OPTIMAL CONTROL UNDER MULTIPLICATIVE AND INTERNAL NOISE WITH MODEL MISMATCH

Anonymous authors
Paper under double-blind review

ABSTRACT

Natural agents interact with their environment through noisy and continuous sensorimotor loops. Stochastic optimal control provides a principled framework for this problem, but existing analytical solutions are restricted to linear dynamics with Gaussian observations and additive noise. They cannot address scenarios with multiplicative noise in control or observations, and with internal noise affecting estimation — features central to biological and robotic systems. We provide a provably convergent algorithm that computes fixed-point controller–filter solutions for linear dynamics with quadratic costs under multiplicative and internal noise. Our method overcomes the limitations of prior analytical approaches and improves the efficiency of state-of-the-art gradient-based methods by more than three orders of magnitude in realistic tasks. Importantly, it also optimizes internal dynamics, relaxing the classical assumption that internal models must match external dynamics. Allowing such model mismatch yields substantially better performance under internal noise. In sum, we provide the first full solution to stochastic optimal linear control with multiplicative and internal noise, covering both matched and mismatched internal models.

1 INTRODUCTION

Understanding the computational mechanisms that govern the sensorimotor system in humans and other animals is a long-standing goal in systems and computational neuroscience (Wolpert et al., 1995; Shadmehr & Krakauer, 2008; Franklin & Wolpert, 2011; Todorov, 2004). Yet, developing formal and mathematically tractable models that accurately capture these mechanisms remains an open problem, with far-reaching implications for fields such as artificial intelligence and robotics. In this context, stochastic optimal control theory provides a powerful mathematical framework for explaining behavior in terms of optimality principles, accounting for uncertainty and variability inherent in biological systems (Todorov & Jordan, 2002; Todorov, 2005; Straub & Rothkopf, 2022; Schultheis et al., 2021; Faisal et al., 2008). The pivotal work in Todorov (2005) extended the classic Linear-Quadratic-Additive-Gaussian – LQAG – framework (usually referred to as Linear-Quadratic-Gaussian – LQG – problem (Davis, 2013)) to incorporate a more biologically realistic noise model of the sensorimotor system. This includes control-dependent noise (Schmidt et al., 1979; Todorov, 2002), signal-dependent sensory feedback noise (Todorov & Jordan, 2002; Harris & Wolpert, 1998), and internal neural noise (Faisal et al., 2008; Moreno-Bote et al., 2014; Churchland et al., 2006) — all of which are essential for reproducing key signatures of human motor behavior (Todorov, 2005; Flash & Hogan, 1985; Harris & Wolpert, 1998; Todorov, 2002; Schmidt et al., 1979).

However, explaining behavior through optimal control requires first obtaining optimal solutions to the underlying problem (Todorov, 2005; Schultheis et al., 2021). In this work, we derive an algorithm that fully solves the stochastic control problem of Todorov (2005); our algorithm exploits coordinate descent, and we prove its monotonic improvement and convergence to a critical point (Sec. 3). This overcomes prior analytical limitations and, unlike the state-of-the-art numerical methods, yields an analytically-derived algorithm for the full problem with speedups of more than three orders of magnitude in realistic tasks (see Prior Work below). Our framework thus provides both a conceptual advance and a major efficiency gain over existing approaches.

A further limitation of current theoretical work on stochastic optimal control is the reliance on two assumptions: (1) a strict separation between estimation and control, and (2) the matched-dynamics

assumption, i.e., that the internal model used for estimation and control perfectly matches the dynamics of the external environment. These limitations underlie both Todorov (2005) and Damiani et al. (2024), where noisy sensory feedback is first processed by a Kalman filter to produce a state estimate – based on the same forward model of the environment – which then guides linear control actions. Within the classical LQAG problem, this methodology is mathematically justified by the separation principle (Davis, 2013). However, once multiplicative and internal noise are included, the separation principle no longer holds, making estimation and control inherently coupled (Todorov, 2005). Moreover, the assumption that the agent's internal model exactly matches the external dynamics strongly limits the realism of this approach, overlooking a substantial body of research emphasizing the role of internal models in motor control (Wolpert et al., 1995; Shadmehr et al., 2010; Körding & Wolpert, 2004; Kawato, 1999; Golub et al., 2015).

Our second main contribution is to relax these assumptions by considering the more general case where the internal dynamics – used by the agent to process sensory stimuli and generate motor outputs – need not match the dynamics of the external world and must themselves be optimized (Sec. 4). We refer to the classical case as Model Match (M-Match), and to our extension as Model Mismatch (M-Mis). We extend the algorithm developed for the M-Match case (Sec. 3.2) to this scenario, providing an analytical solution for mismatched internal models. In Sec. 4.1, we demonstrate that this additional flexibility leads to improved performance relative to M-Match, particularly in the presence of internal noise. Finally, we illustrate the generality of our framework by applying it to the steering of linear neural populations, which connects directly to computational principles underlying reservoir computing (Jaeger & Haas, 2004; Maass et al., 2002) and, more broadly, to recurrent neural network models that generate task-relevant outputs (Sussillo & Abbott, 2009).

Prior Work The seminal study of Todorov (2005) provided the first analytically-derived algorithm for optimal linear control under multiplicative and internal noise. However, Damiani et al. (2024) demonstrated that this solution fails to yield truly optimal results in the presence of internal noise, due to the incorrect assumption of unbiased estimators and its connection with the orthogonality principle (Appendix A.1). To address this limitation, Damiani et al. (2024) introduced a numerical gradient-based algorithm that achieves optimal performance, albeit at high computational cost, making it impractical for inverse optimal control applications. They also proposed an analytical counterpart, the FPOMP algorithm, which solves the problem in the one-dimensional case and, in higher dimensions, only under additive noise, due to the increased mathematical complexity of the full setting. Consequently, no previous work provides a general analytical solution or formal convergence guarantees.

2 Stochastic Linear Optimal Control: Problem Formulation

We first review the standard Linear-Quadratic-Additive-Gaussian (LQAG) problem, then extend the noise model, following Todorov (2005), to include multiplicative observation, control noise, and internal noise, yielding the Linear-Quadratic-Multiplicative-Internal (LQMI) formulation. In both LQAG and LQMI, internal and state dynamics are matched; the more general mismatched case is treated in Sec. 4.

2.1 Stochastic Optimal Control under Multiplicative and Internal Noise

In the standard LQAG formulation, an agent receives noisy observations $y_t \in \mathbb{R}^k$ $(t = 0, 1, ..., T)$ from a state variable $x_t \in \mathbb{R}^m$,

$$y_t = Hx_t + \omega_t \,, \tag{1}$$

where $H \in \mathbb{R}^{k\times m}$ is the observation matrix and $\omega_t \in \mathbb{R}^k$ is a zero-mean noise with covariance Σ_ω. The control problem consists in finding the optimal control signal $u_t(y_{t-1}, ..., y_0) \in \mathbb{R}^p$ that steers the stochastic linear dynamical system

$$x_{t+1} = Ax_t + Bu_t + \xi_t \,, \tag{2}$$

so as to minimize the expected cumulative quadratic cost

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

The dynamics of the state variable, Eq. 2, is assumed to be linear in state and control with matrices $A \in \mathbb{R}^{m\times m}$ and $B \in \mathbb{R}^{m\times p}$ and corrupted by zero-mean noise $\xi_t \in \mathbb{R}^m$ with covariance Σ_ξ. All noises are uncorrelated in time and are not required to be Gaussian. We observe that time-dependent matrices in the dynamics or noise can be trivially incorporated. The initial condition of the dynamics is x_0, usually drawn from a Gaussian distribution. The control signal $u_t(y_{t-1}, ..., y_0)$ at time t is allowed to depend only on previous observations, but not on the state nor on future observations to enforce partial observability and causality, respectively. The expectation in Eq. 3 is over the realizations of the noise and the initial conditions. Each term in the sum is the expected instantaneous cost at time t. The total expected cost C penalizes large control signals – reflecting energetic or metabolic constraints – as well as deviations from desired trajectories or targets, through the symmetric positive semidefinite matrices $R_t \in \mathbb{R}^{p\times p}$, $R_t \geq 0$, and $Q_t \in \mathbb{R}^{m\times m}$, $Q_t \geq 0$, respectively.

The LQAG problem admits an analytical solution (Davis, 2013), which is the combination of a linear Kalman filter, providing optimal estimates $\hat{x}_t \equiv z_t$ of the partially observable state x_t, and a linear feedback controller defined by $u_t = L_t z_t$, which are computed independently, without mathematical dependence between control and filter gains – the so-called separation principle (Davis, 2013). The internal variable becomes a state estimate evolving according to

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

where $K_t \in \mathbb{R}^{m\times k}$ is the Kalman gain at time t. Solving the optimal control problem therefore consists in computing both the optimal filter and control gains, respectively K_t and $L_t \in \mathbb{R}^{p\times m}$, under the constraint that the internal dynamics follows the same forward dynamics as the state variable (matrices A and B; see Appendix A.2.1 for the well-known solutions).

While the analytical tractability of the LQAG framework is a key advantage, it comes at the expense of reduced biological realism. In particular, the noise model does not account for multiplicative noise, also neglecting internal sources of variability (Faisal et al., 2008; Moreno-Bote et al., 2014; Churchland et al., 2006; Franklin & Wolpert, 2011). To consider a more general and realistic noise model, following Todorov (2005), we first introduce multiplicative noise – both control-dependent and observational – into the system and observation dynamics in Eqs. 1,2. This leads to the modified equations

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

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

In this framework, executing a control input u_t adds noise whose magnitude scales with the input itself, Eq. 5. Conversely, sensing the partially observable state x_t introduces sensory noise whose magnitude scales with the state itself, Eq. 6. The matrices $C_i \in \mathbb{R}^{m\times p}$ and $D_i \in \mathbb{R}^{k\times m}$ define fixed gain patterns for the multiplicative noise components, while $\varepsilon_t \in \mathbb{R}^c$ and $\rho_t \in \mathbb{R}^d$ represent zero-mean noise vectors, each with identity covariance, $\Sigma_\varepsilon = \mathbb{I}_{c\times c}$ and $\Sigma_\rho = \mathbb{I}_{d\times d}$. As in the LQAG problem, control and observation noises are assumed to be mutually independent, and also independent from both the additive and multiplicative noise components.

Finding the optimal control signal $u_t(y_{t-1}, ..., y_0)$ that minimizes the cost in Eq. 3 with system and observation dynamics given by Eqs. 5,6 is a challenging problem with no known solutions, even in the case of Gaussian noise. In particular, no sufficient statistic, analogous to $\hat{x}_t \equiv z_t$, is known that would allow for a Kalman filter-like recursion. Following Todorov (2005), we assume that the control signal u_t can only linearly depend on the estimate $z_t \in \mathbb{R}^m$, that is, $u_t = L_t z_t$, with $L_t \in \mathbb{R}^{p\times m}$, and that the state estimate obeys the *matched* dynamical equation

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

with the same terminology as in Eq. 4, but where we have introduced an internal additive noise term $\eta_t \in \mathbb{R}^m$, with zero mean and covariance Σ_η. The internal noise may represent internal neural variability (Faisal et al., 2008; Moreno-Bote et al., 2014; Churchland et al., 2006; Franklin & Wolpert, 2011) or flaws in the filtering process itself, and it is introduced here to obtain a more realistic and general model (Todorov, 2005). Taken together, incorporating multiplicative and internal noise with the assumptions of a linear Kalman filter for state estimation and a linear control policy based on an internal estimate whose forward dynamics match those of the state (matrices A and B) gives rise to the more general Linear–Quadratic–Multiplicative–Internal (LQMI) problem. Solving this problem involves determining the optimal control gains $L_{0,...,T}$ and filter gains $K_{0,...,T}$ that minimize the quadratic cost function in Eq. 3 under the system, observation and estimate dynamics in Eqs. 5,6,7.

3 SOLVING THE LQMI PROBLEM

Here we provide an algorithm guaranteed to converge to a critical point of the cost function in Eq. 3, under the dynamics in Eqs. 5,6,7. The algorithm yields improved pairs of control and filter gains, fully solving the LQMI problem. The pseudocode is shown in Appendix A.3.1.

3.1 FIXED-POINT EQUATIONS OF THE COST FUNCTION

Assuming a linear control $u_t = L_t z_t$, we first rewrite the cost function in Eq. 3 as $C = \sum_{t=0}^{T} \left(\mathrm{tr}(Q_t S_t^{xx}) + \mathrm{tr}(L_t^\top R_t L_t S_t^{zz})\right)$, where we introduce the 2nd-order moment matrices $S_t^{xx} = \int dxdzp_t(x,z)xx^\top$, $S_t^{zz} = \int dxdzp_t(x,z)zz^\top$, and $S_t^{xz} = \int dxdzp_t(x,z)xz^\top$, with $p_t(x,z)$ being the joint distribution of x and z at time t generated by previous control and filter gains and averaging over noises and initial conditions following $p_0(x,z)$. To find the conditions for extrema on the control $L_{0,\dots,T}$ and filter $K_{0,\dots,T}$ gains we add Lagrange multipliers and define the new objective

$$C_{\mathcal{L}} = \sum_{t=0}^{T} \left(\mathrm{tr}(Q_t S_t^{xx}) + \mathrm{tr}(L_t^\top R_t L_t S_t^{zz})\right) - \sum_{t=1}^{T+1} \left(\mathrm{tr}(\Lambda_t G_t^{xx}) + \mathrm{tr}(\Omega_t G_t^{zz}) + \mathrm{tr}(\Gamma_t G_t^{xz})\right) \ , \quad (8)$$

where Λ_t, Ω_t and Γ_t are $\mathbb{R}^{m\times m}$ matrices of Lagrange multipliers. The constraints $G_t^{xx} = G_t^{zz} = G_t^{xz} = 0$ are given by the temporal evolution of S_t^{xx}, S_t^{zz} and S_t^{xz}, respectively, between two consecutive time steps t and $t+1$, obtained from Eqs. 5,6,7 (see Appendix A.2.2 for details), as

$$\begin{aligned}
G_{t+1}^{xx} &= S_{t+1}^{xx} - AS_t^{xx}A^\top - AS_t^{xz}L_t^\top B^\top - BL_t(S_t^{xz})^\top A^\top - BL_tS_t^{zz}L_t^\top B^\top - \Sigma_t^{xx} \\
G_{t+1}^{zz} &= S_{t+1}^{zz} - K_tHS_t^{xx}H^\top K_t^\top - K_tHS_t^{xz}M_t^\top - M_t(S_t^{xz})^\top H^\top K_t^\top - M_tS_t^{zz}M_t^\top - \Sigma_t^{zz} \\
G_{t+1}^{xz} &= S_{t+1}^{xz} - AS_t^{xx}H^\top K_t^\top - BL_tS_t^{zz}M_t^\top - AS_t^{xz}M_t^\top - BL_t(S_t^{xz})^\top H^\top K_t^\top \ , \qquad (9)
\end{aligned}$$

where we have introduced the short-hand notation $M_t = A + BL_t - K_tH$, showing up repetitively, and the noise matrices $\Sigma_t^{xx} = \Sigma_\xi + \sum_i C_iL_tS_t^{zz}L_t^\top C_i^\top$ and $\Sigma_t^{zz} = \Sigma_\eta + K_t\Sigma_\omega K_t^\top + K_t\left(\sum_i D_iS_t^{xx}D_i^\top\right)K_t^\top$. Since the cost function is defined in terms of quadratic terms in x and z and the temporal evolution of moments is closed at 2nd-order, the 2nd-order moments matrices are sufficient statistics of the problem (i.e., $p_t(x,z)$ does not need to be explicitly known), and only the constraints in their temporal evolution suffice.

For convenience, we define the Lagrange multipliers at time $T+1$ to be all equal to zero, $\Lambda_{T+1} = \Omega_{T+1} = \Gamma_{T+1} = 0$ (hereafter 0 meaning a matrix of zeros of consistent dimensions), so the constraints at that time are irrelevant. The introduction of Lagrange multipliers enables to take derivatives with respect the control and filter gains to find the fixed point conditions $\partial C_{\mathcal{L}}/\partial L_t = 0$ and $\partial C_{\mathcal{L}}/\partial K_t = 0$ for extrema without the need to propagate derivatives over the terms in the sum of the cost. The fixed point equations take the form

$$L_t = E_t^{-1}\left(F_tS_t^{xz}(S_t^{zz})^{-1} + J_t\right) \quad (10)$$

$$K_t = \left(S_{AH} + \tilde{\Omega}_{t+1}^{-1}\Gamma_{t+1}S_{LH}\right)S_{HH}^{-1} \ , \quad (11)$$

with matrices defined in Appendix A.2.4 – note that these equations express the control and filter gains as a function of themselves, and therefore they are implicit.

From the conditions $\partial C_{\mathcal{L}}/\partial S_t^{xx} = \partial C_{\mathcal{L}}/\partial S_t^{zz} = \partial C_{\mathcal{L}}/\partial S_t^{xz} = 0$, the Lagrange multipliers themselves obey the set of equations

$$\begin{aligned}
\Lambda_t &= Q_t + A^\top\Lambda_{t+1}A + H^\top K_t^\top\Omega_{t+1}K_tH + H^\top K_t^\top\Gamma_{t+1}A + \sum\nolimits_i D_i^\top K_t^\top\Omega_{t+1}K_tD_i \\
\Omega_t &= L_t^\top R_tL_t + L_t^\top B^\top\Lambda_{t+1}BL_t + M_t^\top\Omega_{t+1}M_t + M_t^\top\Gamma_{t+1}BL_t + \sum\nolimits_i L_t^\top C_i^\top\Lambda_{t+1}C_iL_t \\
\Gamma_t &= L_t^\top B^\top\tilde{\Lambda}_{t+1}A + M_t^\top\tilde{\Omega}_{t+1}K_tH + M_t^\top\Gamma_{t+1}A + L_t^\top B^\top\Gamma_{t+1}^\top K_tH \ . \qquad (12)
\end{aligned}$$

These equations can be solved backwards given control and filter gains, and using the boundary conditions $\Lambda_{T+1} = \Omega_{T+1} = \Gamma_{T+1} = 0$. However, the full solution to Eqs. 10,11,12 would require simultaneously determining gains and multipliers. We bypass this by deriving an iterative algorithm to find fixed point solutions, as described in the next section.

It is worth mentioning that in the derivation of Eqs. 10,11,12 and main algorithm described below we have not assumed the orthogonality principle (OP: $S_t^{xz} = S_t^{zz}$ for all t, equivalent to $\mathbb{E}[(x_t - z_t)z_t^\top] = 0$), which is shown (Sec. 3.3, see also Appendix A.1) to be violated in the general case (specifically, whenever there is internal noise). Secondly, we have not assumed any specific initial distribution $p_0(x, z)$. Also, note that we have not assumed Gaussian noises nor Gaussian distribution on x or z. Further, our algorithm is guaranteed to converge to a fixed-point pair of control and filter gains, and reduce the cost at every step (Sec. 3.2). The algorithm in Todorov (2005) can actually increase the cost in the first iteration step because not for any arbitrary initial filter gain OP is obeyed. Finally, the model described in Eqs. 5,6,7 could be readily extended to the case where i) the internal noise is multiplicative in Eq. 7, ii) when there is x-dependent multiplicative noise in the state dynamics, Eq. 5, and iii) when there is z-dependent multiplicative noise in the feedback dynamics, Eq. 6. However, we refrain from doing so to avoid clutter and because a more general framework (Model Mismatch) is introduced below (Sec. 4).

3.2 Coordinate-Descent Algorithm for Joint Control and Filter Optimization

Here we derive the main algorithm of the paper, a coordinate-descent iterative algorithm that gives a pair of improved, fixed-point control and filter gains. We first start by showing the connection between the Lagrange multipliers and the cost-to-go incurred by starting at fixed x and z.

We define the cost-to-go starting at x and z from time t ($t = 0, ..., T$) up to time T as $C_t(x, z) = \mathrm{tr}(Q_t x x^\top + L_t^\top R_t L_t z z^\top) + \sum_{\tau=t+1}^{T} \mathbb{E}\left[x_\tau^\top Q_\tau x_\tau + u_\tau^\top R_\tau u_\tau\right]$, where the expectation is over the noises with initial conditions fixed at x and z at time t, and for specific control and filter gains from time t onward. This definition is consistent with our definition of cost in Eq. 3, as $C = \int p_0(x, z) C_0(x, z)$, where $p_0(x, z)$ is the distribution of initial conditions over x and z. The cost-to-go obeys the Bellman equation

$$C_t(x, z) = \mathrm{tr}(Q_t x x^\top + L_t^\top R_t L_t z z^\top) + \int dx' dz' C_{t+1}(x', z') p_{x,t+1}(x'|x, z) p_{z,t+1}(z'|x, z) \ , \tag{13}$$

where the transition probability densities $p_{x,t+1}(x'|x, z)$ and $p_{z,t+1}(z'|x, z)$ are defined by equations 5,6,7 with $u_t = L_t z_t$, with means $\mathbb{E}[x'|x, z] = Ax + BL_t z$ and $\mathbb{E}[z'|x, z] = K_t H x + M_t z$, and conditional 2nd-order moments given by Eqs. 30.

The Bellman equation 13 can be solved backwards: noticing that the boundary condition is the final cost-to-go $C_T(x, z) = \mathrm{tr}(Q_T x x^\top + L_T^\top R_T L_T z z^\top)$ and that the 2nd-order moments are closed (that is, no higher nor lower moments appear when propagating backwards the cost-to-go using Eq. 13), we find that the solution is given by

$$C_t(x, z) = \mathrm{tr}(\Lambda_t x x^\top + \Omega_t z z^\top + \Gamma_t x z^\top) + \gamma_t \ , \tag{14}$$

where it can be seen that the coefficients Λ_t, Ω_t and Γ_t are actually the Lagrange multipliers computed in Eqs. 12 with the same boundary conditions (see Appendix A.2.3), and where γ_t can be recursively calculated as

$$\gamma_t = \mathrm{tr}(\Lambda_{t+1}\Sigma_\xi + \Omega_{t+1} K_t \Sigma_\omega K_t^\top + \Omega_{t+1}\Sigma_\eta) + \gamma_{t+1} \ , \tag{15}$$

with boundary condition $\gamma_T = 0$. Eqs. 14,15 correctly captures the cost-to-go expression at time T, and it can be checked that recursively solve Eq. 13.

While Eqs. 14,15 express the exact cost-to-go given control and filter gains if the exact world state x is known, partial observability dictates that our choices of control and filter gains cannot depend on x. Indeed, our assumptions that the filter depends only on time and that the control law depends linearly on the current state estimate z_t, that is, $u_t = L_t z_t$, have already been used in our derivation and problem formalization, and they are subject to partial observability. Because of this, we integrate over the (generally unknown) joint probability density $p_t(x, z)$ given control and filter gains and initial condition $p_0(x, z)$ to define the averaged cost-to-go as

$$C_t = \int dx dz \ p_t(x, z) C_t(x, z) = \mathrm{tr}(\Lambda_t S_t^{xx} + \Omega_t S_t^{zz} + \Gamma_t S_t^{xz}) + \gamma_t \ . \tag{16}$$

We can express the total cost in Eq. 3 as $C = C_0$, and therefore

$$C = C_{<t} + C_t \tag{17}$$

with $C_{<t} = \sum_{\tau=0}^{t-1} \mathrm{tr}(Q_\tau S_\tau^{xx} + L_\tau^\top R_\tau L_\tau S_\tau^{zz})$ is valid for all t. In Eq. 17, C_t is the only term depending on L_t, as $C_{<t}$ does not depend on it. Therefore, we locally optimize L_t as

$$L_t^* = \arg\min_{L_t} C_t \ , \tag{18}$$

while keeping the rest of gains fixed, that is, $L_{0,\dots,t-1,t+1,\dots,T}$ and $K_{0,\dots,T}$ are held constant. A global minimum always exists because C_t is non-negative. After noting that in C_t (Eq. 16) only the Lagrange multipliers depend on L_t (see Eqs. 12), while the 2nd-order moments at time t only depend on previous L_τ with $\tau < t$ (see Eqs. 29), the minimization results in

$$L_t^* = E_t^{-1} \left(F_t S_t^{xz} (S_t^{zz})^{-1} + J_t \right) \ , \tag{19}$$

with matrices identical to those in Eq. 10 and Appendix A.2.4, and whenever matrix inverses exist.

If $L_{0,\dots,T}$ and $K_{0,\dots,T}$ are the values of the control and filter gains before the optimization in Eq. 18, clearly the cost is non-increasing after the optimization,

$$C(L_0, ..., L_{t-1}, L_t^*, L_{t+1}, ..., L_T) \leq C(L_0, ..., L_{t-1}, L_t, L_{t+1}, ..., L_T) \ . \tag{20}$$

Note that after the optimization, the total cost in Eq. 17 becomes

$$C = C_{<t} + \mathrm{tr}(Q_t S_t^{xx} + L_t^{*\top} R_t L_t^* S_t^{zz}) + \mathrm{tr}(\Lambda_{t+1} S_{t+1}^{xx,*} + \Omega_{t+1} S_{t+1}^{zz,*} + \Gamma_{t+1} S_{t+1}^{xz,*}) + \gamma_{t+1} \ , \tag{21}$$

where the new 2nd-order moments at time $t+1$, S_{t+1}^*, are computed from the moments at the previous time t using Eqs. 29 with the optimal L_t^* and noticing that the Lagrange multipliers from $t+1$ onward have not changed. Redefining L_t^* as L_t and the $S_{t+1}^{ab,*}$ as S_{t+1}^{ab}, $ab \in \{xx, zz, xz\}$, we can now proceed to optimize L_{t+1} using the same procedure as above (changing t to $t+1$) to minimize again the total cost $C(L_0, ..., L_t, L_{t+1}^*, ..., L_T) \leq C(L_0, ..., L_t, L_{t+1}, ..., L_T)$ fixing all the gains except L_{t+1}, and consecutively for all t up to T.

Therefore, starting from a set of gains $L^{(n)} \equiv L_{0,\dots,T}^{(n)}$ and $K^{(n)} \equiv K_{0,\dots,T}^{(n)}$, we can optimize L_t in order from $t = 0$ up to time T following the above steps to get a new set of control gains $L^{(n+1)}$, and clearly we have $C(L^{(n+1)}, K^{(n)}) \leq C(L^{(n)}, K^{(n)})$. After this, the Lagrange multipliers in Eq. 12 are recomputed backwards with the updated values of the control gains, $L^{(n+1)}$. In this way, we can express again the cost as in Eq. 17, but with updated values of control gains and multipliers. This represents a full forward pass to sequentially optimize control gains followed by a full backward pass of the multipliers, and we refer to this process as *control pass*.

We can proceed similarly for the filter gains by repeating the above steps but for K_t instead of L_t. We optimize K_t by keeping fixed the remaining filter gains and all control gains by minimizing the cost C in Eq. 17, resulting in

$$K_t^* = \arg\min_{K_t} C_t = \left(S_{AH} + \tilde{\Omega}_{t+1}^{-1} \Gamma_{t+1} S_{LH} \right) S_{HH}^{-1} \ , \tag{22}$$

with matrices as in Eq. 11 and Appendix A.2.4. After updating the cost C with the new K_t^*, we obtain an equation analogous to Eq. 21 having a new γ_{t+1} term. This leads to a non-increasing cost change when going from the old K_t to the optimized K_t^*, $C(K_0, ..., K_t^*, ..., K_T) \leq C(K_0, ..., K_t, ..., K_T)$. Therefore, starting from a set of gains $L^{(n+1)}$ and $K^{(n)}$, we optimize K_t in order for $t = 0, ..., T$ to get a new set of filter gains $K^{(n+1)}$, which will obey $C(L^{(n+1)}, K^{(n+1)}) \leq C(L^{(n+1)}, K^{(n)})$. After this, the Lagrange multipliers are updated. This represents a *filter pass*: full forward pass to sequentially optimize filter gains followed by a full backwards pass to recompute the multipliers.

Starting from arbitrary $L^{(0)}$ and $K^{(0)}$ and distribution of initial conditions $p_0(x, z)$, we can alternate now the control and filter passes, so that $C(L^{(0)}, K^{(0)}) \geq C(L^{(1)}, K^{(0)}) \geq C(L^{(1)}, K^{(1)}) \geq ... \geq C(L^{(n+1)}, K^{(n)}) \geq C(L^{(n+1)}, K^{(n+1)}) \geq ... \geq C_{min} \geq 0$. Since the series is non-negative, it converges to a total cost no higher than the initial one with optimal filters $L^* = L^{(\infty)}$ and $K^* = K^{(\infty)}$. We have thus proven the first part of the following

Theorem 1. *Starting with arbitrary $L^{(0)}$ and K^0 and distribution of initial conditions $p_0(x, z)$, the coordinate descent algorithm defined by iterating in alternation control and filter passes converges to an improved pair of control and filter gains L^* and K^*. The improved pair corresponds to a critical point of the cost function in Eq. 3.*

To complete the last part of the theorem, it is clear that the converged pair of control and filter gains obey the Lagrange Eqs. 29,10,11,12, because Eqs. 19,22, after convergence, are identical to the fixed point Eqs. 10,11. Therefore, the converged pair corresponds a to a fixed point solution of the Lagrangian in Eq. 8, and hence, they must be a critical point of the cost function in Eq. 3.

We note that the Lagrange equations may admit multiple solutions. In practice, our algorithm converges to different critical points depending on the initialization, but when initializing the control and filter matrices trying to impose the orthogonality principle and then freely running the algorithm, the best critical point is found, empirically.

In Appendix A.4.1, we apply our algorithm to a 1D reaching task, modeled as a single-joint reaching movement with a 4D state, using the same setup as in Todorov (2005). We compare against the gradient descent numerical method from Damiani et al. (2024) and observe a substantial speedup: our algorithm (Algorithm 1) runs in ≈ 6 seconds, compared to over 5 hours, on a standard laptop. In Appendix A.4.2, we further evaluate our algorithm on increasingly high-dimensional tasks (up to 100 dimensions in the state variable) to demonstrate its scalability and the growing computational advantage over the same gradient descent numerical method. In the very high-dimensional case, the improvement is particularly pronounced, reducing runtime from more than 2 days to 2.7 seconds.

3.3 Orthogonality Principle Yields a Critical Point at Zero Internal Noise

Theorem 2. *Take initial condition $p_0(x, z)$ such that $S_0^{zz} = S_0^{xz}$. A solution to the Lagrange equations 9,10,11,12 is given by the orthogonality principle $S_t^{zz} = S_t^{xz}$ for $t = 1, ..., T$, iff internal noise is zero, that is, $\Sigma_\eta = 0$. The solution corresponds to a critical point of the cost in Eq. 8*

See the proof in Appendix A.2.5. We note that OP is implied by the unbiasedness condition (Appendix A.1), but not vice versa. While unbiasedness was empirically shown to be violated in Damiani et al. (2024), we have formally demonstrated that only the weaker OP condition is required to obtain a critical point of the cost. In Appendix A.2.6, we further show that, without multiplicative or internal noise, enforcing OP recovers the classical LQAG solution.

4 Optimal Control with Model Mismatch

We have shown that an analytical solution to the LQMI control problem can be derived requiring only standard assumptions: linear Kalman filtering for estimation and linear control laws. However, a central assumption remains unaddressed. By optimizing estimation and control gains ($K_{0,...,T}$ and $L_{0,...,T}$) one implicitly assumes i) that the agent's internal model exactly matches the true dynamics, and ii) that optimal behavior emerges from optimizing estimation and control as a partially decoupled process. This formalization weakens the notion of partial observability by presuming full access to the external world's dynamics. This assumption could result from learning, but it imposes strong constraints on the agent's internal strategy – leaving no room for internal computations structurally independent from the environment. This perspective also risks underestimating the role of internal representations, which are central to many motor control studies (Wolpert et al., 1995; Kawato, 1999; Shadmehr & Krakauer, 2008; Franklin & Wolpert, 2011; Golub et al., 2013; 2015), and could a priori combine estimation and control processes.

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 affected by noise (Hazon et al., 2022; Panzeri et al., 2022; Moreno-Bote et al., 2014). From a mathematical standpoint, in the classical LQAG settings, the separation principle holds: the optimal solution consists of an estimator (a linear Kalman filter) and a controller, designed independently, acting on the estimated state (Davis, 2013). In the more general and realistic LQMI setting, however, such a decomposition can only be assumed, rather than derived (Todorov, 2005). The solution in this case is, thus, not guaranteed to be optimal.

Therefore, we consider a more general control problem where the internal dynamics are also optimized and may become mismatched with the actual forward dynamics of the state variables. We formalize the new *Model Mismatch* framework over an even more general LQMI problem than that

described in Sec. 2, where multiplicative noise is fully generalized: both the state and internal dynamics may be affected by noise that depends on both the state and the internal variable. We then define the control problem as

$$x_{t+1} = Ax_t + BL_t z_t + n_t^x \ , \quad y_t = Hx_t + n_t^y \ , \quad z_{t+1} = W_t z_t + P_t y_t + n_t^z \tag{23}$$
$$n_t^c = \epsilon_t^c + \sum\nolimits_r \eta_t^c U_r^c x_t + \sum\nolimits_l \xi_t^c V_l^c L_t z_t \ , \quad c \in \{x, y, z\} \ ,$$

where notation follows Eqs. 5-7, with appropriate matrix dimensions and noises with covariances $\mathbb{E}[\epsilon_t^c \epsilon_t^{c'}] = \Sigma_{\epsilon^c} \delta_{cc'}$, and i.i.d. one-dimensional noises η_t^c and ξ_t^c with unit variance. We introduce additive and multiplicative noises n_t^c in the dynamics, observation and internal dynamics z_t. Sums over r and l can be c-dependent. We consider control-dependent noise, where the control is given by $u_t = L_t z_t$, rather than modeling the multiplicative noise as directly proportional to z_t. $P_t \in \mathbb{R}^{n \times m}$ is a pseudo-filter matrix that takes the observation y_t and inputs it to the dynamics of the internal variable z_t, which follows a linear system with time-dependent forward dynamics $W_t \in \mathbb{R}^{n \times n}$.

Importantly, in the Model Mismatch framework, the internal variable z_t integrates both control and estimation signals, unlike in the Model Match case where $z_t = \hat{x}_t$ is constrained to represent a state estimate. In the former, since W_t need not match the external dynamics, z_t can evolve independently of x_t and encode dynamics optimized for control rather than estimation. The internal variable z_t has dimension n, while the control signal $u_t = L_t z_t$ is again p-dimensional, with $L_t \in \mathbb{R}^{p \times n}$. The problem consists in optimizing the time-dependent, forward dynamics $W_{0,\dots,T}$, pseudo-filter $P_{0,\dots,T}$ and control $L_{0,\dots,T}$ matrices so as to minimize the cost in Eq. 3, with initial condition $p_0(x, z)$. Following the same procedure as in the Model Match approach (Sec. 3), we derive a coordinate-descent algorithm guaranteed to converge to a critical point of the cost (Appendix A.2.7; see pseudocode in Appendix A.3.2, Algorithm 2).

4.1 From Reaching to Neural Population Steering

We apply the Model Match (M-Match), Model Mismatch (M-Mis), and previous approaches (Todorov, 2005) to a 3D reaching task (with a 6-dimensional state, including positions and velocities, using $m, n, p, k = 6$; see Appendix A.4.3). In Fig. 1, $\tilde{W}_t = A + BL_t - P_t H$ (with P_t corresponding to K_t in Eq. 7) denotes the structure that W_t must take for the Model Mismatch formulation to reduce to the classical Model Match case. Setting $W_t = \tilde{W}_t$ makes the update of z_t in Eq. 23 identical to the Kalman filter dynamics, so then z_t serves as a standard state estimate of x_t.

The coordinate descent algorithm (Algorithm 2) converges reliably across different levels of internal noise σ_η (Fig. 1a). Compared to the solutions found under the Model Match framework (Sec. 3), the Model Mismatch solutions yield significantly better performance as internal noise increases (Fig. 1b). As internal noise grows, the internal variable becomes increasingly reliant on sensory feedback: the pseudo-filter matrices $P_{0,\dots,T}$ induce stronger transformations to compensate for the unreliability of internal dynamics. In contrast, the control matrix L_t induces weaker transformations (in terms of volume scaling) to suppress internal fluctuations when generating the control signal $u_t = L_t z_t$ (Fig. 1c). Notably, this modulation impacts the scaling properties of the system but not the effective embedding dimensionality – i.e., the number of dimensions corresponding to dynamically relevant directions (see Appendix A.4.3) – of the matrices involved (Fig. 1d).

Interestingly, the volume scaling of the internal dynamics (W_t), remains constant (Fig. 1c). What changes with increasing internal noise is the structure of the time-dependent forward dynamics matrix W_t: as internal noise grows, the optimal internal representations no longer aim to replicate the external world, as assumed in the Model Match framework, since the difference between W_t and $\tilde{W}_t$ increases (Fig. 1e). Consequently, the internal variable z_t can no longer be interpreted as an estimate of the state x_t; instead, it becomes a more abstract representation that integrates sensory feedback and past information to support optimal control (Fig. 1f), yet drastically reducing the control cost (see Fig. 1b).

To illustrate the conceptual shift, Appendix A.4.4 outlines example behavioral and neural predictions that distinguish the Model Mismatch and Model Match approaches. To demonstrate generality, Appendix A.4.5 applies the framework to a neural steering task that the classical formulation cannot model, where an unstable population is driven to a target state by another population providing the control signal (Fig. 5a).

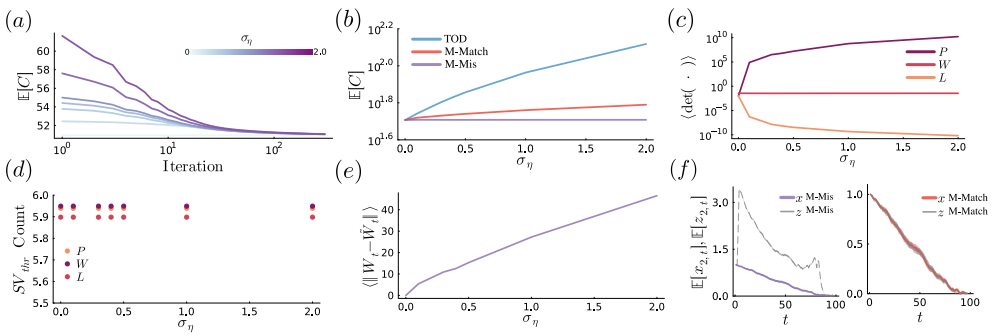

Figure 1: *Cost Reduction via Model Mismatch in 3D Reaching*. **(a)** Convergence of the Model Mismatch algorithm for different internal noise levels σ_η (parameter details are provided in Appendix A.4.3). **(b)** Expected cost for TOD (Todorov, 2005) (blue), Model Match (M-Match, orange), and Model Mismatch (M-Mis, purple). **(c)** Determinant of P_t, W_t, and L_t, averaged over time. **(d)** Embedding dimensionality of the same matrices, averaged over time (see Appendix A.4.3 for details). **(e)** Time-averaged norm of the difference between W_t and $\tilde{W}_t \equiv A + BL_t - P_t H$. **(f)** Second component of the state and internal variable (mean $\pm$ SEM, $\sigma\eta = 0.1$) for M-Mis (left) and M-Match (right).

5 Conclusions

We have introduced a convergent iterative algorithm (Sec. 3) that fully solves stochastic optimal control problems under a general noise model with both multiplicative and internal noise, assuming linear control with a quadratic cost – the so-called LQMI problem (Sec. 2). This goes beyond previous analytical approaches, which remained incomplete (Todorov, 2005; Damiani et al., 2024). Our algorithm also outperforms existing state-of-the-art gradient-based methods (Damiani et al., 2024) by more than three orders of magnitude in efficiency when applied to realistic tasks (Appendix A.4.1, Appendix A.4.2), making it particularly well-suited for inverse optimal control.

Moreover, our framework relaxes two central assumptions in stochastic control: (1) the partial decoupling of estimation and control, and (2) the requirement that internal forward dynamics match the actual state dynamics. By allowing internal dynamics — used to generate control signals — to be optimized jointly with control and pseudo-filter gains (Sec. 4), our framework broadens the solution space. Notably, we find that mismatched forward dynamics can outperform matched dynamics in the presence of internal noise. This suggests that internal representations need not faithfully track the state variable; instead, mixed representations of estimation and control signals can provide superior performance (Fig. 2). Furthermore, the Model Mismatch framework of Sec. 4 extends the applicability of stochastic optimal control to the control of neural populations (Appendix A.4.5).

Overall, our work expands stochastic optimal control to a more general, powerful, and realistic setting, with direct applications to neuroscience and robotics, while preserving analytical tractability and interpretability.

Limitations and Future Work We assume linear dynamics, linear control, and a quadratic cost, which yield closed-form second-order moments and analytical tractability but might not capture all problems of interest. Nevertheless, the framework accommodates time-varying dynamics, which can approximate nonlinearities. The Model Mismatch framework allows internal dimensionality to be freely chosen – a promising but unexplored direction that could support nonlinear strategies via linear representations (Korda & Mezić, 2018; Brunton et al., 2016). The coordinate descent algorithm that we have derived could handle suboptimal configurations of forward dynamics, control, or filters (e.g., to model constraints or impairments) by fixing some of these components during the updates.

REFERENCES

Vivek R Athalye, Preeya Khanna, Suraj Gowda, Amy L Orsborn, Rui M Costa, and Jose M Carmena. Invariant neural dynamics drive commands to control different movements. *Current Biology*, 33 (14):2962–2976, 2023.

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.

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

Tiago Costa, Juan R Castineiras de Saa, and Alfonso Renart. Optimal control of spiking neural networks. *bioRxiv*, pp. 2024–10, 2024.

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

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

Kenji Doya. *Bayesian brain: Probabilistic approaches to neural coding*. MIT press, 2007.

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

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

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

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

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

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

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.

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

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.

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

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

Konrad P Körding and Daniel M Wolpert. Bayesian integration in sensorimotor learning. *Nature*, 427(6971):244–247, 2004.

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

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.

Chiara Mastrogiuseppe and Ruben Moreno Bote. Controlled maximal variability along with reliable performance in recurrent neural networks. *Advances in Neural Information Processing Systems*, 37:24569–24600, 2024.

Francesca Mastrogiuseppe and Srdjan Ostojic. Linking connectivity, dynamics, and computations in low-rank recurrent neural networks. *Neuron*, 99(3):609–623, 2018.

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

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.

Kanaka Rajan, LF Abbott, and Haim Sompolinsky. Stimulus-dependent suppression of chaos in recurrent neural networks. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, 82(1):011903, 2010.

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.

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.

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

Reza Shadmehr, Maurice A Smith, and John W Krakauer. Error correction, sensory prediction, and adaptation in motor control. *Annual review of neuroscience*, 33(1):89–108, 2010.

Filip S Slijkhuis, Sander W Keemink, and Pablo Lanillos. Closed-form control with spike coding networks. *IEEE Transactions on Cognitive and Developmental Systems*, 2023.

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

Anderson Speed, Joseph Del Rosario, Navid Mikail, and Bilal Haider. Spatial attention enhances network, cellular and subthreshold responses in mouse visual cortex. *Nature communications*, 11 (1):505, 2020.

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

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

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

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

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

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

Martin Vinck, Renata Batista-Brito, Ulf Knoblich, and Jessica A Cardin. Arousal and locomotion make distinct contributions to cortical activity patterns and visual encoding. *Neuron*, 86(3):740–754, 2015.

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

A Appendix

A.1 Unbiasedness and Orthogonality: Clarifications and Implications

Here we briefly review related work on stochastic optimal control in the presence of multiplicative and internal noise (LQMI problem, Sec. 2.1). The influential work of Todorov (2005) introduced an iterative algorithm that alternates between optimizing the control and filter gains until convergence. A key assumption in this derivation is *unbiased estimation*, i.e., $\mathbb{E}[x_t \mid z_t] = z_t$, used to constrain the control policy to depend solely on the internal estimate z_t, in line with the problem's partial observability.

However, Damiani et al. (2024) empirically showed that this unbiasedness condition is generally violated, with the discrepancy growing as internal noise increases. They also proposed an alternative numerical algorithm that avoids assuming unbiasedness and empirically outperforms the original approach under internal noise.

The reason the method in Todorov (2005) still performs optimally when internal noise is absent is that unbiasedness implies the *orthogonality principle* (Davis, 2013; Damiani et al., 2024), which characterizes the optimal estimator in that specific case. Importantly, orthogonality does not imply unbiasedness, so the converse does not hold. Thus, the success of Todorov (2005) in the zero internal noise regime stems from its implicit reliance on orthogonality, which breaks down otherwise.

In Appendix A.2.5, we provide a formal proof that the orthogonality principle corresponds to a critical point of the cost function in Eq. 3 only in the absence of internal noise, extending and mathematically validating the empirical observations in Damiani et al. (2024). Moreover, in Appendix A.2.6, we demonstrate that the orthogonality principle actually leads to the global optimum for the classic LQAG problem.

A.2 Mathematical Derivations

A.2.1 Solutions of the Classic LQAG problem

The optimal $L_{0,\dots,T}$ and $K_{0,\dots,T}$, for the classic LQAG problem — defined in Sec. 2.1 – are given by (Doya, 2007; Davis, 2013; Todorov, 2005)

$$L_t = (2R_t + B^\top S_{t+1} B)^{-1} B^\top S_{t+1} A \tag{24}$$

$$S_t = 2Q_t + A^\top S_{t+1}(A + BL_t) \tag{25}$$

$$K_t = A\Sigma_t^e H^\top (H\Sigma_t^e H^\top + \Sigma_\omega)^{-1} \tag{26}$$

$$\Sigma_{t+1}^e = \Sigma_\xi + (A - K_t H)\Sigma_t^e A^\top \,. \tag{27}$$

A detailed derivation can be found in Doya (2007), Chapter 12, Sections 4 and 5. We observe that the only differences with the Eqs. in Doya (2007) arise from slightly different conventions: in the standard LQAG formulation, there is a prefactor of $1/2$ in front of the cost function, and the control signal is defined as $u_t = -L_t z_t$, meaning the control gain has the opposite sign compared to our convention.

In Appendix A.2.6, we prove that the solutions derived in Sec. 3 recover these classical results in the absence of multiplicative and internal noise.

A.2.2 Deriving the Propagation of Second-Order Moments

Here we derive the temporal evolution of the 2nd-order moment matrices. We first rewrite Eqs. 5,6,7 in a more compact form by inserting the observation in the state estimate variable and grouping terms as

$$\begin{aligned} x_{t+1} &= Ax_t + BL_t z_t + \xi_t + \sum\nolimits_i \varepsilon_t^i C_i L_t z_t \\ z_{t+1} &= M_t z_t + K_t H x_t + \eta_t + K_t \omega_t + K_t \sum\nolimits_i \rho_t^i D_i x_t \end{aligned} \tag{28}$$

with $M_t = A + BL_t - K_t H$.

The 2nd-order moments at time t can be computed based on those in the previous time step t by using the appropriate averages and interactions between terms in Eqs. 28. The result is

$$\begin{aligned} S^{xx}_{t+1} &= AS^{xx}_t A^\top + AS^{xz}_t L_t^\top B^\top + BL_t (S^{xz}_t)^\top A^\top + BL_t S^{zz}_t L_t^\top B^\top + \Sigma^{xx}_t \\ S^{zz}_{t+1} &= K_t H S^{xx}_t H^\top K_t^\top + K_t H S^{xz}_t M_t^\top + M_t (S^{xz}_t)^\top M_t^\top + M_t S^{zz}_t M_t^\top + \Sigma^{zz}_t \\ S^{xz}_{t+1} &= AS^{xx}_t H^\top K_t^\top + BL_t S^{zz}_t M_t^\top + AS^{xz}_t M_t^\top + BL_t (S^{xz}_t)^\top H^\top K_t^\top \ . \end{aligned} \tag{29}$$

with $M_t = A + BL_t - K_t H$ and noise covariances $\Sigma^{xx}_t = \Sigma_\xi + \sum_i C_i L_t S^{zz}_t L_t^\top C_i^\top$ and $\Sigma^{zz}_t = \Sigma_\eta + K_t \Sigma_\omega K_t^\top + K_t \left(\sum_i D_i S^{xx}_t D_i^\top \right) K_t^\top$.

The conditional second-order moments at time $t+1$ conditioned on x and z at time t are defined as

$$\begin{aligned} \hat{S}^{xx}_t &= \int dx' dz' x' x'^\top p_{x,t+1}(x'|x,z) p_{z,t+1}(z'|x,z) \\ \hat{S}^{zz}_t &= \int dx' dz' z' z'^\top p_{x,t+1}(x'|x,z) p_{z,t+1}(z'|x,z) \\ \hat{S}^{xz}_t &= \int dx' dz' x' z'^\top p_{x,t+1}(x'|x,z) p_{z,t+1}(z'|x,z) \ , \end{aligned}$$

where the transition probabilities $p_{x,t+1}(x'|x,z)$ and $p_{z,t+1}(z'|x,z)$ are defined by equations 5,6,7 (with $u_t = L_t z_t$), or, equivalently, by Eqs. 28. The conditional second-order moments at time $t+1$ are obtained simply by replacing the second-order moments on the right hand side of Eqs. 29 by their corresponding non-averaged x and z as

$$\begin{aligned} \hat{S}^{xx}_{t+1} &= Axx^\top A^\top + Axz^\top L_t^\top B^\top + BL_t zx^\top A^\top + BL_t zz^\top L_t^\top B^\top + \hat{\Sigma}^{xx}_t \\ \hat{S}^{zz}_{t+1} &= K_t H xx^\top H^\top K_t^\top + K_t H xz^\top M_t^\top + M_t zx^\top M_t^\top + M_t zz^\top M_t^\top + \hat{\Sigma}^{zz}_t \\ \hat{S}^{xz}_{t+1} &= Axx^\top H^\top K_t^\top + BL_t zz^\top M_t^\top + Axz^\top M_t^\top + BL_t zx^\top H^\top K_t^\top \ . \end{aligned} \tag{30}$$

with conditional noise covariances $\hat{\Sigma}^{xx}_t = \Sigma_\xi + \sum_i C_i L_t zz^\top L_t^\top C_i^\top$ and $\hat{\Sigma}^{zz}_t = \Sigma_\eta + K_t \Sigma_\omega K_t^\top + K_t \left(\sum_i D_i xx^\top D_i^\top \right) K_t^\top$.

A.2.3 Consistency of the Cost-to-Go Solution

The cost-to-go obeys the Bellman equation

$$C_t(x,z) = \mathrm{tr}(Q_t xx^\top + L_t^\top R_t L_t zz^\top) + \int dx' dz' C_{t+1}(x',z') p_{x,t+1}(x'|x,z) p_{z,t+1}(z'|x,z) \ , \tag{31}$$

identical to Eq. 13 in the main paper. The transition probability densities $p_{x,t+1}(x'|x,z)$ and $p_{z,t+1}(z'|x,z)$ are defined by equations 5,6,7 with $u_t = L_t z_t$, with means $\mathbb{E}[x'|x,z] = Ax + BL_t z$ and $\mathbb{E}[z'|x,z] = K_t H x + M_t z$, and 2nd-order moments given by Eqs. 30. These will be important to compute averages as needed.

We propose a solution to the Bellman equation of the form

$$C_t(x,z) = \mathrm{tr}(\Lambda_t xx^\top + \Omega_t zz^\top + \Gamma_t xz^\top) + \gamma_t \ , \tag{32}$$

identical to Eq. 14 in the main paper. Our goal is to show that it is possible to find a solution with such a form, and that the expression of the coefficients Λ_t, Ω_t and Γ_t are actually identical to the Lagrange multipliers in Eqs. 12 with the same boundary conditions. In addition we want to show that γ_t follows Eq. 15 with boundary condition $\gamma_T = 0$.

We first note that Eq. 32 is true for $t = T$, because $C_T(x,z)$ should be $C_T(x,z) = \mathrm{tr}(Q_T xx^\top + L_T^\top R_T L_T zz^\top)$ and indeed this coincides with Eq. 32 when taking $\Lambda_T = Q_T$, $\Omega_T = L_T^\top R_T L_T$, $\Gamma_T = 0$ and $\gamma_T = 0$, which in turn are consistent with the Lagrange multiplier expression in Eq. 12 for $t = T$.

Now, assume that Eq. 32 is true for some $t+1$. Let us show that then it is true for t. We insert Eq. 32 for $t+1$ into Eq. 31 and use the expression of the conditional 2nd-order moments in Eqs. 30 to

obtain

$$
\begin{aligned}
C_t(x,z) &= \mathrm{tr}(Q_t x x^\top + L_t^\top R_t L_t z z^\top) \\
&\quad + \int dx' dz' \left(\mathrm{tr}(\Lambda_{t+1} x' x'^\top + \Omega_{t+1} z' z'^\top + \Gamma_{t+1} x' z'^\top) + \gamma_{t+1}\right) p_{x,t+1}(x'|x,z) p_{z,t+1}(z'|x,z) \\
&= \mathrm{tr}(Q_t x x^\top + L_t^\top R_t L_t z z^\top) \\
&\quad + \mathrm{tr}[\Lambda_{t+1}(A x x^\top A^\top + B L_t z z^\top L_t^\top B^\top + A x z^\top L_t^\top B^\top + B L_t z x^\top A^\top + \hat{\Sigma}_t^{xx})] \\
&\quad + \mathrm{tr}[\Omega_{t+1}(K_t H x x^\top H^\top K_t^\top + M_t z z^\top M_t^\top + K_t H x z^\top M_t^\top + M_t z x^\top H^\top K_t^\top + \hat{\Sigma}_t^{zz})] \\
&\quad + \mathrm{tr}[\Gamma_{t+1}(A x x^\top H^\top K_t^\top + B L_t z z^\top M_t^\top + A x z^\top M_t^\top + B L_t z x^\top H^\top K_t^\top)] \\
&\quad + \gamma_{t+1} \,. \qquad (33)
\end{aligned}
$$

Grouping terms proportional to $xx^\top$, $xz^\top$ and $zz^\top$ and constant, we find that the cost-to-go can be written as Eq. 32 where the coefficients obey the Lagrange multiplier equations in Eqs. 12 at time t. In addition, γ_t is computed using Eq. 15.

By induction, then we have that Eq. 32 is true for all t and that the coefficients are indeed the Lagrange multipliers defined in Eqs. 12 and Eq. 15.

A.2.4 Fixed-Point Equations for Control and Filter Derivatives

The fixed point equations $\partial C_{\mathcal{L}}/\partial L_t = 0$ and $\partial C_{\mathcal{L}}/\partial K_t = 0$ for the extrema of the Lagrangian 8 take the form

$$
\begin{aligned}
\frac{\partial C_{\mathcal{L}}}{\partial L_t} &= \left[2R_t L_t + B^\top \left(\tilde{\Lambda}_{t+1} B L_t + \tilde{\Omega}_{t+1} M_t + \Gamma_{t+1} B L_t + \Gamma_{t+1}^\top M_t\right) + \sum\nolimits_i C_i^\top \tilde{\Lambda}_{t+1} C_i L_t\right] S_t^{zz} \\
&\quad + B^\top \left[\tilde{\Lambda}_{t+1} A + \tilde{\Omega}_{t+1} K_t H + \Gamma_{t+1} A + \Gamma_{t+1}^\top K_t H\right] S_t^{xz} = 0 \,, \qquad (34) \\
\frac{\partial C_{\mathcal{L}}}{\partial K_t} &= \left[\tilde{\Omega}_{t+1} K_t H + \Gamma_{t+1} A\right] S_t^{xx} H^\top - \left[\tilde{\Omega}_{t+1} M_t + \Gamma_{t+1} B L_t\right] S_t^{zz} H^\top - \tilde{\Omega}_{t+1} K_t H S_t^{xz} H^\top \\
&\quad + \tilde{\Omega}_{t+1} M_t (S_t^{xz})^\top H^\top - \Gamma_{t+1} A S_t^{xz} H^\top + \Gamma_{t+1} B L_t (S_t^{xz})^\top H^\top + \tilde{\Omega}_{t+1} K_t \Sigma_\omega \\
&\quad + \tilde{\Omega}_{t+1} K_t \sum\nolimits_i D_i S_t^{xx} D_i^\top = 0 \,, \qquad (35)
\end{aligned}
$$

with symmetric matrices $\tilde{\Lambda}_t = \Lambda_t + \Lambda_t^\top$ and $\tilde{\Omega}_t = \Omega_t + \Omega_t^\top$, after using elementary properties of the trace operator and its derivatives.

The fixed point equations can be further manipulated to express L_t and K_t as

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

where

$$
\begin{aligned}
E_t &= 2R_t + B^\top (\tilde{\Lambda}_{t+1} + \tilde{\Omega}_{t+1} + \Gamma_{t+1} + \Gamma_{t+1}^\top) B + \sum\nolimits_i C_i^\top \tilde{\Lambda}_{t+1} C_i \,, \\
F_t &= -B^\top (\tilde{\Lambda}_{t+1} A + \tilde{\Omega}_{t+1} K_t H + \Gamma_{t+1} A + \Gamma_{t+1}^\top K_t H) \,, \\
J_t &= -B^\top (\tilde{\Omega}_{t+1} + \Gamma_{t+1})(A - K_t H) \,,
\end{aligned}
$$

and

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

with

$$
\begin{aligned}
S_{AH} &= (A + B L_t)(S_t^{zz} - (S_t^{xz})^\top) H^\top \,, \\
S_{LH} &= \left(-A(S_t^{xx} - S_t^{xz}) + B L_t (S_t^{zz} - (S_t^{xz})^\top)\right) H^\top \,, \\
S_{HH} &= H(S_t^{xx} + S_t^{zz} - S_t^{xz} - (S_t^{xz})^\top) H^\top + \Sigma_\omega + \sum\nolimits_i D_i S_t^{xx} D_i^\top \,.
\end{aligned}
$$

Note that the equation for L_t explicitly depends on K_t on the right side, while the equation for K_t depends on L_t on the right side. This property enables the coordinate-descent algorithm described in the paper. The above expressions coincide with Eqs. 10,11.

A.2.5 Orthogonality Principle Yields a Critical Point if and Only if Internal Noise Vanishes

Theorem 2. *Take the initial condition $p_0(x,z)$ such that $S_0^{zz} = S_0^{xz}$. A solution to the Lagrange equations 9,10,11,12 is given by the orthogonality principle $S_t^{zz} = S_t^{xz}$ for $t = 1, ..., T$, iff internal noise is zero, that is $\Sigma_\eta = 0$. The solution corresponds to a critical point of the cost in Eq. 8*

Proof. We first show that (1) assuming OP ($S_t^{xz} = S_t^{zz}$ for $t = 0, ..., T$) is true, we prove that the satisfaction of the Lagrange equations for the multipliers, Eqs. 12, and the equation for the fixed point of L_t, Eq. 10, for all t implies that the *Lagrange equality*, $\Gamma_t = -\tilde{\Omega}_t$ for all t ($\tilde{\Omega}_t \equiv \Omega_t + \Omega_t^\top$), is true, regardless of the value of internal noise. Next, we show that (2) OP and the Lagrange equality imply satisfaction of the fixed point equation for K_t, Eq. 11, and the 2nd-order moments equations, Eqs. A.2.2, if and only if internal noise is zero, $\Sigma_\eta = 0$. This will show that OP solves all Lagrange equations iff internal noise is zero, and therefore it will correspond to a critical point of the cost function in Eq. 8.

(1) Assume that OP holds. From the boundary condition of the Lagrange equations for the multipliers we have that $\Lambda_{T+1} = \Omega_{T+1} = \Gamma_{T+1} = 0$. Therefore, at time $T+1$ the Lagrange equality $\Gamma_{T+1} = -\tilde{\Omega}_{T+1}$ is true. Let us prove by induction that the equality holds for all t. Assume that the Lagrange equality is true for some $t+1$, that is, $\Gamma_{t+1} = -\tilde{\Omega}_{t+1}$ (note that Γ_{t+1} is then symmetric). Then, from the Lagrange multipliers Eqs. 12 we can write

$$
\begin{aligned}
\Gamma_t &= L_t^\top B^\top \tilde{\Lambda}_{t+1} A + M_t^\top \tilde{\Omega}_{t+1} K_t H + M_t^\top \Gamma_{t+1} A + L_t^\top B^\top \Gamma_{t+1}^\top K_t H \\
&= L_t^\top B^\top \tilde{\Lambda}_{t+1} A - M_t^\top \Gamma_{t+1} K_t H + M_t^\top \Gamma_{t+1} A + L_t^\top B^\top \Gamma_{t+1} K_t H \\
\tilde{\Omega}_t &= 2 L_t^\top R_t L_t + L_t^\top B^\top \tilde{\Lambda}_{t+1} B L_t + M_t^\top \tilde{\Omega}_{t+1} M_t + M_t^\top \Gamma_{t+1} B L_t + L_t^\top B^\top \Gamma_{t+1} M_t \\
&\quad + \sum\nolimits_i L_t^\top C_i^\top \tilde{\Lambda}_{t+1} C_i L_t \\
&= 2 L_t^\top R_t L_t + L_t^\top B^\top \tilde{\Lambda}_{t+1} B L_t - M_t^\top \Gamma_{t+1} M_t + M_t^\top \Gamma_{t+1} B L_t + L_t^\top B^\top \Gamma_{t+1} M_t \\
&\quad + \sum\nolimits_i L_t^\top C_i^\top \tilde{\Lambda}_{t+1} C_i L_t \,,
\end{aligned}
$$

where we have replaced $\tilde{\Omega}_{t+1}$ by $-\Gamma_{t+1}$ and using that Γ_{t+1} is symmetric. Now, summing we have

$$
\begin{aligned}
\Gamma_t + \tilde{\Omega}_t &= 2 L_t^\top R_t L_t + L_t^\top B^\top \tilde{\Lambda}_{t+1} (A + B L_t) + M_t^\top \Gamma_{t+1} (A + B L_t - K_t H - M_t) \\
&\quad + L_t^\top B^\top \Gamma_{t+1}^\top (A + B L_t) + \sum\nolimits_i L_t^\top C_i^\top \tilde{\Lambda}_{t+1} C_i L_t \\
&= L_t^\top \left[2 R_t L_t + B^\top \tilde{\Lambda}_{t+1} (A + B L_t) + B^\top \Gamma_{t+1}^\top (A + B L_t) + \sum\nolimits_i C_i^\top \tilde{\Lambda}_{t+1} C_i L_t \right] \,,
\end{aligned}
\tag{36}
$$

where we have realized that the last term in the first line is zero.

Now, the solution for which OP holds should satisfy all other Lagrange equations, in particular the one for the fixed point equation for L_t, Eq. 10. As OP is assumed to be true at all times, and in particular at time t, and the Lagrange equality is assumed to be true for $t+1$, Eq. 10 (see Sec. A.2.4) largely simplifies to

$$
L_t = \bar{E}_t^{-1} \bar{F}_t \,, \tag{37}
$$

with $\bar{E}_t = 2R_t + B^\top (\tilde{\Lambda}_{t+1} + \Gamma_{t+1}) B + \sum_i C_i^\top \tilde{\Lambda}_{t+1} C_i$ and $\bar{F}_t = -B^\top (\tilde{\Lambda}_{t+1} + \Gamma_{t+1}) A$. Then, it is clear that the bracket in the last line of Eq. 36 is zero, and therefore the Lagrange equality $\Gamma_t = -\tilde{\Omega}_t$ is true. Therefore, by induction we conclude that the Lagrange equality is true for all t and that Lagrange equations for the multipliers and L_t are solved. Notice that the above results are true regardless of the presence of internal noise.

(2) Still we have not used the Lagrange equation for K_t, Eq. 11, nor the Lagrange equations for the 2nd-order moments, Eqs. 29. These equations must also be satisfied by the OP condition. First, from OP (and the implied Lagrange equality shown in (1)) the expression for K_t (see Sec. A.2.4) largely simplifies to

$$
K_t = A \left(S_t^{xx} - S_t^{zz} \right) H^\top \bar{S}_{HH}^{-1} \,, \tag{38}
$$

with $\bar{S}_{HH} = H (S_t^{xx} - S_t^{zz}) H^\top + \Sigma_\omega + \sum_i D_i S_t^{xx} D_i^\top$.

Now, this expression of K_t must solve the Lagrange equations for the 2nd-order moments. The equation for S_t^{xx} is trivially satisfied, but the equations for S_t^{xz} and S_t^{zz} should be such that $S_t^{xz} = S_t^{zz}$ for all t – otherwise, our OP initial assumption would be inconsistent; no other restrictions are imposed by the Lagrange equations of the 2nd-order moments. This is only possible iff the difference $S_{t+1}^{zz} - S_{t+1}^{xz}$ equals zero:

$$S_{t+1}^{zz} - S_{t+1}^{xz} = \left[-(A - K_t H)(S_t^{xx} - S_t^{zz})H^\top + K_t \Sigma_\omega + K_t \sum\nolimits_i D_i^\top S_t^{xx} D_i\right] K_t^\top + \Sigma_\eta = 0 \,, \tag{39}$$

for all t (this expression has been obtained using the 2nd-order moments in Eqs. 29 after several cancellations). In this expression, the bracket equals zero after using Eq. 38. Therefore, consistency of OP and satisfaction of the 2nd-order moments are satisfied if and only if internal noise is zero, $\Sigma_\eta = 0$.

This concludes the proof, because iff $\Sigma_\eta = 0$ we have a full satisfaction of all Lagrange equations for all t under the sole assumption of OP for all t. □

A.2.6 Recovery of Classical LQAG Solutions

In this section, we demonstrate that the solutions derived in Sec. 3 exactly recover the classical analytical solutions of the standard LQAG problem (see Appendix A.2.1) when both multiplicative and internal noise terms vanish. To illustrate this, we examine the solutions presented in Appendix A.2.5. As empirically validated in Damiani et al. (2024), the optimal solutions, when internal noise is absent, satisfy the orthogonality principle (OP). Thus, by setting the multiplicative noise terms to zero, we can directly verify whether these solutions converge to the classic LQAG solutions. Additionally, this provides a proof that the orthogonality principle indeed corresponds to the global optimum of the cost function for the standard LQAG problem.

The optimal controller derived under the orthogonality principle in Appendix A.2.5 is given by Eq. 37. When both multiplicative and internal noise terms are turned off, we obtain

$$L_t = -[2R_t + B^\top(\tilde{\Lambda}_{t+1} + \Gamma_{t+1})B]^{-1}[B^\top(\tilde{\Lambda}_{t+1} + \Gamma_{t+1})A] \,, \tag{40}$$

which corresponds to the optimal L_t for the classic LQAG case (see solutions in Sec. A.2.1) if $S_t = (\Gamma_t + \tilde{\Lambda}_t)$. Using Eq.12 and imposing the OP (setting $\Gamma_t = -\tilde{\Omega}_t$ – see Appendix A.2.5) we obtain

$$\Gamma_{t+1} + \tilde{\Lambda}_{t+1} = 2Q_t + (A + BL_t)^\top(\tilde{\Lambda}_t + \Gamma_t)A \,. \tag{41}$$

Now we observe, as discussed in Appendix A.2.5, that Γ_t is symmetric and the same holds for $\tilde{\Lambda}_t$ (by definition), therefore we can rewrite Eq. 41 as

$$\Gamma_{t+1} + \tilde{\Lambda}_{t+1} = 2Q_t + A^\top(\tilde{\Lambda}_t + \Gamma_t)(A + BL_t) \,. \tag{42}$$

which corresponds to the formula for S_t in Sec. A.2.1, therefore proving the equality between the two optimal solutions.

The optimal Kalman filter derived under the OP in Appendix A.2.5 is given by Eq. 38, corresponding to

$$K_t = A\,(S_t^{xx} - S_t^{zz})\,H^\top[H(S_t^{xx} - S_t^{zz})H^\top + \Sigma_\omega]^{-1} \,, \tag{43}$$

when neither internal nor multiplicative noise is considered. We note that this solution corresponds to the one presented in Sec. A.2.1 when $\Sigma_t^e = S_t^{xx} - S_t^{zz}$, which is automatically satisfied when the OP, stating $S_t^{zz} = S_t^{xz}$, holds.

Therefore, the solutions derived in Appendix A.2.5 correspond to the globally optimal solutions of the classic LQAG problem in the absence of multiplicative and internal noise.

A.2.7 Joint Optimization of Forward Dynamics, Pseudo-Filter, and Control with Model Mismatch

Model and Moments The Model Mismatch approach is defined by the equations

$$\begin{aligned} x_{t+1} &= Ax_t + BL_t z_t + n_t^x \,, \quad y_t = Hx_t + n_t^y \,, \quad z_{t+1} = W_t z_t + P_t y_t + n_t^z \\ n_t^c &= \epsilon_t^c + \sum\nolimits_r \eta_t^c U_r^c x_t + \sum\nolimits_l \xi_t^c V_l^c L_t z_t \,, \quad c \in \{x, y, z\} \,, \end{aligned} \tag{44}$$

identical to Eqs. 23. The goal is to optimize the forward dynamics $W_t \in \mathbb{R}^{n\times n}$, pseudo-filter $P_t \in \mathbb{R}^{n\times m}$ 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

$$C = \sum_{t=0}^{T} \mathbb{E}\left[x_t^\top Q_t x_t + z_t^\top L_t^\top R_t L_t z_t\right] = \sum_{t=0}^{T} \left(\text{tr}(Q_t S_t^{xx}) + \text{tr}(L_t^\top R_t L_t S_t^{zz})\right) \ , \tag{45}$$

with initial condition $p_0(x, z)$.

Eqs. 44 can be put in a more compact form as

$$\begin{aligned} x_{t+1} &= Ax_t + BL_t z_t + n_t^x \\ z_{t+1} &= W_t z_t + P_t H x_t + P_t n_t^y + n_t^z \\ n_t^c &= \epsilon_t^c + \sum\nolimits_r \eta_t^c U_r^c x_t + \sum\nolimits_l \xi_t^c V_l^c L_t z_t \ , \quad c \in \{x, y, z\} \ , \end{aligned} \tag{46}$$

from where it is more obvious that the system consists of two coupled linear dynamical systems with free parameters W_t, P_t and L_t chosen so as the minimize the cost. The sums $\sum_r$ and $\sum_l$ 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 Mismatch framework is strictly more general than the Model Match one because one always is free to choose in Eqs. 46 $P_t = K_t$ and $W_t = A + BL_t - K_t H$, leading exactly to the Model Match approach in Eqs. 5,6,7. The reverse, mapping the Model Mismatch approach into the Model Match one, is in general not possible.

The 2nd-order moments, appearing in the cost 45, obey

$$\begin{aligned} S_{t+1}^{xx} &= AS_t^{xx}A^\top + BL_t S_t^{zz} L_t^\top B^\top + AS_t^{xz} L_t^\top B^\top + BL_t (S_t^{xz})^\top A^\top + \Sigma_t^x \\ S_{t+1}^{zz} &= P_t H S_t^{xx} H^\top P_t^\top + W_t S_t^{zz} W_t^\top + P_t H S_t^{xz} W_t^\top + W_t (S_t^{xz})^\top H^\top P_t^\top + P_t \Sigma_t^y P_t^\top + \Sigma_t^z \\ S_{t+1}^{xz} &= AS_t^{xx} H^\top P_t^\top + BL_t S_t^{zz} W_t^\top + AS_t^{xz} W_t^\top + BL_t (S_t^{xz})^\top H^\top P_t^\top \ , \end{aligned} \tag{47}$$

with $\Sigma_t^c = \Sigma_{\epsilon^c} + \sum_r U_r^c S_t^{xx} (U_r^c)^\top + \sum_l V_l^c L_t S_t^{zz} L_t^\top (V_l^c)^\top$, $c \in \{x, y, x\}$.

Even though the Model Mismatch approach is more general than the Model Match 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 $C_t(x, z) = \text{tr}(Q_t x x^\top + L_t^\top R_t L_t z z^\top) + \sum_{\tau=t+1}^{T} \mathbb{E}\left[x_\tau^\top Q_\tau x_\tau + z_\tau^\top L_t^\top R_\tau L_t z_\tau\right]$, 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

$$C_t(x, z) = \text{tr}(Q_t x x^\top) + \text{tr}(L_t^\top R_t L_t z z^\top) + \int dx' dz' C_{t+1}(x', z') p_{x,t+1}(x'|x, z) p_{z,t+1}(z'|x, z) \ , \tag{48}$$

where 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 equations 44. Using backwards induction, and following similar steps to those in Secs. A.2.2 and A.2.3, it is not difficult to show that the cost-to-go can be written for all t ($t = 0, ..., T$) as

$$C_t(x, z) = \text{tr}(\Lambda_t x x^\top) + \text{tr}(\Omega_t z z^\top) + \text{tr}(\Gamma_t x z^\top) + \gamma_t \ , \tag{49}$$

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

$$\begin{aligned} \Lambda_t &= Q_t + A^\top \Lambda_{t+1} A + H^\top P_t^\top \Omega_{t+1} P_t H + H^\top P_t^\top \Gamma_{t+1} A \\ &\quad + \sum\nolimits_r (U_r^x)^\top \Lambda_{t+1} U_r^x + \sum\nolimits_r (U_r^y)^\top P_t^\top \Omega_{t+1} P_t U_r^y + \sum\nolimits_r (U_r^z)^\top \Omega_{t+1} U_r^z \ , \\ \Omega_t &= L_t^\top R_t L_t + L_t^\top B^\top \Lambda_{t+1} B L_t + W_t^\top \Omega_{t+1} W_t + W_t^\top \Gamma_{t+1} B L_t \\ &\quad + \sum\nolimits_r L_t^\top (V_r^x)^\top \Lambda_{t+1} V_r^x L_t + \sum\nolimits_r L_t^\top (V_r^y)^\top P_t^\top \Omega_{t+1} P_t V_r^y L_t + \sum\nolimits_r L_t^\top (V_r^z)^\top \Omega_{t+1} V_r^z L_t \ , \\ \Gamma_t &= L_t^\top B^\top (\Lambda_{t+1} + \Lambda_{t+1}^\top) A + W_t^\top (\Omega_{t+1} + \Omega_{t+1}^\top) P_t H + W_t^\top \Gamma_{t+1} A + L_t^\top B^\top \Gamma_{t+1}^\top P_t H \ , \\ \gamma_t &= \text{tr}(\Lambda_{t+1} \Sigma_{\epsilon^x}) + \text{tr}(P_t^\top \Omega_{t+1} P_t \Sigma_{\epsilon^y}) + \text{tr}(\Omega_{t+1} \Sigma_{\epsilon^z}) + \gamma_{t+1} \ , \end{aligned} \tag{50}$$

with boundary conditions $\Lambda_T = Q_T$, $\Omega_T = L_T^\top R_T L_T$, $\Gamma_T = 0$ and $\gamma_T = 0$ (in this way the boundary condition that $C_T(x,z) = \mathrm{tr}(Q_T x x^\top) + \mathrm{tr}(L_T^\top R_T L_T z z^\top)$ is satisfied).

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

$$C_t \equiv \int dx dz p_t(x,z) C_t(x,z) = \mathrm{tr}(\Lambda_t S_t^{xx}) + \mathrm{tr}(\Omega_t S_t^{zz}) + \mathrm{tr}(\Gamma_t S_t^{xz}) + \gamma_t \ , \tag{51}$$

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 in Eq. 45 can be written as

$$C = C_0 \equiv \int dx dz p_0(x,z) C_0(x,z) = \mathrm{tr}(\Lambda_0 S_0^{xx}) + \mathrm{tr}(\Omega_0 S_0^{zz}) + \mathrm{tr}(\Gamma_0 S_t^{xz}) + \gamma_0 \ , \tag{52}$$

which it can also be expressed as

$$C = C_{<t} + C_t \tag{53}$$

with $C_{<t} = \sum_{\tau=0}^{t-1} \mathrm{tr}(Q_\tau S_\tau^{xx} + L_\tau^\top R_\tau L_\tau S_\tau^{zz})$. It is important to note that Eq. 53 is valid for all t.

Algorithm Building an algorithm to find an improved triplet of time-dependent forward dynamics, pseudo-filter and control matrices is slightly simpler than in the case of the Model Match approach because W_t and P_t only appear in the internal variable dynamical equation and L_t only appears in the state variable dynamics. In contrast, in the Model Match approach, L_t appeared both in the state and state estimate dynamics, complicating the mathematical derivations.

Indeed, we note from Eqs. 50 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. 47. Therefore, choosing an arbitrary t, in Eq. 53 only the term C_t depends on W_t, and in that term, Eq. 51, only the coefficients Λ_t, Ω_t, Γ_t and γ_t can depend on W_t. In conclusion, starting with a set of $W_{0,\ldots,T}$, $P_{0,\ldots,T}$ and $L_{0,\ldots,T}$, we can improve the value of W_t as

$$W_t^* = \arg\min_{W_t} C = \arg\min_{W_t} C_t \ , \tag{54}$$

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

$$W_t^* = -P_t H S_t^{xz} (S_t^{zz})^{-1} - (\Omega_{t+1} + \Omega_{t+1}^\top)^{-1} \Gamma_{t+1} \left(B L_t + A S_t^{xz} (S_t^{zz})^{-1} \right) \ . \tag{55}$$

Note that if S_0^{zz} is not invertible, then W_0^* is not well defined, and thus we can take any arbitrary matrix. This might correspond to $z_0 = 0$. After the optimization, we must have

$$C^* = C(W_0, ..., W_t^*, ..., W_t) \leq C(W_0, ..., W_t, ..., W_T) \ , \tag{56}$$

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

$$C^* = C_{<t+1} + C_{t+1}^* = C_{<t+1} + \mathrm{tr}(\Lambda_{t+1} S_{t+1}^{xx,*}) + \mathrm{tr}(\Omega_{t+1} S_{t+1}^{zz,*}) + \mathrm{tr}(\Gamma_{t+1} S_{t+1}^{xz,*}) + \gamma_{t+1} \tag{57}$$

where the coefficients at time $t+1$ do not need to be updated (as they do not depend on W_t^*), but where the $S_{t+1}^{ab,*}$ need to be updated using Eqs. 47 with the new W_t^*.

Redefining W_t^* as W_t and the $S_{t+1}^{ab,*}$ as S_{t+1}^{ab}, 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 $C(W_0, ..., W_t, W_{t+1}^*, ..., W_T) \leq C(W_0, ..., W_t, W_{t+1}, ..., W_T)$ fixing $P_{0,\ldots,T}$, $L_{0,\ldots,T}$ and all W_τ except for $\tau = t$. 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 value of the coefficients in Eqs. 50 have to be recomputed so that Eq. 52 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 $W^{(n)} = W_{0,\ldots,T}^{(n)}$, $P^{(n)} = P_{0,\ldots,T}^{(n)}$ and $L^{(n)} = L_{0,\ldots,T}^{(n)}$, the W-pass leads to a new set of forward dynamics matrices $W^{(n+1)}$ such that the cost is non-increasing, $C(W^{(n+1)}, P^{(n)}, L^{(n)}) \leq C(W^{(n)}, P^{(n)}, L^{(n)})$. 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)

$$P_t^* = -\left[W_t(S_t^{xz})^\top + (\Omega_{t+1} + \Omega_{t+1}^\top)^{-1}\Gamma_{t+1}\left(AS_t^{xx} + BL_t(S_t^{xz})^\top\right)\right] H^\top E_t^{-1} \,, \tag{58}$$

with $E_t = HS_t^{xx}H^\top + \sum_l U_l^y S_t^{xx}(U_l^y)^\top + \sum_r V_l^y L_t S_t^{zz} L_t^\top (V_l^y)^\top + \Sigma_{\epsilon^y}$. Starting from $W^{(n+1)} = W_{0,...,T}^{(n+1)}$, $P^{(n)} = P_{0,...,T}^{(n)}$ and $L^{(n)} = L_{0,...,T}^{(n)}$, the P-pass leads to a new set of pseudo-filter matrices $P^{(n+1)}$ such that the cost is non-increasing, $C(W^{(n+1)}, P^{(n+1)}, L^{(n)}) \leq C(W^{(n+1)}, P^{(n)}, L^{(n)})$. 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)

$$L_t^* = -F_t^{-1}B^\top\left\{\tilde{\Lambda}_{t+1}AS_t^{xz}(S_t^{zz})^{-1} + \Gamma_{t+1}^\top\left[P_tHS_t^{xz}(S_t^{zz})^{-1} + W_t\right]\right\} \,, \tag{59}$$

with $F_t = 2R_t + B^\top\tilde{\Lambda}_{t+1}B + \sum_l (V_l^x)^\top\tilde{\Lambda}_{t+1}V_l^x + \sum_l (V_l^y)^\top P_t^\top\tilde{\Omega}_{t+1}P_tV_l^y + \sum_l (V_l^z)^\top\tilde{\Omega}_{t+1}V_l^x$, where we have defined $\tilde{\Lambda}_t = \Lambda_t + \Lambda_t^\top$ and $\tilde{\Omega}_t = \Omega_t + \Omega_t^\top$. Starting from $W^{(n+1)} = W_{0,...,T}^{(n+1)}$, $P^{(n+1)} = P_{0,...,T}^{(n+1)}$ and $L^{(n)} = L_{0,...,T}^{(n)}$, the L-pass leads to a new set of control matrices $L^{(n+1)}$ such that the cost is non-increasing, $C(W^{(n+1)}, P^{(n+1)}, L^{(n+1)}) \leq C(W^{(n+1)}, P^{(n+1)}, L^{(n)})$.

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

$$\begin{aligned} C(W^{(0)}, P^{(0)}, L^{(0)}) &\geq C(W^{(1)}, P^{(0)}, L^{(0)}) \geq C(W^{(1)}, P^{(1)}, L^{(0)}) \geq ... \\ &\geq C(W^{(n+1)}, P^{(n)}, L^{(m)}) \geq C(W^{(n+1)}, P^{(n+1)}, L^{(n)}) \\ &\geq C(W^{(n+1)}, P^{(n+1)}, L^{(n+1)}) \geq ... \geq C_{min} \geq 0 \,. \end{aligned} \tag{60}$$

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 3. *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. 45.*

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 45, we must show that they solve all fixed points equations of the Lagrangian,

$$C_{\mathcal{L}} = \sum_{t=0}^{T}\left(\text{tr}(Q_tS_t^{xx}) + \text{tr}(R_tS_t^{zz})\right) - \sum_{t=1}^{T+1}\left(\text{tr}(\Lambda_tG_t^{xx}) + \text{tr}(\Omega_tG_t^{zz}) + \text{tr}(\Gamma_tG_t^{xz})\right) \,, \tag{61}$$

where Λ_t, Ω_t and Γ_t are matrices of Lagrange multipliers. The constraints $G_t^{xx} = G_t^{zz} = G_t^{xz} = 0$ are given by the temporal evolution of S_t^{xx}, S_t^{zz} and S_t^{xz}, respectively, between two consecutive time steps t and $t+1$, and can be computed using Eqs. 47 similarly as in Eqs. 9. Indeed, the fixed point equations of the Lagrangian $\partial C_{\mathcal{L}}/\partial W_t = 0$ and $\partial C_{\mathcal{L}}/\partial P_t = 0$ are identical to Eqs. 55,58,59, respectively, which must be satisfied after convergence by the improved triplet W^*, P^* and L^*. After some work, the Lagrange equations $\partial C_{\mathcal{L}}/\partial S_t^{xx} = 0$, $\partial C_{\mathcal{L}}/\partial S_t^{xx} = 0$ and $\partial C_{\mathcal{L}}/\partial S_t^{xx} = 0$ can be seen to lead exactly to the coefficient Eqs. 50, which, again, are satisfied by the improved triplet. Finally, the derivatives of the Lagrangian with respect to the multipliers reduce to the second-order moment Eqs. 47, which are satisfied by the improved triplet. Thus, the improved triplet is a fixed-point solution of the Lagrangian 61 and therefore a critical point of the cost function 45.

A.3 Algorithms Implementation: Pseudocodes

A.3.1 Pseudocode – Model Match framework

Algorithm 1 Model Match (M-Match) approach

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. 12 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. 10,
    $S^{xx}_{t+1}, S^{xz}_{t+1}, S^{zz}_{t+1} \leftarrow$ Eqs. 29 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. 12 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. 11,
    $S^{xx}_{t+1}, S^{xz}_{t+1}, S^{zz}_{t+1} \leftarrow$ Eqs. 29 using $L_t^{(k)}$ and $K_t^{(k)}$
  end for
end for
$L^*_{0,\dots,T} \leftarrow L^{(k)}_{0,\dots,T}$; $K^*_{0,\dots,T} \leftarrow K^{(k)}_{0,\dots,T}$

The pseudocode above implements the algorithm of Sec. 3.2, referred to as the Model Match (M-Match) approach, in contrast to the Model Mismatch (M-Mis) method of Sec. 4.

A.3.2 Pseudocode – Model Mismatch framework

Algorithm 2 Model Mismatch (M-Mis) approach

Input: $S_0^{xx}, S_0^{xz}, S_0^{zz}$; initial guesses $L^{(0)}_{0,\dots,T}, P^{(0)}_{1,\dots,T}, W^{(0)}_{1,\dots,T}$; system parameters.
Output: Optimal matrices $L^*_{0,\dots,T}, P^*_{1,\dots,T}, W^*_{1,\dots,T}$.
Steps:
for each iteration $k = 1, \dots,$ optimization steps **do**
  $\Lambda_{1,\dots,T}, \Omega_{1,\dots,T}, \Gamma_{1,\dots,T} \leftarrow$ Eqs. 50 using $P^{(k-1)}_{1,\dots,T}$, $W^{(k-1)}_{1,\dots,T}$ and $L^{(k-1)}_{0,\dots,T}$ (backward equations)
  for each iteration $t = 0, \dots, T-1$ **do**
    $P_t^{(k)} \leftarrow$ Eq. 58,
    $S^{xx}_{t+1}, S^{xz}_{t+1}, S^{zz}_{t+1} \leftarrow$ Eqs. 47 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. 50 using $P^{(k)}_{1,\dots,T}$, $W^{(k-1)}_{1,\dots,T}$ and $L^{(k-1)}_{0,\dots,T}$ (backward equations)
  for each iteration $t = 0, \dots, T-1$ **do**
    $W_t^{(k)} \leftarrow$ Eq. 55,
    $S^{xx}_{t+1}, S^{xz}_{t+1}, S^{zz}_{t+1} \leftarrow$ Eqs. 47 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. 50 using $P^{(k)}_{1,\dots,T}$, $W^{(k)}_{1,\dots,T}$ and $L^{(k-1)}_{0,\dots,T}$ (backward equations)
  for each iteration $t = 0, \dots, T-1$ **do**
    $L_t^{(k)} \leftarrow$ Eq. 59,
    $S^{xx}_{t+1}, S^{xz}_{t+1}, S^{zz}_{t+1} \leftarrow$ Eqs. 47 using $P_t^{(k)}$, $W_t^{(k)}$ and $L_t^{(k)}$
  end for
end for
$P^*_{1,\dots,T} \leftarrow P^{(k)}_{1,\dots,T}$; $W^*_{1,\dots,T} \leftarrow W^{(k)}_{1,\dots,T}$; $L^*_{0,\dots,T} \leftarrow L^{(k)}_{0,\dots,T}$

The pseudocode above outlines the Model Mismatch (M-Mis) approach, introduced in Sec. 4 and detailed in Appendix A.2.7. While the order of optimization for P, W, and L differs from that in Appendix A.2.7, all variants converge to a critical point of the cost function in Eq. 45.

A.3.3 Implementations Details

Here we report the algorithms' hyper-parameters, as selected for the experiments described in Sec. A.4.

For the single-joint reaching task used to evaluate the algorithm derived in Sec. 3 (Algorithm 1) – and to compare it with the gradient-based numerical method from Damiani et al. (2024) (referred to as GD) – we use the parameters listed in Table 1. Note that, in line with Damiani et al. (2024), the GD algorithm is implemented using the GradientDescent() function from the Optim.jl Julia package.

Table 1: Hyper-parameters of the algorithms used in the single-joint reaching task (Sec. A.4.1)

Algorithm	Description	value
GD (Damiani et al., 2024)	Number of iterations of the "GradientDescent()" function	50000
M-Match (Algorithm 1)	Number of iterations of the estimation-control optimization	100

For the 3D reaching task, detailed in Sec. A.4.3 we use

Table 2: Hyper-parameters of the algorithms used in the 3D reaching task (Sec. A.4.3)

Algorithm	Description	value
TOD (Todorov, 2005)	Number of iterations of the estimation-control optimization	100
M-Match (Algorithm 1)	Number of iterations of the estimation-control optimization	100
M-Mis (Algorithm 2)	Number of iterations of the M-Mis optimization	100

while for the neural population steering task of Sec. A.4.5 we selected the following hyper-parameters

Table 3: Hyper-parameters of the algorithm used in the neural population steering task (Sec. A.4.5)

Algorithm	Description	value
M-Mis (Algorithm 2)	Number of iterations of the $L_{0,\dots,T}$ optimization	20

A.4 Experimental Details and Supplementary Results

A.4.1 Optimal Model Match Control for Goal-Directed Behavior

We evaluated the algorithm introduced in Sec. 3.2 on a single-joint reaching task, using the same problem formulation as in (Todorov, 2005; Damiani et al., 2024). The system features a four-dimensional state and one-dimensional control and sensory feedback, i.e., $m = 4$, $p = k = 1$.

The discrete-time dynamics is given by Todorov (2005),

$$
\begin{aligned}
p(t + \Delta t) &= p(t) + \dot{p}(t)\Delta t \\
\dot{p}(t + \Delta t) &= \dot{p}(t) + f(t)\Delta t/m \\
f(t + \Delta t) &= f(t)(1 - \Delta t/\tau_2) + g(t)\Delta t/\tau_2 \\
g(t + \Delta t) &= g(t)(1 - \Delta t/\tau_1) + u(t)(1 + \sigma_\varepsilon \varepsilon_t)\Delta t/\tau_1
\end{aligned}
$$

with

$$A = \begin{pmatrix} 1 & \Delta t & 0 & 0 \\ 0 & 1 & \Delta t/m & 0 \\ 0 & 0 & 1 - \Delta t/\tau_2 & \Delta t/\tau_2 \\ 0 & 0 & 0 & 1 - \Delta t/\tau_1 \end{pmatrix}$$

$$B = (0 \quad 0 \quad 0 \quad \Delta t/\tau_1)^\top$$

$$C = (0 \quad 0 \quad 0 \quad \sigma_\varepsilon \Delta t/\tau_1)^\top$$

$$H = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$D = \begin{pmatrix} \sigma_\rho & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$Q_{1,\cdots,T-1} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$Q_T = \vec{p}\vec{p}^\top + \vec{v}\vec{v}^\top + \vec{f}\vec{f}^\top$$

$$R_{1,\cdots,T-1} = \frac{r}{T-1}$$

$$R_T = 0$$

$$\vec{p} = (1 \quad 0 \quad 0 \quad 0)$$

$$\vec{v} = (0 \quad w_v \quad 0 \quad 0)$$

$$\vec{f} = (0 \quad 0 \quad w_v \quad 0)$$

$$\Sigma_\xi = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \sigma_\xi^2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\Sigma_\omega = \sigma_\omega^2$$

$$\Sigma_\eta = \begin{pmatrix} \sigma_\eta^2 & 0 & 0 & 0 \\ 0 & \sigma_\eta^2 & 0 & 0 \\ 0 & 0 & \sigma_\eta^2 & 0 \\ 0 & 0 & 0 & \sigma_\eta^2 \end{pmatrix}$$

with the initial conditions given by

$$\mathbb{E}[x_1] = (x_1 \quad 0 \quad 0 \quad 0)^\top$$

$$\mathbb{E}[z_1] = \mathbb{E}[x_1]$$

$$\Sigma_{x_1} = \begin{pmatrix} \sigma_x^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\Sigma_{z_1} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

The parameters of the problem are listed in Table 4 (std = standard deviation).

Table 4: Parameters of the single-joint reaching task

Name	Description	Value
Δt	time-step (s)	0.010
m	mass of the hand (Kg)	1
τ_1	first time constant of the second order low pass filter	0.04
τ_2	second time constant of the second order low pass filter	0.04
r	Auxiliary variable for control-dependent cost	$1e^{-5}$
w_v	Auxiliary variable for task-related cost	0.2
w_f	Auxiliary variable for task-related cost	0.01
T	time steps	100
x_1	Target position	0.15
σ_x	Target position standard deviation	0.0
σ_ξ	std of dynamics noise ξ_t	0.1
σ_ω	std of the sensory noise ω_t	0.1
σ_ε	std of the control-dependent noise ε_t	0.5
σ_ρ	std of the sensory-dependent noise ρ	0.5
σ_η	std of the additive internal noise η_t	0.1

Fig 2a shows that Algorithm 1 consistently decreases the initial expected cost and converges to a critical point of the cost function in Eq. 3 after approximately 100 iterations. Fig. 2b compares the optimal controller obtained with Algorithm 1 (red lines, left panel) to the numerical solution proposed in Damiani et al. (2024) (green lines, right panel), which represents the state-of-the-art for the LQMI control problem. Both algorithms converge to the same solution, but with sensibly different computational costs: our method completes in approximately 6 seconds, whereas the approach in Damiani et al. (2024) takes over 5 hours on a standard laptop.

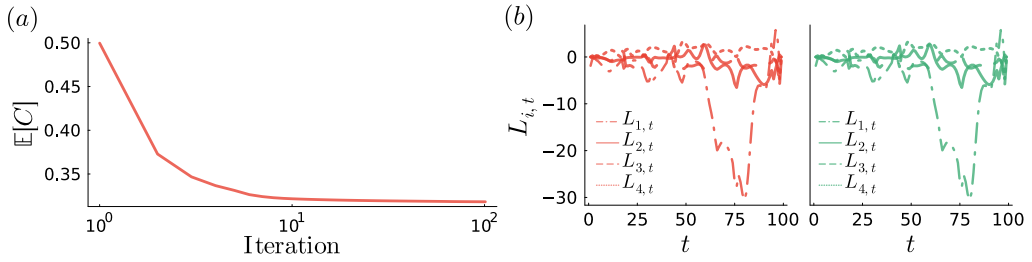

Figure 2: *Model Match Approach Converges to the Optimal LQMI Solution*.**(a)** Expected accumulated cost C (Eq. 3), computed via Eq. 16, during the joint optimization of control and filter gains using Algorithm 1, described in Sec. 2.1. **(b)** Optimal control gains L_t obtained using the algorithm from Sec. 2.1 (red lines, left panel) and the numerical method from Damiani et al. (2024) (green lines, right panel). Here, $L_{i,t}$ denotes the i-th component of the 4-dimensional control gain vector at time t (note that $p = 1$).

A.4.2 Computational Efficiency and Dimensionality Scaling: Comparison with Prior Work

As additional evidence for computational efficiency of the Algorithm of Sec. A.3.1, we present a dimensionality-scaling study comparing computation times with the numerical algorithm in Damiani et al. (2024), extending the analysis up to $m = 100$. This complements the results in Sec. A.4.1, which already demonstrates a pronounced gap in runtime (6 s vs. 5 h).

To isolate the effect of dimensionality, we set $m = k = p = n_{\text{shared}}$. Matrices A, B, C, and D are drawn from zero-mean, unit-variance Gaussian distributions and rescaled to ensure spectral radius < 1 for stability. We fix $T = 6$ and $\sigma_\xi = \sigma_\omega = \sigma_\rho = \sigma_\epsilon = \sigma_\eta = 0.2$, and vary $n_{\text{shared}} \in \{5, 10, 15, 40, 100\}$. We then compare the total computation time of our method (Secs. 3.2 and A.3.1) with the numerical approach in Damiani et al. (2024), initializing both with optimal gains

from Todorov (2005) to ensure a fair comparison. All results were obtained on a MacBook Pro (Apple M1, 16 GB RAM).

Table 5: Comparison of runtime between this work and the numerical algorithm in Damiani et al. (2024) as a function of the number of shared dimensions n_{shared}.

n_{shared}	**This work**	**GD (Damiani et al., 2024)**
5	1.15 s	8.4 min
10	1.25 s	75.7 min
15	1.40 s	6.4 h
40	2.7 s	$>$ 2 days
100	14 s	–

Here, s = seconds, min = minutes, and h = hours. These results highlight the scalability of our method. Similar time gaps also emerge in lower-dimensional settings as trial duration T increases, due to the linear growth in optimization parameters with T.

This computational advantage is critical for applying stochastic optimal control to real-world problems, particularly in Inverse Optimal Control (Schultheis et al., 2021; Straub & Rothkopf, 2022), which requires solving many control problems across parameter settings. The high cost of Damiani et al. (2024) renders it impractical for realistic tasks such as that in Sec. A.4.1, first described in Todorov (2005).

A.4.3 3D Reaching Task: Model and Parameters

The problem is defined by the following matrices:

$$A = \begin{pmatrix} 1 & 0 & 0 & \Delta t & 0 & 0 \\ 0 & 1 & 0 & 0 & \Delta t & 0 \\ 0 & 0 & 1 & 0 & 0 & \Delta t \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$$B = I_6$$

$$C = \sigma_\varepsilon \cdot I_6$$

$$H = I_6$$

$$D = \sigma_\rho \cdot I_6$$

$$\Sigma_\xi = \sigma_\xi^2 \cdot I_6$$

$$\Sigma_\omega = \sigma_\omega^2 \cdot I_6$$

$$\Sigma_\eta = \sigma_\eta^2 \cdot I_6$$

$$Q_{1,\dots,T-1} = 0_{6\times 6}$$

$$Q_T = \begin{pmatrix} 10 & 0 & 0 & 0 & 0 & 0 \\ 0 & 10 & 0 & 0 & 0 & 0 \\ 0 & 0 & 10 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$$R_t = r \cdot I_6 \quad \text{for } t = 1, \dots, T-1$$

$$R_T = 0\ ,$$

where I_6 denotes the 6×6 identity matrix, and $0_{6\times6}$ denotes the 6×6 zero matrix. The initial conditions are given by:

$$\begin{aligned}\mathbb{E}[x_1] &= \begin{pmatrix}1.5 & 1.0 & 2.5 & 10^{-5} & 10^{-5} & 10^{-5}\end{pmatrix}^\top \\ \mathbb{E}[z_1] &= \mathbb{E}[x_1] \\ \Sigma_{x_1} &= 0_{6\times6} \\ \Sigma_{z_1} &= 0_{6\times6}\end{aligned}$$

The parameters of the problem are listed in Table 6 (std = standard deviation).

Table 6: Parameters of the 3D reaching task

Name	Description	Value
Δt	Time step (s)	0.010
T	Time steps	100
m	Dimension of state x_t	6
n	Dimension of internal state z_t (for NSC)	6
p	Dimension of observation y_t	6
k	Dimension of control u_t	6
r	Control cost scaling	0.0001
σ_ξ	Std of dynamics noise ξ_t	0.5
σ_ω	Std of additive sensory noise ω_t	0.5
σ_ρ	Std of multiplicative sensory noise ρ	0.4
σ_ε	Std of multiplicative control noise ε_t	0.4
σ_η	Std of additive internal noise η_t	$\{0.0, 0.1, 0.3, 0.4, 0.5, 1.0, 2.0\}$

In this experiment, we set the control matrix to $B = I_6$ and use a control signal with dimensionality equal to the state ($p = m = 6$), enabling full control of the system. This choice is primarily motivated by numerical considerations: it avoids instabilities in our Model Mismatch algorithm related to matrix inversions that arise when B is not full-rank or poorly conditioned.

Although this means that control directly affects all state variables – including positions – this can be interpreted as an idealized feedback mechanism. The dynamics matrix A still captures the physical structure, with positions evolving from velocities over time. Our focus is on assessing algorithmic performance under internal and multiplicative noise, rather than enforcing strict biomechanical realism. Nonetheless, the setup remains rich enough to support meaningful behavioral predictions and comparisons with biological control strategies.

Embedding Dimensionality In Fig. 1d, we plot the embedding dimensionality of the matrices P, W, and L. For each time step t, we compute the number of singular values of P_t, W_t, and L_t that are larger than $0.01 \cdot \max_{\sigma_i \in SV}\{\sigma_i\}$, where SV denotes the set of singular values of the matrix under consideration. We then average this count across time steps to obtain a measure of effective dimensionality. Formally, we define:

$$SV_{thr}\text{Count} = \sum_{\sigma_i \in SV} \theta\left(\sigma_i \geq 0.01 \cdot \max_{\sigma_j \in \text{SV}} \sigma_j\right)$$

where $\theta(x)$ is the Heaviside step function. This quantity provides an estimate of the "effective" dimensionality of the transformation induced by the matrix, relative to its dominant singular values. This method accounts for changes in scale – such as reductions or increases in determinant magnitude due to varying levels of internal noise (Fig. 1c) – and thus provides a more meaningful estimate of dimensionality across different values of σ_η.

A.4.4 Distinct Neural and Behavioral Signatures of Model Match and Model Mismatch approaches

While our main focus is to introduce an analytical solution to stochastic optimal control problems with multiplicative and internal noise, the two frameworks considered here – Model Match and Model Mismatch – also lead to distinct, experimentally testable predictions. Below we outline illustrative examples that highlight these differences and the importance of choosing between the two approaches.

Divergence of internal dynamics In the 3D reaching task (Fig. 1), the Model Mismatch approach exhibits qualitatively different strategies from the Model Match one. With internal noise, optimal control (Fig.1b) is achieved when internal dynamics diverge from external ones (Fig. 1e), leading to z_t that no longer tracks x_t (Fig. 1f). This suggests a fundamentally different way of handling internal fluctuations. Using inverse optimal control (Schultheis et al., 2021; Straub & Rothkopf, 2022), behavior can be fit under both Model Match and Model Mismatch approaches, allowing one to test whether neural activity aligns more closely with the inferred internal dynamics of one framework. If it resembles M-Match's z_t, it may reflect state estimation (e.g., posterior parietal cortex or cerebellum); if it resembles M-Mis's z_t, it may reflect control-optimized representations, possibly in premotor or motor areas.

Noise-Dependent Control Magnitude From a behavioral perspective, in the same task as above, the magnitude of the control signal is strongly modulated by internal noise in the Model Match approach (Fig. 3a). In contrast, the Model Mismatch approach maintains a stable temporal profile of control magnitude across noise levels (Fig. 3a), likely due to flexible internal representations not constrained to track the external state (Fig. 1f). Internal fluctuations could in principle be experimentally influenced or estimated (Speed et al., 2020; Vinck et al., 2015), making this prediction possibly testable.

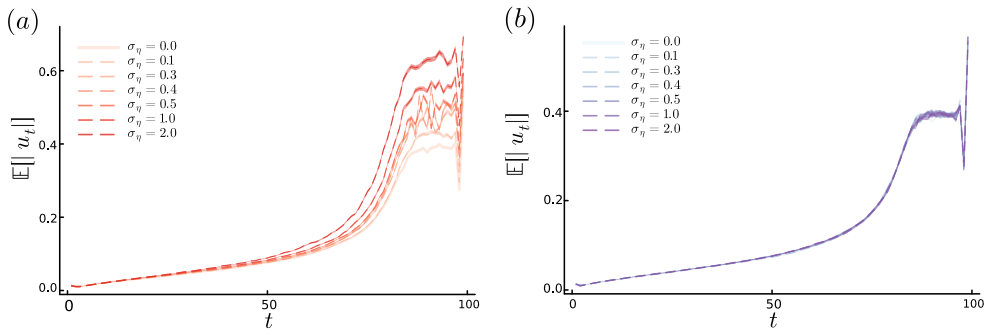

Figure 3: *Noise-dependent control magnitude in the two approaches.* **(a)** Expected control magnitude $|u_t|$, averaged over 10,000 realizations while varying internal noise σ_η in the Model Match framework (shaded areas indicate the standard error of the mean). **(b)** Same as (a), but for the Model Mismatch framework.

Perturbation Responses To further probe the distinction between the Model Match and Model Mismatch approaches, we simulated the 3D reaching task from Fig. 1 with a transient bump of magnitude $d = 2.0$ applied to the second component of x_t at $t = 20$, without reoptimizing. Both methods successfully compensate for the perturbation (Fig. 4a), as expected from their respective optimal solutions. Moreover, the behavioral output does not show visible qualitative differences across approaches (Fig. 4a). However, the internal dynamics diverge: in M-Mis, z_t shows a nonlinear, non-monotonic response with a slower return to baseline (Fig. 4b), strongly modulated by internal noise σ_η (Fig. 4c). In contrast, M-Match displays a Kalman-like profile, where z_t follows the perturbation magnitude and decays smoothly and monotonically (Fig. 4b), largely independent of noise (Fig. 4d). These findings suggest that M-Match and M-Mis could yield distinguishable neural signatures following perturbations, even when behavioral outputs remain similar.

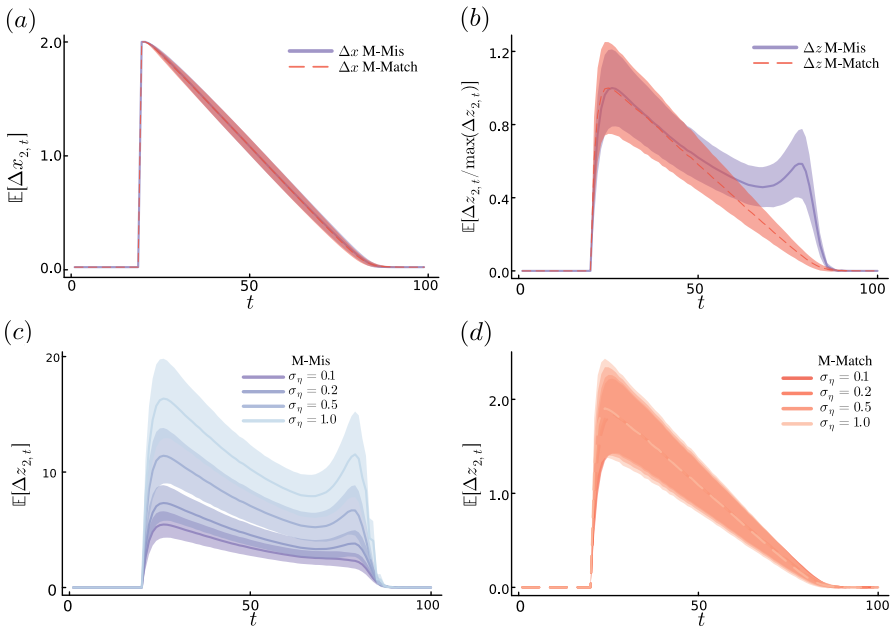

Figure 4: *Perturbation Responses in Model Match and Model Mismatch.* **(a)** Difference in the second component of the state (y-coordinate) between perturbed and unperturbed trials (same noise realization), averaged over 10,000 trials for the Model Match and Model Mismatch approaches, with $\sigma_\eta = 0.5$. **(b)**. Difference in the second component of the internal estimate between perturbed and unperturbed trials (same noise realization), averaged over 10,000 realizations for both approaches, normalized to their maximum, with $\sigma_\eta = 0.5$. **(c)**. Difference in the second component of the internal estimate between perturbed and unperturbed trials (same noise realization), averaged over 10,000, for the Model Mismatch approach at different levels of internal noise. **(d)**. Same as (c), but for the Model Match approach. In all panels, shaded areas indicate the standard error of the mean.

A.4.5 Neural Population Steering via Model Mismatch Control

We also apply our framework to a neural population steering task, where an unstable recurrent network is driven toward a target state via optimized linear readouts from another population – a setup reminiscent of biologically inspired machine learning methods (Jaeger & Haas, 2004; Maass et al., 2002; Sussillo & Abbott, 2009). This task connects to a growing body of work applying optimal control to neural population dynamics (Costa et al., 2024; Kao et al., 2021; Slijkhuis et al., 2023; Athalye et al., 2023), as well as related reinforcement learning approaches (Mastrogiuseppe & Moreno Bote, 2024).

Classical approaches (Todorov, 2005; Damiani et al., 2024) constrain the internal variable z_t to act as a Kalman filter estimate of x_t, enforcing the structural condition $W_t = A + BL_t - P_tH$ in Eq. 23 so that the dynamics of z_t match Eq. 7. In contrast, the Model Mismatch framework relaxes this constraint by allowing W_t to be freely optimized. This flexibility lets us treat z_t and x_t as distinct neural populations with independent connectivity matrices W and A (Fig. 5a).

Our algorithm (Appendix A.3.2) also supports partial optimization: for example, one can fix W and P (e.g., as random or biologically plausible) and optimize only L_t. Such setups are incompatible with the classic Model Match framework, which ties z_t's connectivity directly to x_t and forces W_t to vary over time, making it difficult to simulate realistic interactions between distinct neural populations.

We consider two populations of $N_{\text{units}} = 100$ linear neurons, each with sparse random connectivity. The recurrent connectivity within the first population (x_t) is given by

$$A_{ij} \sim \mathcal{N}\left(0, \frac{g_A}{\sqrt{N_{\text{units}}}}\right), \quad i, j = 1, \ldots, N_{\text{units}},$$

and similarly, the recurrent connectivity of the second population (z_t) is drawn from

$$W_{ij} \sim \mathcal{N}\left(0, \frac{g_W}{\sqrt{N_{\text{units}}}}\right), \quad i, j = 1, \ldots, N_{\text{units}} .$$

Note that internal dynamics is fixed over time, $W_{0,\ldots,T} = W$. The activity of the second population is linearly read out through a time-varying matrix L_t, which is optimized to steer the activity of the first population toward a desired target state while minimizing control effort (see Fig. 5a). The population z_t receives input from x_t through sparse random projections defined by

$$P_{ij} \sim \mathcal{N}\left(0, \frac{g_P}{\sqrt{N_{\text{units}}}}\right), \quad i, j = 1, \ldots, N_{\text{units}} .$$

Again we consider $P_{0,\ldots,T} = P$. To conform this setup to our control framework, we set $m = n = p = k = N_{\text{units}}$, and define

$$\begin{aligned} B = H &= I_{N_{\text{units}}} \\ D = \Sigma_\omega &= 0_{N_{\text{units}} \times N_{\text{units}}} . \end{aligned}$$

The cost and noise structure of the problem are defined by the following matrices

$$\begin{aligned} C &= \sigma_\varepsilon \cdot I_{N_{\text{units}}}, \\ \Sigma_\xi &= \sigma_\xi^2 \cdot I_{N_{\text{units}}}, \\ \Sigma_\eta &= \sigma_\eta^2 \cdot I_{N_{\text{units}}}, \\ Q_{1,\ldots,T-1} &= q_{<T} \cdot I_{N_{\text{units}}}, \\ Q_T &= q_T \cdot I_{N_{\text{units}}}, \\ R_t &= r \cdot I_{N_{\text{units}}}, \quad \text{for } t = 1, \ldots, T-1, \\ R_T &= 0 . \end{aligned}$$

The initial conditions are given by:

$$\begin{aligned} \mathbb{E}[x_1] &\sim \mathcal{N}\left(0, g_{x_1}^2 I_{N_{\text{units}}}\right), \\ \mathbb{E}[z_1] &\sim \mathcal{N}\left(0, g_{z_1}^2 I_{N_{\text{units}}}\right), \\ \Sigma_{x_1} &= 0_{N_{\text{units}} \times N_{\text{units}}}, \\ \Sigma_{z_1} &= 0_{N_{\text{units}} \times N_{\text{units}}} . \end{aligned}$$

The choice of Gaussian-distributed connectivity for the recurrent matrices A, W, and the feedforward matrix P is grounded in principles from dynamical mean-field theory, which describes the macroscopic behavior of large, sparsely connected networks of rate neurons (Sompolinsky et al., 1988; Rajan et al., 2010). We set $g_A = 1.1$ to ensure that the state dynamics in x_t are intrinsically unstable – this choice is deliberate, as our objective is to stabilize the system through control. Since we define the desired target state as zero, using it as a reference point, the initial condition effectively coincides with the goal. In this setting, a naturally decaying (stable) dynamics would trivially converge to the target without requiring active control. Instead, by inducing unstable dynamics, we create a scenario where control is essential to prevent divergence from the desired state. The internal dynamics gain $g_W = 0.9$ places the latent population z_t in a subcritical regime, supporting stable internal representations of the external dynamics. Lastly, the feedforward gain $g_P = 0.3$ models sparse and weak inter-population connectivity. These structured random matrices instantiate biologically inspired constraints that the Model Mismatch framework naturally accommodates while enabling effective control.

The parameters of the problem are listed in Table 7 (std = standard deviation).

Table 7: Parameters of the Neural Steering task

Name	Description	Value
T	Time steps	50
m	Dimension of state x_t	100
n	Dimension of internal state z_t	100
p	Dimension of observation y_t	100
k	Dimension of control u_t	100
r	Control cost scaling	0.001
$q_{<T}$	Task-related cost scaling	0.001
q_T	Task-related cost scaling	0.1
g_{x_1}	Initial condition scaling for x_1	10.0
g_{z_1}	Initial condition scaling for z_1	0.2
g_A	Scaling of random connectivity of population x_t	1.1
g_W	Scaling of random connectivity of population z_t	0.9
g_P	Scaling of random connections from population x_t to population z_t	0.3
σ_ξ	Std of dynamics noise ξ_t	0.5
σ_ε	Std of multiplicative control noise ε_t	0.0
σ_η	Std of additive internal noise η_t	0.2

Note that the "dynamics noise" ξ_t now represents the internal noise affecting the population x_t, analogous to the role of η_t for the population z_t. We also observe that the initial condition of the population z_t reflects spontaneous activity arising from internal fluctuations; accordingly, we set $g_{z_1} = \sigma_\eta$ to match the scale of this variability.

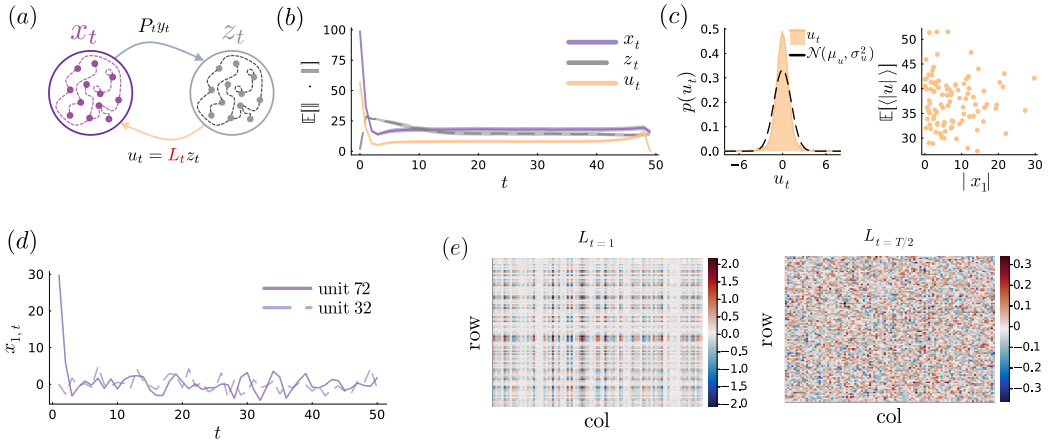

Figure 5: *Model Mismatch approach for Neural Population Steering*. **(a)** Sketch of the neural population steering task. **(b)** Average (over noise realizations) norm of x_t, z_t, and of the control signal $u_t = L_t z_t$ with error bars (standard error of the mean). **(c)** Distribution of the control signal over time and realizations with Gaussian fit (left), and average control magnitude (over time and realizations) received by each unit as a function of its initial absolute activity (right). **(d)** Activity of two units from the population vector x_t in a single trial. **(e)** Heatmaps of the matrices L_t at two time points: early (left) and mid-trial (right).

We optimize only the time-varying readout weights $L_{0,\dots,T}$ using the algorithm described in Sec. 4, keeping all other parameters fixed – Fig. 5a. By doing so, the activity x_t is successfully steered toward the target state (Fig. 5b) through a distributed control strategy: all units in the x population receive, on average, a similar amount of control, regardless of their initial distance from the target (Fig. 5c). Despite this relatively uniform distribution, the control effectively targets the units that require it most – namely, those starting furthest from zero, which represents the target in the current

reference frame (Fig. 5d). This selective modulation likely arises from the interaction between the recurrent dynamics in x and the structure of $L_{0,\dots,T}$. At the beginning of the trial, L_1 is highly structured and low-rank (Mastrogiuseppe & Ostojic, 2018), strongly directing activity toward the target. After the initial transient, $t \geq \tilde{t}$, $L_{\tilde{t},\dots,T}$ becomes sparse and high-rank, stabilizing the system and maintaining x_t near the target despite internal instability and noise (Fig. 5e). Interestingly, this result parallels findings in the control of recurrent networks through reinforcement learning (Mastrogiuseppe & Moreno Bote, 2024).

The Model Mismatch framework thus extends stochastic control beyond agent–environment formulations and enables the study of neural computation. For instance, in the simplified setting of this Sections, z_t can be interpreted as a premotor population driving a downstream motor population x_t toward a target – consistent with studies where premotor activity initializes motor cortex before movement (Kao et al., 2021; Logiaco et al., 2021). While not intended as an exhaustive biological mapping, this example illustrates how the framework can model complex strategies such as low-to-high rank transitions, selective modulation, and stabilization of unstable dynamics—phenomena that classical Model Match approaches cannot accommodate.

A.5 LLM USAGE

Large Language Models (LLMs) were used solely as a general-purpose writing assistant to improve clarity and language in some parts of this manuscript. They were not used for research ideation, derivations, analysis, or experiments. The authors take full responsibility for the content of the paper.