective control (Fig. 6j) by independently optimizing internal dynamics, input channels, and output channels. This flexibility allows task-relevant variables to be encoded in non-physical coordinates, giving rise to internal representations suited to robust control rather than to estimation alone. Consequently, the optimized internal model is not simply a copy of the external forward dynamics, but a task-specific dynamical system shaped by the joint demands of sensory processing, control, and robustness to internal noise.

Methods For each optimized solution, we compute the readout-visible subspace from the singular value decomposition $L_t = U_t \Sigma_t^L V_t^\top$, denoting by $\sigma_i^L(t)$ the singular values of L_t. We define $P_{\mathcal{R}}(t) = V_{\mathcal{R}}(t) V_{\mathcal{R}}(t)^\top$, where $V_{\mathcal{R}}(t)$ contains the right singular vectors satisfying $\sigma_i^L(t) > \epsilon_{\mathrm{readout}} \max_j \sigma_j^L(t)$, with $\epsilon_{\mathrm{readout}} = 10^{-3}$. The effective readout-null projector is then

$P_{\mathcal{N}}(t) = I - P_{\mathcal{R}}(t)$. Recurrent routing is quantified by the normalized block strengths

$$s_{\alpha\leftarrow\beta}(t) = \frac{|P_\alpha(t+1)W_t P_\beta(t)|_F^2}{|W_t|_F^2}, \qquad \alpha,\beta \in \{\mathcal{R},\mathcal{N}\},$$

and averaged over time.

To quantify variability in each subspace, we propagate the first and second moments of the optimized dynamics and extract the internal covariance $\Sigma_z(t)$. We then compute the fractions of covariance in the readout-visible and readout-null subspaces as

$$\frac{\mathrm{Tr}(P_{\mathcal{R}}(t)\Sigma_z(t)P_{\mathcal{R}}(t))}{\mathrm{Tr}(\Sigma_z(t))}, \qquad \frac{\mathrm{Tr}(P_{\mathcal{N}}(t)\Sigma_z(t)P_{\mathcal{N}}(t))}{\mathrm{Tr}(\Sigma_z(t))}.$$

Noise compression is measured by comparing the variance in the readout-visible subspace with the variance in random subspaces of the same rank. In our simulations, the effective rank of $P_{\mathcal{R}}(t)$ is consistently equal to three across noise levels. We therefore compute

$$\frac{\mathrm{Tr}(P_{\mathcal{R}}(t)\Sigma_z(t)P_{\mathcal{R}}(t))}{\mathrm{Tr}(P_{\mathrm{rand}}(t)\Sigma_z(t)P_{\mathrm{rand}}(t))},$$

where $P_{\mathrm{rand}}(t)$ is the orthogonal projector onto a random three-dimensional subspace of the six-dimensional latent space. To generate each random subspace, we first sample a matrix $X \in \mathbb{R}^{6\times 3}$ with independent Gaussian entries,

$$X_{ij} \sim \mathcal{N}(0,1).$$

We then compute its QR decomposition,

$$X = QR,$$

where the columns of Q are orthonormal and R is upper triangular. The first three columns of Q, denoted $Q_{\mathrm{rand}} \in \mathbb{R}^{6\times 3}$, define an orthonormal basis for a random three-dimensional subspace. The corresponding orthogonal projector is

$$P_{\mathrm{rand}} = Q_{\mathrm{rand}} Q_{\mathrm{rand}}^\top.$$

For each time point, noise level, and optimized solution, we sample 100 independent random projectors. We compute the ratio between readout-visible variance and random-subspace variance for each random projector, average these ratios over random projectors at fixed time, and then average the resulting quantity over time. Values below one indicate that less variability lies in readout-visible directions than expected for random subspaces of equal dimension.

We quantify the effective volume of the output and input channels by computing $\log \mathrm{pdet}(L_t)$ and $\log \mathrm{pdet}(K_t)$. For a generic matrix $A_t \in \{L_t, K_t\}$, we denote its singular values by $\sigma_i^A(t)$ and compute

$$\log \mathrm{pdet}(A_t) = \sum_{i\in\mathcal{H}_t} \log \sigma_i^A(t).$$

with

$$\mathcal{H}_t = \left\{ i : \sigma_i^A(t) > \epsilon_{\mathrm{pdet}} \sigma_{\max}^A(t) \right\}, \qquad \sigma_{\max}^A(t) = \max_j \sigma_j^A(t), \qquad \epsilon_{\mathrm{pdet}} = 10^{-10}.$$

Here, $\sigma_i^A(t)$ denotes the i-th singular value of A_t, and the threshold excludes numerically negligible singular directions.

For the geometric comparisons, we use the same angle decoder defined in the linear decoding analysis. The physical joint-angle subspace is defined directly in latent coordinates as the span of the first three axes,

$$P_\theta = \begin{pmatrix} I_3 & 0 \\ 0 & 0 \end{pmatrix},$$

corresponding to the angle components of $x_t = (\theta_t, \dot\theta_t)$. We then define the decoder-angle subspace as the row space of the angle decoder $\Phi_t^{\mathcal{I}}$, with $\mathcal{I} = \{1,2,3\}$. Specifically, from the singular value decomposition

$$\Phi_t^{\mathcal{I}} = U_t \Sigma_t V_t^\top,$$

we retain the right singular vectors associated with singular values above the threshold $\epsilon_{\text{decoder}} = 10^{-8}$ and define

$$P_{\Phi_\theta}(t) = V_{\Phi_\theta}(t) V_{\Phi_\theta}(t)^\top.$$

Thus, P_θ represents the physical joint-angle coordinate subspace, whereas $P_{\Phi_\theta}(t)$ represents the latent subspace from which joint angles are optimally decoded. We compare the readout-visible subspace with both reference subspaces using normalized Frobenius distances,

$$\frac{|P_{\mathcal{R}}(t) - P_\theta|_F}{|P_{\mathcal{R}}(t)|_F}, \qquad \frac{|P_{\mathcal{R}}(t) - P_{\Phi_\theta}(t)|_F}{|P_{\mathcal{R}}(t)|_F},$$

and average these quantities over time.

A.4.7 Comparison with Previous Optimization Methods

To further assess the methodological contribution of our approach, we compared model-free control with the analytical algorithm of [24] and the gradient-based method of [25]. We considered the redundant-arm task with a shorter horizon ($T = 100$) and internal noise $\sigma_\eta = 0.4$. The corresponding expected task costs were 146.2, 105.3, and 94.8, respectively, with model-free control achieving the lowest cost. Unlike the previous methods, which optimize model-matched controllers, our approach allows the recurrent internal dynamics to be optimized independently, providing additional flexibility for reducing task cost.

Our framework complements previous learning-based approaches to optimal control [71, 72] by providing an analytically derived optimization method for internal dynamics and control under multiplicative and internal noise.

A.4.8 Behavioral Signatures of Model-Matched and Model-Free Control

Beyond differences in expected task cost, an important question is whether model-matched and model-free control can produce distinguishable behavioral signatures that could be measured experimentally. To investigate this, we simulated 1000 trials per condition in the redundant-arm task and quantified joint-angle and angular-velocity variability, mean absolute peak velocity, and motor-command statistics across internal-noise levels. All measures were evaluated over $t = 51, \ldots, 250$, excluding initial and final transients.

Without internal-activity regularization ($\lambda = 0$), the two approaches produce essentially identical behavior at zero internal noise. As σ_η increases, however, model-free control maintains approximately constant behavioral variability, consistent with the internal-noise attenuation mechanism discussed above. In contrast, model-matched control exhibits increasing joint-angle variability, with standard deviation rising from 31.2° to 35.4°, becoming 13.4% larger than under model-free control at $\sigma_\eta = 0.40$. Differences also emerge in peak velocity, which varies non-monotonically under model-matched control, initially increasing and then decreasing with internal noise, while remaining approximately constant under model-free control. Differences in angular-velocity variability and motor-command amplitudes are more modest. Thus, the two approaches produce distinct relationships between internal-noise magnitude and movement kinematics.

This distinction persists when both approaches are optimized with internal-activity regularization ($\lambda = 10^{-3}$). As σ_η increases from 0 to 0.40, joint-angle standard deviation increases from 31.10° to 35.46° under model-matched control, compared with 31.08° to 31.98° under model-free control. At the highest noise level, model-matched joint-angle variability is therefore 10.9% larger. Although regularization causes peak velocity to increase with internal noise under model-free control, this dependence remains monotonic, in contrast to the non-monotonic trend observed under model-matched control. Thus, the two approaches retain qualitatively distinct behavioral signatures even when internal activity is penalized.

Together, these results provide proof-of-principle examples showing that model-matched and model-free control can produce distinguishable behavioral signatures. Importantly, these simulations are intended to illustrate potential qualitative differences between the two approaches, rather than provide quantitative predictions for specific experimental settings, as the model parameters were not calibrated to experimental data. Nevertheless, they suggest that movement variability and other kinematic statistics could help distinguish the two architectures experimentally, potentially through inverse optimal control. Additional examples of distinguishable responses to internal state perturbations are presented in Appendix A.4.9.

A.4.9 Perturbation Responses

To show that model-matched and model-free control not only implement different computational strategies but also lead to qualitatively different behavioral predictions, we perform a perturbation experiment in the redundant arm task, in which the internal state z_t is directly perturbed. Rather than perturbing arbitrary internal directions, we target behaviorally relevant directions, defined as those that can affect the control signal through the optimized readout. To do so, we inject perturbations in the readout-visible subspace of the internal state, defined from the singular value decomposition of the readout matrix L_t. Let $L_t = U_t \Sigma_t^L V_t^\top$ be the singular value decomposition of the control readout at time t. The right singular vectors associated with nonzero singular values span the subspace of internal states that is visible to the control output $u_t = L_t z_t$. We denote these vectors by $V_{\mathcal{R}}(t)$ and define the corresponding readout-visible projector as $P_{\mathcal{R}}(t) = V_{\mathcal{R}}(t) V_{\mathcal{R}}(t)^\top$. This projector isolates the components of z_t that can influence the control signal. Perturbing this subspace therefore provides a direct test of how each optimized architecture responds to internal fluctuations that are potentially behaviorally consequential.

For each internal-noise level σ_η, for each approach $s \in \{\mathrm{model-matched}, \mathrm{model-free}\}$, and for each perturbation time t_p, we generate random perturbation directions in the readout-visible subspace. To do so, we sample a random vector

$$\zeta \sim \mathcal{N}(0, I_m), \tag{68}$$

where m is the dimension of the internal state z_t and I_m is the $m \times m$ identity matrix. We then project this vector onto the readout-visible subspace:

$$\delta z(t_p) = P_{\mathcal{R}}^s(t_p)\zeta. \tag{69}$$

This produces a random direction in the subspace of internal activity that can affect the control readout. We normalize this projected vector to obtain the injected perturbation

$$\delta z_{\mathrm{inj}}(t_p) = \gamma_s(t_p) \frac{\delta z(t_p)}{|\delta z(t_p)|}, \tag{70}$$

where $\gamma_s(t_p)$ is a scaling factor specified below. By construction, $\delta z_{\mathrm{inj}}(t_p) \in \mathrm{row}(L_{t_p}^s)$, and therefore the perturbation directly affects the control signal through $\delta u_{t_p} = L_{t_p}^s \delta z_{\mathrm{inj}}(t_p)$.

Because model-matched and model-free control can produce internal activity with very different magnitudes and geometries, we scale the perturbation separately for each approach and perturbation time to obtain a fair comparison. Specifically, we set

$$\gamma_s(t_p) = \alpha\, \mathbb{E}\left[|P_{\mathcal{R}}^s(t_p) z_{t_p}^{s,\mathrm{base}}| \right], \tag{71}$$

with $\alpha = 1.0$, to ensure a significant effect of the perturbation (Figs. 7a-d). Here, $z_{t_p}^{s,\mathrm{base}}$ denotes the unperturbed internal state. Thus, the perturbation has the same relative magnitude with respect to the ongoing readout-visible activity of each solution, rather than the same absolute magnitude in internal coordinates. This is important because internal activity can differ substantially between model-matched and model-free control. We then verify that this normalization leads to comparable readout-level effects by measuring the relative control perturbation $\mathbb{E}[|\Delta u_t|]/\mathbb{E}[|u_t^{\mathrm{base}}|]$, where $\Delta u_t = u_t^{\mathrm{pert}} - u_t^{\mathrm{base}}$, and the expectation is taken over trials and perturbation directions (Fig. 7e).

Perturbations are applied during a finite stimulation window, $t_p \in \{100, \ldots, 130\}$, in a task of duration $T = 300$ time steps with $\Delta t = 0.01$. At each perturbation time, the perturbation is added before computing the control signal:

$$z_t^{\mathrm{pert}} \leftarrow z_t^{\mathrm{pert}} + \delta z_{\mathrm{inj}}(t). \tag{72}$$

For each internal-noise level, we simulate 100 random perturbation directions and 500 trials per direction. Baseline and perturbed trajectories use common random numbers: the same initial condition and the same realizations of process noise, sensory noise, motor noise, and internal noise. Therefore, differences between perturbed and baseline trajectories isolate the causal effect of the injected perturbation.

Across the time points and internal-noise levels analyzed here, both model-matched and model-free control have a three-dimensional readout-visible subspace. However, the orientation of this subspace

differs between the two approaches. During the stimulation window, the readout-visible subspace is also stable both in rank and in orientation: the rank remains equal to three, and the projector $P_{\mathcal{R}}(t)$ changes only weakly over time. This means that the effective perturbation target does not require substantial retuning during the stimulation period. In an experimental setting, this stability would make the perturbation more feasible, since a stimulation pattern designed to affect the readout-visible subspace at the beginning of the stimulation window would remain approximately aligned with that subspace over the perturbation interval.

This approach is reminiscent of recent causal perturbation studies in motor control, where optogenetic stimulation and intracortical microstimulation have been used to test how neural perturbations interact with task-relevant population dynamics during reaching [73]. Our perturbation experiment is not intended to model the biophysical details of a specific stimulation protocol. Instead, it formalizes a targeted perturbation that is informative for distinguishing the two optimized architectures: perturbations aligned with the readout-visible subspace are expected to be behaviorally consequential, whereas perturbations in readout-null directions should have weaker immediate effects on the control signal.

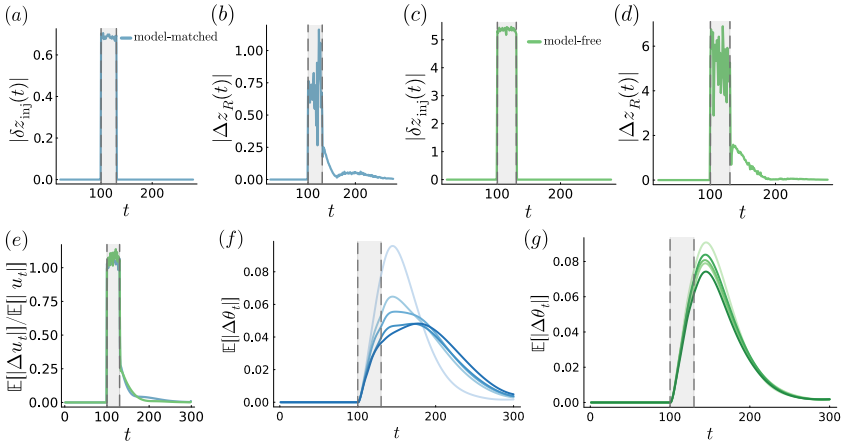


Figure 7: *Readout-Visible Internal Perturbations in the Redundant Arm Task.* **(a)** Example injected perturbation magnitude $|\delta z_{\mathrm{inj}}(t)|$ for model-matched control, shown for one perturbation direction and one trial at $\sigma_\eta = 0.09$. The gray shaded region indicates the stimulation window, $t \in \{100, \ldots, 130\}$, and dashed lines mark its onset and offset. **(b)** Resulting readout-visible internal deviation for the same model-matched trial, $|\Delta z_R(t)| = |P_{\mathcal{R}}(t)(z_t^{\mathrm{pert}} - z_t^{\mathrm{base}})|$, computed by subtracting a baseline (i.e., unperturbed) trajectory with matched noise. **(c,d)** Same as (a,b), for model-free control. **(e)** Relative control deviation, $\mathbb{E}[|\Delta u_t|]/\mathbb{E}[|u_t^{\mathrm{base}}|]$, where u_t^{base} denotes the control signal in the unperturbed trajectory, shown at $\sigma_\eta = 0.09$ for model-matched and model-free control. This confirms that the perturbation produces comparable readout-level effects on the control signal in the two approaches. **(f)** Behavioral effect of the perturbation in model-matched control, measured as the absolute deviation in the task-relevant joint angles, $\mathbb{E}[|\Delta\theta_t|]$, across internal-noise levels $\sigma_\eta \in \{0, 0.03, 0.06, 0.09, 0.13\}$. Darker colors indicate higher internal-noise levels. **(g)** Same as (f), for model-free control. Expectations in (e–g) are computed over 100 perturbation directions and 500 trials per direction.

For each trial and perturbation direction, we compute the readout-visible internal deviation $\Delta z_{\mathcal{R},t} = P_{\mathcal{R}}(t)\left(z_t^{\mathrm{pert}} - z_t^{\mathrm{base}}\right)$, the control deviation $\Delta u_t = L_t z_t^{\mathrm{pert}} - L_t z_t^{\mathrm{base}}$, and the behavioral deviation in the task-relevant variables (only joint angles are strongly penalized through the state-cost matrix Q_t, and only joint angles are directly observable), $\Delta\theta_t = \theta_t^{\mathrm{pert}} - \theta_t^{\mathrm{base}}$, where θ_t denotes the three joint-angle components of the six-dimensional arm state.

The perturbation responses reveal qualitative differences between the two architectures. In model-matched control, increasing internal noise substantially changes the behavioral response to perturbations in the readout-visible subspace: the induced deviation in the task-relevant variables, $\Delta\theta_t$, persists for longer transients, indicating slower recovery from internal perturbations as internal noise increases (Fig. 7f).

In contrast, in model-free control, the temporal profile of $\Delta\theta_t$ remains much closer to the zero-internal-noise response across noise levels, with faster decay of the perturbation-induced transient than in model-matched control. Since model-matched and model-free control solutions coincide in the zero-internal-noise limit, this indicates that model-free control preserves a more noise-invariant perturbation-response profile as internal noise increases (Fig. 7g), consistent with its more flexible internal representation and with the robust behavior shown in Fig. 6j.

A.4.10 Control Under Multiplicative Internal Noise

To investigate the robustness of our results to signal-dependent internal noise, we repeated the three-dimensional reaching and redundant-arm tasks using the internal state-dependent multiplicative noise already included in the general formulation of Eq. 8. Specifically, we considered a single multiplicative noise source with $Z_1^z = \sigma_\nu I_n$ and $\nu_{t,1}^z = \nu_t$, where $\nu_t \sim \mathcal{N}(0,1)$ and $\sigma_\nu \geq 0$ controls the noise magnitude. All other Z_j^c were set to zero. The same noise configuration was used for both model-matched and model-free control.

All task, plant, observation, cost, and other noise parameters were kept unchanged. In particular, additive internal noise was fixed at $\sigma_\eta = 0.2$ in the reaching task and $\sigma_\eta = 0.05$ in the redundant-arm task, while σ_ν was varied over ten values between 0 and 0.7. No internal-activity regularization was applied ($\lambda = 0$) in either approach, and performance was evaluated using the original task cost.

Across all tested multiplicative-noise levels, model-free control outperformed model-matched control in both tasks. The relative cost improvement increased from 2.1% to 16.6% in the reaching task and from approximately 1% to 9.8% in the redundant-arm task. Since multiplicative noise scales with internal state activity, these results provide further evidence that the model-free advantage does not rely solely on suppressing additive internal noise through internal state amplification.

A.4.11 Generalization Across Tasks and Noise Conditions

To investigate whether the advantage of model-free control extends beyond the conditions under which it is optimized, we performed four additional experiments using the redundant-arm model. We tested generalization across task objectives, reward magnitudes, and target locations by keeping the learned internal architecture fixed and reoptimizing only the control gain L_t. Specifically, the sensory gain P_t and recurrent dynamics W_t remained fixed in model-free control, while in model-matched control the estimator gain K_t remained fixed and the recurrent dynamics continued to satisfy $W_t = A + BL_t - K_tH$. We also examined robustness to unexpected changes in internal noise without reoptimizing any controller parameters. The experiments in Fig. 8 were performed without internal-activity regularization ($\lambda = 0$) for both approaches. An additional comparison, summarized below, examined how regularization with $\lambda = 10^{-3}$ affects model-free generalization.

Generalization Across Task Objectives. We first optimized both approaches on the original redundant-arm stabilization task, which penalizes deviations from the target posture throughout the movement. We then transferred the solutions to a finite-horizon reaching task, in which only the final state is penalized. Specifically, the second joint was initially displaced by $0.30\,\mathrm{rad}$ from the target, the running state penalties were removed ($Q_t = 0$ for $t = 0, \ldots, T-1$), and the terminal penalty was set to $Q_T = \mathrm{diag}(10, 10, 10, 1, 1, 1)$. All other parameters, including the control cost and time horizon, were unchanged. After reoptimizing only L_t, model-free control achieved lower task costs than model-matched control, with relative improvements of 1.5%, 2.5%, 5.6%, and 7.8% at $\sigma_\eta = 0.10$, 0.15, 0.30, and 0.40, respectively (Fig. 8a). Thus, the internal architecture learned by model-free control can support a qualitatively different task objective through readout-only adaptation.

Generalization Across Reward Magnitudes. We next tested whether the learned internal architecture could adapt to different terminal-reward magnitudes. Both approaches were first fully optimized on the terminal-reaching task described above, with $\sigma_\eta = 0.10$. We then varied the terminal penalty according to $Q_T(\alpha) = \mathrm{diag}(10\alpha, 10\alpha, 10\alpha, 1, 1, 1)$, with $\alpha \in \{1, 2, 4, 8\}$, while leaving all other parameters unchanged. Reoptimizing only L_t, model-free control retained its advantage at every reward magnitude, with relative task-cost improvements of 1.4%, 1.9%, 2.3%, and 2.5%, respectively (Fig. 8b). These results indicate that the same internal architecture can be reused as the relative importance of endpoint accuracy changes.

Generalization Across Target Locations. We also examined transfer to different static targets. Since the redundant-arm task is expressed in target-centered coordinates, changing the target while keeping the initial posture fixed is equivalent to changing the initial target-relative displacement. For each internal-noise level, both approaches were first fully optimized for zero initial displacement and subsequently tested on five displacements of the second joint, $x_2^{\mathrm{init}} \in \{-0.30, -0.15, 0.10, 0.20, 0.30\}\,\mathrm{rad}$. We reoptimized only L_t and compared the resulting costs with those obtained by fully reoptimizing each approach for the new target. Across the complete internal-noise sweep, readout-only adaptation increased the mean task cost by less than 0.067% for model-matched control and 0.013% for model-free control, relative to full reoptimization (Fig. 8c). Moreover, model-free control retained a lower cost at every positive internal-noise level, with its relative advantage increasing from 0.70% at $\sigma_\eta = 0.04$ to 16.23% at $\sigma_\eta = 0.40$. Thus, both approaches generalize effectively across target locations, without a loss of transferability for model-free control.

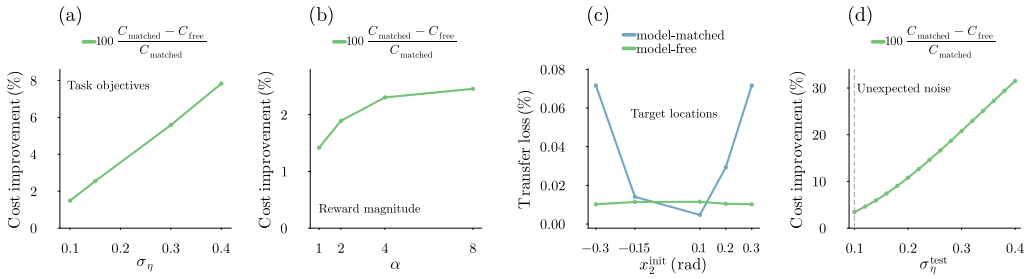


Figure 8: *Model-Free Control Generalizes Across Tasks and Noise Conditions.* All experiments were performed in the redundant-arm task without internal-activity regularization ($\lambda = 0$). **(a)** Generalization from stabilization to terminal-only reaching, with an initial displacement of $0.30\,\mathrm{rad}$ in the second joint. After optimizing both approaches on the original stabilization task, only the control gain L_t was reoptimized for the new objective. The panel shows the relative task-cost improvement of model-free over model-matched control, $100(C_{\mathrm{matched}} - C_{\mathrm{free}})/C_{\mathrm{matched}}$, as a function of internal-noise magnitude σ_η. **(b)** Generalization across terminal-reward magnitudes. Controllers were first fully optimized for terminal reaching at $\sigma_\eta = 0.10$ and $\alpha = 1$, then adapted by reoptimizing only L_t as the terminal angle penalty was scaled by $\alpha \in \{1, 2, 4, 8\}$. The panel shows the relative task-cost improvement of model-free over model-matched control, $100(C_{\mathrm{matched}} - C_{\mathrm{free}})/C_{\mathrm{matched}}$, as a function of α. **(c)** Generalization across target locations. Controllers optimized for zero initial displacement were transferred to five target-relative displacements of the second joint, keeping the learned internal architecture fixed and reoptimizing only L_t. The panel shows the relative transfer loss, $100(C_{\mathrm{transfer}} - C_{\mathrm{full}})/C_{\mathrm{full}}$, for each approach, where C_{transfer} is the cost obtained through readout-only adaptation and C_{full} is the cost obtained by fully reoptimizing all method-specific parameters for each new target. For each displacement, the relative loss is averaged across nine positive internal-noise levels, $\sigma_\eta \in \{0.04, 0.09, 0.13, 0.18, 0.22, 0.27, 0.31, 0.36, 0.40\}$. **(d)** Robustness to unexpected internal noise. Both controllers were optimized at $\sigma_\eta^{\mathrm{train}} = 0.10$ (vertical dashed line) and evaluated at different internal-noise magnitudes $\sigma_\eta^{\mathrm{test}}$, without reoptimizing any parameters. The panel shows the relative task-cost improvement of model-free over model-matched control, $100(C_{\mathrm{matched}} - C_{\mathrm{free}})/C_{\mathrm{matched}}$, as a function of $\sigma_\eta^{\mathrm{test}}$. In panels (a–c), readout-only adaptation keeps P_t and W_t fixed for model-free control, while K_t remains fixed for model-matched control.

Robustness to Unexpected Internal Noise. Finally, we examined whether the optimized controllers remain effective when the actual internal-noise magnitude differs from that assumed during optimization. Both approaches were optimized at $\sigma_\eta^{\mathrm{train}} = 0.10$ and subsequently evaluated, without any reoptimization, at $\sigma_\eta^{\mathrm{test}} \in \{0.10, 0.12, \ldots, 0.40\}$. Model-free control consistently outperformed model-matched control throughout this range, with relative task-cost improvements of 3.46%, 10.80%, 20.80%, and 31.52% at $\sigma_\eta^{\mathrm{test}} = 0.10$, 0.20, 0.30, and 0.40, respectively (Fig. 8d). These results indicate that the model-free advantage persists even when internal noise is substantially greater than anticipated during optimization.

Generalization was also retained with internal-activity regularization ($\lambda = 10^{-3}$), although performance deteriorated, consistent with penalizing the internal state scaling mechanism discussed above.

Compared with model-free control at $\lambda = 0$, and expressing all percentage increases relative to its task cost at the same test condition, regularized model-free control incurred task-cost increases of 2.8–10.0% across task objectives, 2.2–4.9% across reward magnitudes, 0.4–4.1% across target locations, and 1.0–14.7% under unexpected internal noise. These comparisons support retained transfer capabilities under regularization, with a performance cost that is most pronounced under unexpected noise.

Together, these experiments show that the internal architectures learned by model-free control can be reused across different task objectives, rewards, and targets, while remaining robust to changes in internal noise. Thus, in the unregularized setting, the improved performance of model-free control did not come at the expense of generalization within the conditions examined here, while regularized solutions retained transfer capabilities with reduced performance.

A.4.12 Neural Population Steering with Model-Free Control

The flexibility of the model-free framework allows stochastic optimal control to be directly applied to a broader class of problems than those covered by classical formulations [24]. In particular, the internal variable is not required to behave as a Kalman-filter estimate of the controlled state, and its dynamics need not match the dynamics of the external system. This makes it possible to formulate control problems in which the controlled process and the internal process correspond to two distinct interacting neural populations, each with its own recurrent structure.

In classical approaches, the internal variable z_t is constrained to evolve as an estimate of the controlled state x_t. This is achieved by imposing the matched-dynamics structure $W_t = A + BL_t - P_tH$, so that the dynamics of z_t are tied to the external dynamics of x_t and to the current control and feedback gains. By contrast, model-free control removes this structural constraint. The internal dynamics can be chosen or optimized independently of the external dynamics, allowing z_t and x_t to represent distinct neural populations with independent recurrent connectivity matrices. Moreover, the framework naturally supports partial optimization: for example, the recurrent connectivity of the internal population and the projection from x_t to z_t can be kept fixed, while only the readout from z_t to the control input is optimized. Such configurations are not naturally compatible with the matched formulation, where the internal connectivity is forced to depend on A, L_t, and P_t.

To illustrate this point, we consider a neural population steering task in which one recurrent linear population is driven toward a target state through optimized readouts from another recurrent linear population, Fig. 9a. This setting is related in spirit to reservoir computing and recurrent network approaches in which high-dimensional neural dynamics are linearly read out to generate task-relevant signals [74, 75, 76]. Here, however, the readout is obtained through the stochastic optimal control objective, which makes it possible to optimize control directly in the presence of noise and explicit control costs.

Although this task is not intended as a biologically realistic circuit model, it can be qualitatively reminiscent of thalamo-cortical control loops, in which an upstream controller population helps initialize or reshape downstream motor population dynamics before movement initiation [77, 78]. In this interpretation, z_t plays the role of an internal controller population driving the downstream population x_t.

We consider two linear populations of $N_{\text{units}} = 100$ neurons. The first population, with activity $x_t \in \mathbb{R}^{N_{\text{units}}}$, is the population to be controlled. The second population, with activity $z_t \in \mathbb{R}^{N_{\text{units}}}$, acts as an internal or controller population. Their dynamics are

$$x_{t+1} = Ax_t + BL_tz_t + \xi_t + \epsilon_tCL_tz_t, \tag{73}$$

$$z_{t+1} = W_tz_t + P_tHx_t + P_t\omega_t + \rho_tP_tDx_t + \eta_t. \tag{74}$$

In this task, we fix the internal dynamics and optimize only the readout matrices L_t. More precisely, we set

$$W_t = W, \qquad P_t = P, \qquad t = 0, \dots, T-1, \tag{75}$$

where W is the recurrent connectivity of the z_t population and P is a fixed projection from the x_t population to the z_t population. The matrices L_t are time-dependent and are the only variables optimized by the algorithm. Thus, z_t is not interpreted as an estimate of x_t, but rather as a distinct dynamical population whose activity is linearly read out to generate the control input $u_t = L_tz_t$. We have $m = n = p = k = N_{\text{units}}$, and we set

$$B = I_{N_{\text{units}}}, \qquad H = I_{N_{\text{units}}}. \tag{76}$$

The matrices A, W, and P are sampled independently as dense Gaussian random matrices with the standard $1/\sqrt{N_{\text{units}}}$ scaling used in dynamical mean-field theory of large random recurrent networks [79]. Specifically, we sample

$$A_{ij} \sim \mathcal{N}\left(0, \frac{g_A^2}{N_{\text{units}}}\right), \tag{77}$$

$$W_{ij} \sim \mathcal{N}\left(0, \frac{g_W^2}{N_{\text{units}}}\right), \tag{78}$$

$$P_{ij} \sim \mathcal{N}\left(0, \frac{g_P^2}{N_{\text{units}}}\right). \tag{79}$$

Here, A defines the recurrent dynamics of the controlled population, W defines the recurrent dynamics of the internal population, and P defines the fixed projection from x_t to z_t. We choose $g_A = 1.1$, so that the uncontrolled dynamics of the x_t population are unstable, making active control necessary. The internal population is instead initialized in a stable regime with $g_W = 0.9$, while the coupling from x_t to z_t is weaker, with $g_P = 0.3$. The cost function penalizes deviations of x_t from the origin and the magnitude of the control input

$$C = \sum_{t=0}^{T} \mathbb{E}\left[x_t^\top Q_t x_t + u_t^\top R_t u_t\right]. \tag{80}$$

The target state is therefore the origin, and the cost matrices are chosen as

$$Q_t = q_{<T} I_{N_{\text{units}}}, \qquad t = 0, \ldots, T-1, \tag{81}$$

$$Q_T = q_T I_{N_{\text{units}}}, \tag{82}$$

$$R_t = r I_{N_{\text{units}}}, \qquad t = 0, \ldots, T-1, \tag{83}$$

$$R_T = 0. \tag{84}$$

In the simulations shown here, we use $q_{<T} = 10^{-3}$, $q_T = 1$, and $r = 10^{-3}$.

The noise covariances are

$$\Sigma_\xi = \sigma_\xi^2 I_{N_{\text{units}}}, \tag{85}$$

$$\Sigma_\omega = \sigma_\omega^2 I_{N_{\text{units}}}, \tag{86}$$

$$\Sigma_\eta = \sigma_\eta^2 I_{N_{\text{units}}}, \tag{87}$$

with $\sigma_\xi = 0.5$, $\sigma_\eta = 0.3$, and $\sigma_\omega = 0$. In addition, multiplicative sensory and control noise are set to zero,

$$\sigma_\rho = 0, \qquad \sigma_\epsilon = 0, \tag{88}$$

so that in this example only additive noise in the controlled population and additive internal noise in the internal population are present. We consider deterministic initial conditions that are sampled once from Gaussian distributions and then kept fixed across Monte Carlo trials. Specifically, we draw

$$x_0 \sim \mathcal{N}(0, g_{x_0}^2 I_{N_{\text{units}}}), \tag{89}$$

$$z_0 \sim \mathcal{N}(0, g_{z_0}^2 I_{N_{\text{units}}}), \tag{90}$$

and use the sampled vectors as fixed initial activities for the two populations. Thus, trial-to-trial variability arises from the dynamical and internal noise terms, rather than from variability in the initial condition. Equivalently, in the moment propagation we set the initial covariances to zero, $\Sigma_{x_0} = \Sigma_{z_0} = 0$. We use $g_{x_0} = 10$ and $g_{z_0} = \sigma_\eta = 0.3$. The large value of g_{x_0} initializes the controlled population far from the target or, equivalently, in a coordinate system centered on the target, far from zero. This makes active control necessary. In contrast, z_t is initialized close to zero, representing spontaneous internal activity whose scale is set by the internal noise level. We use a horizon of $T = 50$ time steps. The readouts were optimized for 500 iterations while keeping $P_t = P$ and $W_t = W$ fixed. After optimization, we simulate 1000 Monte Carlo trials to characterize the resulting trajectories.

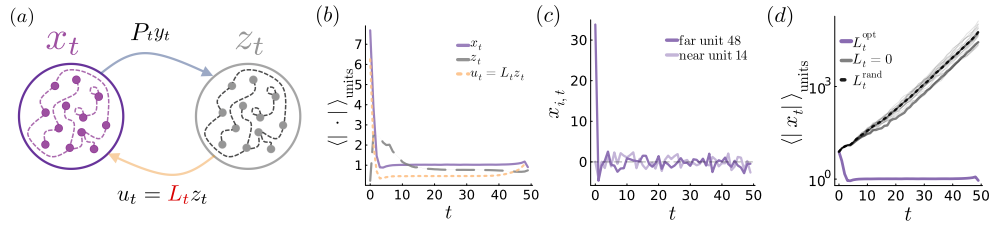


Figure 9: *Neural Population Steering with Model-Free Control.* **(a)** Schematic of the task. The activity of the controlled population x_t is projected to the internal/controller population z_t through a fixed mapping, while the control signal is generated as $u_t = L_t z_t$ and fed back to the controlled population. **(b)** Mean absolute activity of the controlled population x_t, the internal population z_t, and the control signal u_t, averaged over units and Monte Carlo trials. **(c)** Activity of two representative units from the controlled population in a single trial: one unit with large initial absolute activity (far unit) and one unit with initial activity close to zero (near unit). **(d)** Comparison of the optimized solution with two baselines: no control ($L_t = 0$) and random matched-norm readouts (L_t^{rand}). The plotted quantity is the mean absolute activity of the controlled population, $\langle |x_t| \rangle_{\text{units}}$, shown on a logarithmic scale. For the random-readout baseline, thin black traces show individual random readout realizations, while the black dotted line shows the mean across random realizations.

Figs. 9b–c show that the controlled population x_t is effectively steered toward the target through the activity of the internal population z_t, while optimizing only the readout matrices L_t. At the population level, the mean absolute activity of x_t rapidly decreases over time, indicating convergence toward the target state, while z_t and the control signal $u_t = L_t z_t$ remain bounded throughout the trial (Fig. 9b). The same behavior is visible at the single-unit level, where units with different initial displacements are progressively brought toward the target (Fig. 9c).

Importantly, this stabilization does not result simply from injecting a control signal of sufficient magnitude. As shown in Fig. 9d, removing control entirely ($L_t = 0$) leads to an explosive growth of the controlled population activity, consistent with the unstable recurrent dynamics of A. We also compare the optimized solution with random readouts matched in magnitude: at each time step, each random matrix $\tilde{L}_t^{\text{rand}}$ is sampled with independent Gaussian entries, $(\tilde{L}_t^{\text{rand}})_{ij} \sim \mathcal{N}(0, 1)$, and then rescaled as $L_t^{\text{rand}} = |L_t^{\text{opt}}|_F \tilde{L}_t^{\text{rand}} / |\tilde{L}_t^{\text{rand}}|_F$, so that $|L_t^{\text{rand}}|_F = |L_t^{\text{opt}}|_F$. We simulated 20 such random readout sequences, with 300 Monte Carlo trials per sequence. These matched-norm random controls also failed to stabilize the system. Similar results were obtained when the random readout entries were sampled with positive or negative means, rather than zero mean, before Frobenius-norm normalization. Thus, successful steering depends on the structured, time-dependent readout found by the model-free optimization, rather than on control magnitude alone.

For the baseline comparison in Fig. 9d, we generate 20 random readout sequences. At each time step, each random matrix is sampled with independent Gaussian entries and then rescaled to match the Frobenius norm of the optimized readout, $|L_t^{\text{rand}}|_F = |L_t^{\text{opt}}|_F$. The optimized trajectories in Figs. 9b–c are estimated from 1000 Monte Carlo trials. The no-control and random-readout baselines in Fig. 9d are simulated using 300 Monte Carlo trials for each readout, and the random-readout curve reports the mean across the 20 random readouts.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The claims made in the introduction and abstract are supported by the results presented in Section 4.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Limitations are discussed in the Conclusions.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

Justification: The complete set of assumptions and proofs is provided in the main manuscript and in the Appendix.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: All details necessary to reproduce the results are provided in the Appendix.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: The codes to reproduce the results are provided in the Supplementary Material.

Guidelines:

- The answer [N/A] means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: All parameters of the experiments are listed in the Appendix.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: When applicable, error bars represent the standard error of the mean across trials.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.

- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [No]

   Justification: All algorithms can be run on a standard laptop without issue.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: We present a theoretical study aimed at solving stochastic optimal control problems. This work does not have any direct ethical implications.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [N/A]

    Justification: Given its purely theoretical nature, this study is not expected to have any direct societal impact.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.

- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: This study does not pose such a risk.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [N/A]

    Justification: The paper does not use existing assets.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.

- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: The paper does not release new assets.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing and research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: The core method development in this research does not involve LLMs as any important, original, or non-standard components.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.