Solving Stochastic Control Under Multiplicative and Internal Noise via Constrained Optimization

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

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

Dmytro Grytskyy
Center for Brain and Cognition,
Department of Engineering
Pompeu Fabra University
Barcelona, ES
dmytro.grytskyy@upf.edu

Abstract

Multiplicative motor and observation noise, along with internal noise affecting computation, are central features of the sensorimotor system in biological and robotic agents. Yet, analytical solutions for stochastic optimal control are largely restricted to additive noise models and neglect internal noise. Here, we consider the problem of finding optimal control and filter laws for partially observable stochastic linear systems under quadratic costs with multiplicative control and observation noise, as well as internal noise. We provide an efficient, analytically derived coordinate-descent algorithm that computes mutually optimal linear control and filter laws by reformulating the problem as a constrained optimization. Our method guarantees monotonic decrease of the expected cost and convergence to a critical point, while addressing limitations of prior analytical approaches, which can yield suboptimal or incomplete solutions. Compared to state-of-the-art numerical methods, it achieves orders-of-magnitude computational speedups. Our solution also proves instrumental in revealing novel internal-noise-dependent, task-structured control strategies in redundant-arm tasks, even when internal noise does not dominate behavioral variability.

1 Introduction

Understanding the computational mechanisms that govern the sensorimotor system in humans and other animals is a long-standing goal [1, 2, 3, 4], with far-reaching applications in robotics and autonomous systems [5, 6, 7, 8]. Yet, developing mathematically tractable models that accurately capture these mechanisms remains an open problem. A principled mathematical framework for addressing this challenge is provided by partially observable Markov decision processes (POMDPs), which formalize decision-making under uncertainty [9]. Sensorimotor control naturally fits this formulation: the true state of the body and environment is not directly accessible and must be inferred

*These authors contributed equally.
†Corresponding author.

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

from noisy sensory feedback, with these estimates guiding actions that in turn shape uncertain future states and observations. In continuous-state settings, this framework is typically formulated within stochastic optimal control (SOC) theory, which has long served as a normative account of behavior in biological and robotic systems [10, 11, 1, 5, 6].

Under the assumptions of linear system dynamics, quadratic cost functions, and additive Gaussian noise affecting both state evolution and observations – the so-called Linear-Quadratic-Gaussian (LQG) problem [12] – SOC admits analytical solutions. Owing to its analytical tractability, the LQG framework has been extensively used to uncover computational principles of sensorimotor processing [10, 13, 14]. However, the assumption of purely additive noise limits the realism and scope of these models. Indeed, experimental evidence supports the presence of multiplicative noise in both motor execution [15, 16, 17, 18] and sensory feedback [19, 20, 18], as well as additive internal noise in neural processes, reflecting endogenous fluctuations in neural activity [21, 22, 23] or unreliability of internal computations [24]. Incorporating multiplicative and internal noise sources has proven essential for reproducing key signatures of human motor behavior [18, 25, 17, 26, 16]. Moreover, related noise models have also been adopted in robotics to capture actuator and sensor variability [27, 5].

A SOC framework incorporating the previously missing multiplicative and internal noise contributions – referred to here as the Linear-Quadratic-Multiplicative-Internal (LQMI) problem – was first introduced in the seminal work of [18]. This work also provided the first analytically derived solution for the LQMI problem, which has since been widely used from inverse optimal control to brain–machine interfaces [11, 14, 28, 29, 30, 31, 32]. As a result, the LQMI framework is often favored over the classical LQG formulation for modeling sensorimotor control.

Limitations of Existing Solutions Despite its historical relevance and wide applications, the solution derived in [18] relies on assumptions that are not guaranteed to hold in the general LQMI setting. In particular, recent work has demonstrated that it yields suboptimal policies in the presence of internal noise [33]. The suboptimality stems from the assumption of unbiased estimation and the associated reliance on the *orthogonality principle*, which does not hold at optimality when internal noise is present (see Appendix A.1.1). Related methods, including iterative LQG (iLQG) and differential dynamic programming (DDP), rely on similar assumptions [5].

To overcome these limitations, [33] proposed a gradient-based numerical algorithm that achieves state-of-the-art cost minimization under multiplicative and internal noise. However, this approach incurs a substantial computational burden as state, observation, control, and horizon dimensions increase, rendering it prohibitive for inverse optimal control – which requires solving many control problems across parameter settings [11] – and for large-scale applications. An analytical alternative, termed FPOMP [33], partially addresses these issues but is restricted either to one-dimensional systems or to higher-dimensional settings with purely additive noise (possibly including internal but not multiplicative noise), due to the complexity of the general case. Consequently, no existing method provides an analytically grounded and computationally efficient solution to the full LQMI problem.

Contributions In this paper, we analytically derive a coordinate-descent algorithm that computes mutually optimal linear control and filter laws for the LQMI problem. We reformulate SOC under partial observability as a constrained optimization problem within a Lagrangian framework, yielding tractable fixed-point equations. These equations define an iterative procedure with provable monotonic decrease of the expected cost and convergence to a critical point. Our method applies to the full LQMI setting in arbitrary dimensions, without imposing structural assumptions beyond linear filtering and linear control policies – assumptions standard in prior theoretical work [18]. Our algorithm improves upon previous analytical approaches [18] by eliminating unnecessary assumptions and achieving lower cost under internal noise. Reformulating the problem as constrained optimization over second-order moments avoids the gradient-propagation complexity that restricted the analytical method of [33] to one-dimensional systems. As a consequence, our algorithm achieves orders-of-magnitude speedups over state-of-the-art gradient-based numerical methods [33], while attaining the same cost-minimization performance, enabling scalable applications in high-dimensional systems and inverse optimal control. We also provide a formal proof that the orthogonality principle fails under internal noise – a phenomenon previously supported only by numerical evidence. Finally, when applied to redundant arm control tasks, our solution reveals novel internal-noise-dependent control strategies that were practically inaccessible with previous approaches, thereby elucidating new feedback control principles emerging from the modulation of internal state dynamics.

In the following, Sec. 2 formulates the stochastic optimal linear control problem with multiplicative and internal noise. Sec. 3 derives the convergent coordinate-descent algorithm and formalizes the breakdown of the orthogonality principle. Sec. 4 presents experiments comparing our approach with existing methods and revealing novel control strategies in structured sensorimotor tasks.

2 Stochastic Optimal Linear Control: Problem Formulation

We first review the classic LQG problem – referred to here as the Linear-Quadratic-Additive-Gaussian (LQAG) setting, as in [33] – and then, following [18], extend the noise model to include multiplicative control and observation noise, as well as additive internal noise, yielding the Linear–Quadratic–Multiplicative–Internal (LQMI) formulation. In the standard LQAG formulation, an agent receives noisy observations $y_t \in \mathbb{R}^k$ ($t = 0, 1, \dots, T$) 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 $u_t(y_{t-1}, \dots, y_0) \in \mathbb{R}^p$ 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, are 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 noise processes are uncorrelated in time and, in the LQAG framework, Gaussian (see Sec. 3.3 for a brief discussion on the Gaussianity assumption in our framework). We note that time-dependent matrices in the dynamics and noise can be trivially incorporated. The initial condition of the dynamics, x_0, is usually drawn from a Gaussian distribution. The control signal $u_t(y_{t-1}, \dots, y_0)$ 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, 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 [12] that combines a linear Kalman filter – providing optimal estimates $z_t \equiv \hat{x}_t$ of the partially observable state x_t – with a linear feedback controller of the form $u_t = L_t z_t$ (see Fig. 1a). Crucially, the control and filter gains are computed independently, a property known as the *separation principle* [12]. The internal variable then 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, with initial condition $z_0 = \mathbb{E}[x_0]$. Solving the optimal control problem therefore consists in computing both the optimal filter and control gains, respectively K_t and $L_t \in \mathbb{R}^{p\times m}$ (see Appendix A.1.2 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 [21, 22, 23, 3]. To consider a more general and realistic noise model, following [18], we first introduce control and observation multiplicative noise into the system and observation dynamics in Eqs. 1, 2. This leads to the modified equations

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

$$y_t = Hx_t + \omega_t + \sum\nolimits_j \rho_t^j D_j x_t \ . \tag{6}$$

In this framework, executing a control input u_t adds noise whose magnitude scales with the input itself [15, 16, 17], Eq. 5. Conversely, sensing the partially observable state x_t introduces sensory

noise whose magnitude scales with the state itself [19, 20], Eq. 6. The matrices $C_i \in \mathbb{R}^{m\times p}$, with $i = 1, \ldots, c$, and $D_j \in \mathbb{R}^{k\times m}$, with $j = 1, \ldots, d$, 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, all noise terms are assumed to be mutually independent.

Finding the optimal control signal $u_t(y_{t-1}, \ldots, y_0)$ 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. Following [18, 33], we assume that the control signal u_t can only linearly depend on the state 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 linear evolution 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.

Taken together, incorporating multiplicative and internal noise while assuming a linear filter for state estimation and a linear control policy (as in previous approaches [18, 33]) gives rise to the more general LQMI problem, fully defined by Eqs. 3, 5, 6, and 7. Solving this problem involves determining the optimal control $L_{0,\ldots,T}$ and filter $K_{0,\ldots,T}$ gains that minimize the quadratic cost function in Eq. 3 under the system, observation and estimate dynamics in Eqs. 5,6,7. Unlike in the classical LQAG setting, linearity of the optimal filter and controller is not guaranteed in the LQMI problem. Restricting both to linear forms, together with the quadratic cost, allows the dynamics of the second-order moments to close exactly, enabling the analytical tractability of our approach, as discussed in the next section. More general nonlinear or adaptive policies could potentially achieve lower costs, but would require different solution methods, potentially involving approximations. Notably, beyond the existence of finite second-order moments, we make no assumptions on the noise distributions.

Empirical Motivation and Interpretation of Internal Noise The internal-noise term η_t in Eq. 7 represents variability arising within the internal computations used for estimation, memory, and control, beyond noise affecting sensory inputs or motor execution [18]. Several lines of evidence motivate its inclusion. Neural computations are implemented by a biological substrate that exhibits variability at cellular, synaptic, and population levels [21, 34], while, more broadly, the reliability of internal representations may also depend on network state, external stimuli, attention, computational resources, memory demands, and alterations in cognitive or sensorimotor function [35, 36, 37, 38, 39, 40, 41, 42, 43]. Moreover, variability arising within neural processes can be behaviorally relevant [44, 23, 45]. As discussed by [18], human localization uncertainty approaches a nonzero asymptote despite repeated sensory observations [46], suggesting a continuously acting source of internal uncertainty. In memory-guided movements, behavioral variability can increase with the time over which information must be maintained, as observed in reaching, saccades, and grasping [47, 48, 49, 50], consistent with stochastic degradation of internal representations [51, 52]. Related evidence from perceptual and decision-making tasks indicates that variability in internal computations can substantially limit performance beyond sensory noise alone [53, 24]. Through the perception–action loop, these perceptual and decision-related effects may also influence sensorimotor behavior. Thus, η_t provides a tractable effective description of endogenous variability in internal computation, rather than a direct model of a specific physiological noise source. Its introduction does not imply that all neural variability is independent and Gaussian. Other factors, such as imperfect inference, may also contribute to behavioral variability without being equivalent to stochastic internal noise [54] (see Appendix A.3.2).

3 Solving the LQMI Problem via Constrained Optimization

We derive an iterative algorithm to compute control $L_{0,\ldots,T}$ and filter $K_{0,\ldots,T}$ gains for the LQMI problem that guarantees a monotone decrease of the cost in Eq. 3 and convergence to a critical point. This guarantee is nontrivial because optimization over gains is generally non-convex: even in the particular case of full state observability and zero dynamics noise (known as Linear-Quadratic Regulator) the objective is non-convex in feedback gains [55].

In the LQMI problem, a naive approach consisting in directly differentiating Eq. 3 with respect to the full gains sequence and solving the resulting stationarity conditions (see [33]) requires backprop-

agating how each gain affects all future state/estimate statistics, which is analytically prohibitive in high dimensions. We instead reformulate the problem in an expanded variable space by treating the second-order moments of (x_t, z_t) – sufficient statistics for the quadratic cost – as independent variables, and enforcing the moment dynamics as equality constraints in a Lagrangian formulation. This avoids explicit backpropagation and yields analytically tractable stationarity conditions in arbitrary dimension. The proposed method can be interpreted as an extension of Pontryagin's Maximum Principle (PMP), a classical variational framework for deterministic optimal control in which the system dynamics are imposed as constraints. In particular, PMP can be derived from a Lagrangian formulation in which the state dynamics are enforced via Lagrange multipliers [56]. Here, we extend the same constrained variational logic to the level of second-order moments (see Appendix A.1.7).

3.1 Lagrangian Formulation and Control-Filter Fixed-Point Equations

We first rewrite the cost in Eq. 3 as $C = \sum_{t=0}^{T} \left(\mathrm{tr}(Q_t S_t^{xx}) + \mathrm{tr}\big(L_t^\top R_t L_t S_t^{zz}\big)\right)$, where we assumed a linear control law $u_t = L_t z_t$ and we introduced the second-order moment matrices $S_t^{xx} = \mathbb{E}[x_t x_t^\top]$, $S_t^{zz} = \mathbb{E}[z_t z_t^\top]$, and $S_t^{xz} = \mathbb{E}[x_t z_t^\top]$. The expectation is taken with respect to the joint distribution $p_t(x_t, z_t)$ induced by previous control and filter gains and the initial distribution $p_0(x_0, z_0)$. To find the stationarity conditions for the control and filter gains, we introduce the Lagrangian objective

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

where Λ_t, Ω_t, and $\Gamma_t \in \mathbb{R}^{m\times m}$ are matrices of Lagrange multipliers. The constraints $G_t^{xx} = 0$, $G_t^{zz} = 0$, and $G_t^{xz} = 0$ enforce the correct temporal evolution of the second-order moments S_t^{xx}, S_t^{zz}, and S_t^{xz} between consecutive time steps, as implied by Eqs. 5,6,7 (see Eq. 33 in Appendix A.1.3),

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

where we have introduced the short-hand notation $M_t = A + BL_t - K_t H$, used repetitively, and the noise matrices $\Sigma_t^{xx} = \Sigma_\xi + \sum_i C_i L_t S_t^{zz} L_t^\top C_i^\top$ and $\Sigma_t^{zz} = \Sigma_\eta + K_t \Sigma_\omega K_t^\top + K_t \left(\sum_i D_i S_t^{xx} D_i^\top\right) K_t^\top$. Above, and from now on, 0 denotes a zero matrix of appropriate dimensions. Because the cost is quadratic in x_t and z_t and the moment dynamics close at second order, the second-order moments are sufficient statistics of the problem, and it suffices to enforce their temporal evolution constraints (i.e., the full joint distribution $p_t(x_t, z_t)$ need not be specified explicitly). For convenience, we set the terminal Lagrange multipliers to zero, $\Lambda_{T+1} = \Omega_{T+1} = \Gamma_{T+1} = 0$. The stationarity conditions $\partial C_{\mathcal{L}}/\partial L_t = 0$ and $\partial C_{\mathcal{L}}/\partial K_t = 0$ then yield (see Appendix A.1.4) the following fixed-point equations

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

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

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

$$E_t = 2R_t + B^\top (\tilde{\Lambda}_{t+1} + \tilde{\Omega}_{t+1} + \Gamma_{t+1} + \Gamma_{t+1}^\top) B + \sum\nolimits_i C_i^\top \tilde{\Lambda}_{t+1} C_i , \quad (12)$$

$$F_t = -B^\top (\tilde{\Lambda}_{t+1} A + \tilde{\Omega}_{t+1} K_t H + \Gamma_{t+1} A + \Gamma_{t+1}^\top K_t H) , \quad (13)$$

$$J_t = -B^\top (\tilde{\Omega}_{t+1} + \Gamma_{t+1}^\top)(A - K_t H) , \quad (14)$$

and

$$S_{AH} = (A + BL_t)(S_t^{zz} - (S_t^{xz})^\top) H^\top , \quad (15)$$

$$S_{LH} = \left(-A(S_t^{xx} - S_t^{xz}) + B L_t (S_t^{zz} - (S_t^{xz})^\top)\right) H^\top , \quad (16)$$

$$S_{HH} = H(S_t^{xx} + S_t^{zz} - S_t^{xz} - (S_t^{xz})^\top) H^\top + \Sigma_\omega + \sum\nolimits_i D_i S_t^{xx} D_i^\top . \quad (17)$$

Stationarity with respect to the moment variables, $\partial C_{\mathcal{L}}/\partial S_t^{xx} = \partial C_{\mathcal{L}}/\partial S_t^{zz} = \partial C_{\mathcal{L}}/\partial S_t^{xz} = 0$, yields the following backward equations for the Lagrange multipliers

$$\begin{aligned}
\Lambda_t &= Q_t + A^\top \Lambda_{t+1} A + H^\top K_t^\top \Omega_{t+1} K_t H + H^\top K_t^\top \Gamma_{t+1} A + \sum\nolimits_i D_i^\top K_t^\top \Omega_{t+1} K_t D_i \\
\Omega_t &= L_t^\top R_t L_t + L_t^\top B^\top \Lambda_{t+1} B L_t + M_t^\top \Omega_{t+1} M_t + M_t^\top \Gamma_{t+1} B L_t + \sum\nolimits_i L_t^\top C_i^\top \Lambda_{t+1} C_i L_t \\
\Gamma_t &= L_t^\top B^\top \tilde{\Lambda}_{t+1} A + M_t^\top \tilde{\Omega}_{t+1} K_t H + M_t^\top \Gamma_{t+1} A + L_t^\top B^\top \Gamma_{t+1}^\top K_t H \ . \qquad (18)
\end{aligned}$$

Given fixed control and filter gains, the Lagrange multipliers can be computed by backward recursion using the terminal conditions $\Lambda_{T+1} = \Omega_{T+1} = \Gamma_{T+1} = 0$. However, the fixed-point Eqs. 10, 11, 18 and 33 are mutually coupled, and solving them requires determining moments, gains and multipliers simultaneously. We bypass this difficulty by observing that Eqs. 10, 11, 18 and 33 naturally define an iterative *coordinate-descent procedure*, which is described next.

3.2 Iterative Algorithm and Monotonic Convergence to a Control-Filter Critical Point

Here we derive the main algorithm of the paper, hereafter referred to as the Lagrangian Constrained Optimization (LCO) algorithm, a coordinate-descent iterative scheme that yields a fixed-point pair of control and filter gains. To build the iterative algorithm to solve the fixed-point Eqs. 10,11,18,33 we first define the cost-to-go from time t (for $t = 0, \ldots, T$) starting from fixed values x and z as $C_t(x,z) = \mathrm{tr}(Q_t x x^\top + L_t^\top R_t L_t z z^\top) + \sum_{\tau=t+1}^{T} \mathbb{E}\left[x_\tau^\top Q_\tau x_\tau + u_\tau^\top R_\tau u_\tau\right]$. Here, the expectation is taken over future noise realizations, conditioned on $(x_t, z_t) = (x, z)$ and on fixed control and filter gains L_τ and K_τ for $\tau \geq t$. Note that the total cost in Eq. 3 can be expressed as the expected value of the cost-to-go from time $t = 0$, averaged over initial conditions as $C = \int p_0(x,z)\, C_0(x,z)\, dx\, dz$. The cost-to-go obeys the Bellman equation

$$C_t(x,z) = \mathrm{tr}(Q_t x x^\top + L_t^\top R_t L_t z z^\top) + \int dx' dz' C_{t+1}(x',z') p_{x,t+1}(x'|x,z) p_{z,t+1}(z'|x,z)\ , \quad (19)$$

where the transition probability densities $p_{x,t+1}(x'|x,z)$ and $p_{z,t+1}(z'|x,z)$ from time t to $t+1$ are induced by the dynamics in Eqs. 5, 6, and 7 under the control law $u_t = L_t z_t$. The Bellman equation 19 can be solved by backward induction, with terminal condition $C_T(x,z) = \mathrm{tr}\big(Q_T x x^\top + L_T^\top R_T L_T z z^\top\big)$. Exploiting the closure of the second-order moments in Eq. 33, we obtain the key result that the cost-to-go takes the form (see Appendix A.1.5)

$$C_t(x,z) = \mathrm{tr}(\Lambda_t x x^\top + \Omega_t z z^\top + \Gamma_t x z^\top) + \gamma_t\ , \quad (20)$$

where the coefficients Λ_t, Ω_t and Γ_t are actually the Lagrange multipliers in Eq. 18 with the same boundary conditions, and where γ_t can be recursively calculated, with boundary condition $\gamma_T = 0$, as

$$\gamma_t = \mathrm{tr}(\Lambda_{t+1}\Sigma_\xi + \Omega_{t+1} K_t \Sigma_\omega K_t^\top + \Omega_{t+1}\Sigma_\eta) + \gamma_{t+1}\ . \quad (21)$$

Because partial observability precludes any dependence of the control or filter gains on the true state x, we define the averaged cost-to-go by integrating $C_t(x,z)$ over the joint distribution $p_t(x,z)$ induced by the control and filter gains and the initial distribution $p_0(x,z)$

$$C_t = \int dx dz\ p_t(x,z) C_t(x,z) = \mathrm{tr}(\Lambda_t S_t^{xx} + \Omega_t S_t^{zz} + \Gamma_t S_t^{xz}) + \gamma_t\ , \quad (22)$$

using Eq. 20 and the definition of the second-order moments. The total cost in Eq. 3 can then be decomposed as $C = C_{<t} + C_t$, for any $0 < t < T$, where $C_{<t} = \sum_{\tau=0}^{t-1} \mathrm{tr}\big(Q_\tau S_\tau^{xx} + L_\tau^\top R_\tau L_\tau S_\tau^{zz}\big)$ is the cost up to time $t-1$. Importantly, $C_{<t}$ depends on $L_{0,\ldots,t-1}$ but not on L_t or later control gains. In turn, C_t contains the Lagrange multipliers and the second-order moments at time t. The Lagrange multipliers depend only on L_t and later control gains, $L_{t+1,\ldots,T}$, through Eq. 18. In contrast, the second-order moments at time t, which are computed from Eq. 33 given the fixed initial distribution $p_0(x,z)$, depend only on previous control gains, $L_{0,\ldots,t-1}$. Therefore, while keeping all other gains fixed, i.e., L_τ for $\tau \neq t$ and $K_{0,\ldots,T}$, we can locally optimize L_t by solving

$$L_t^* = \arg\min_{L_t} C = \arg\min_{L_t} (C_{<t} + C_t) = \arg\min_{L_t} C_t\ . \quad (23)$$

A global minimum exists for this local optimization problem since the cost-to-go is a non-negative quadratic function of L_t. Indeed, the Lagrange multipliers depend quadratically on the control gains L_t (see Eq. 18). The minimization yields the same expression as the fixed-point equation Eq. 10,

$$L_t^* = E_t^{-1}\left(F_t S_t^{xz} (S_t^{zz})^{-1} + J_t\right)\ . \quad (24)$$

Let $L_{0,\dots,T}$ and $K_{0,\dots,T}$ denote the control and filter gains before the update in Eqs. 23,24. Since L_t^* minimizes C_t with all other gains fixed, the total cost satisfies

$$C^* \equiv C(L_0, \dots, L_{t-1}, L_t^*, L_{t+1}, \dots, L_T) \;\le\; C(L_0, \dots, L_{t-1}, L_t, L_{t+1}, \dots, L_T) \equiv C. \quad (25)$$

Therefore, the cost decreases or remains unchanged after the local optimization. After this update, the cost can be expressed as

$$C^* = C_{<t} + \mathrm{tr}\big(Q_t S_t^{xx} + L_t^{*\top} R_t L_t^* S_t^{zz}\big) + \mathrm{tr}\big(\Lambda_{t+1} S_{t+1}^{xx,*} + \Omega_{t+1} S_{t+1}^{zz,*} + \Gamma_{t+1} S_{t+1}^{xz,*}\big) + \gamma_{t+1}, \quad (26)$$

where the updated moments $S_{t+1}^{ab,*}$, $ab \in \{xx, zz, xz\}$, are obtained by propagating the moments from time t to $t+1$ using Eq. 33 with the optimal gain L_t^* of Eq. 24. The Lagrange multipliers from time $t+1$ onward remain unchanged, since they do not depend on L_t (see Eq. 18), the only control gain updated at this step.

By redefining L_t^* as L_t and $S_{t+1}^{ab,*}$ as S_{t+1}^{ab}, we can repeat the same argument for $t+1$ and proceed sequentially up to T. Therefore, starting from gain sequences $L^{(n)} \equiv L_{0,\dots,T}^{(n)}$ and $K^{(n)} \equiv K_{0,\dots,T}^{(n)}$, this yields an updated sequence $L^{(n+1)}$ satisfying $C\big(L^{(n+1)}, K^{(n)}\big) \;\le\; C\big(L^{(n)}, K^{(n)}\big)$. After the forward sweep, the Lagrange multipliers can be recomputed by a backward recursion using Eq. 18. We refer to this forward control gain sweep, followed by the backward multiplier sweep, as a *control pass*. Hence, each control pass monotonically decreases the objective or leaves it unchanged.

An analogous argument applies to the filter gains. Fixing all control and all filter gains except K_t, we minimize C_t with respect to K_t, which yields

$$K_t^* = \arg\min_{K_t} C_t = \Big(S_{AH} + \tilde{\Omega}_{t+1}^{-1} \Gamma_{t+1} S_{LH}\Big)\, S_{HH}^{-1}, \quad (27)$$

identical to the fixed-point Eq. 11. As in the control case, replacing K_t with K_t^* does not increase the objective, since C_t is a non-negative quadratic function of K_t. Proceeding sequentially for $t = 0, \dots, T$ yields an updated sequence $K^{(n+1)}$ satisfying $C\big(L^{(n+1)}, K^{(n+1)}\big) \;\le\; C\big(L^{(n+1)}, K^{(n)}\big)$. Analogously to the control pass, we define a *filter pass* as a forward sweep over the filter gains followed by a backward sweep over the Lagrange multipliers. Therefore, each *filter pass* monotonically decreases the cost or leaves it unchanged.

Thus, starting from arbitrary initial gains $L^{(0)}$ and $K^{(0)}$ and any initial distribution $p_0(x, z)$, alternating control and filter passes yields a non-increasing sequence of cost values, $C(L^{(0)}, K^{(0)}) \ge C(L^{(1)}, K^{(0)}) \ge C(L^{(1)}, K^{(1)}) \ge \cdots \ge C(L^{(n+1)}, K^{(n)}) \ge C(L^{(n+1)}, K^{(n+1)}) \ge \cdots \ge 0$. Since the cost is non-negative, this sequence converges. At convergence, the gains satisfy the fixed-point Eqs. 10, 11, 18, and 33. Therefore, the limiting gains $L^* = L^{(\infty)}$ and $K^* = K^{(\infty)}$ constitute a fixed point of the Lagrangian in Eq. 8, and hence a critical point of the cost in Eq. 3. We have thus proven the following

Theorem 1. *Starting from arbitrary initial gains $L^{(0)}$ and $K^{(0)}$, and any initial distribution $p_0(x, z)$, the coordinate-descent algorithm defined by alternating control and filter passes (see Algorithm 1) converges to a pair of gains (L^*, K^*) satisfying $C(L^*, K^*) \le C(L^{(0)}, K^{(0)})$ and constituting a critical point of the cost function in Eq. 3.*

3.3 Breakdown of the Orthogonality Principle With Internal Noise and Other Notes

As discussed in Sec. 1 (see also Appendix A.1.1 for more details), the seminal work of [18] addressed the LQMI problem by assuming unbiased estimation, which implicitly enforces the orthogonality principle (OP) [18, 12]. While OP is implied by unbiasedness, the converse does not hold. Unbiasedness was empirically shown to fail in [33], whereas the weaker OP condition was empirically found to yield optimal solutions only in the absence of internal noise. Here, we formally characterize this phenomenon by proving that OP is sufficient to obtain a critical point of the cost function if and only if internal noise is absent, thereby providing a rigorous justification for its conditional validity in the LQMI problem. In particular, the following theorem holds.

Theorem 2. *Assume an initial distribution $p_0(x, z)$ such that $S_0^{xz} = S_0^{zz}$. Then the stationarity equations Eqs. 10, 11, and 18 admit a solution satisfying the orthogonality principle $S_t^{xz} = S_t^{zz}$ for all $t = 1, \dots, T$ if and only if internal noise is zero, i.e. $\Sigma_\eta = 0$. Moreover, this solution corresponds to a critical point of the cost in Eq. 3.*

See Appendix A.1.6 for the proof. We emphasize that the initial condition $S_0^{xz} = S_0^{zz}$, ensuring that the orthogonality principle holds at the initial time, is required only for Theorem 2. Algorithm 1 holds for arbitrary valid initial second-order moments. In Appendix A.1.8, we further show that, in the absence of multiplicative and internal noise, the iterative procedure defined in Sec. 3.2 recovers the classical LQAG solution when OP is enforced.

Convergence Properties Because the LQMI problem is non-convex, the fixed-point equations in Sec. 3.1 may admit multiple solutions. In contrast to previous approaches [18, 33], our algorithm provides formal guarantees of monotonic cost improvement and convergence to a critical point. However, the critical point found might not be the global optimum, and the obtained solution may depend on initialization (see Appendices A.2.2 and A.3.1).

Robustness to Non-Gaussian Noise As shown in Eqs. 18, 24, 27, and 33, the algorithm depends only on the first and second moments of the noise processes. No assumptions beyond finite second moments are required, and the method therefore applies to arbitrary noise distributions with finite covariance. In Appendix A.3.4 we empirically validate this robustness using non-Gaussian noise.

4 Experiments

In what follows, we denote the gradient-based method of [33] by GD, the iterative method of [18] by TOD, and the algorithm derived in Sec. 3 by LCO (Lagrangian Constrained Optimization). As discussed in Sec. 1, GD provides state-of-the-art numerical solutions for LQMI cost minimization, whereas TOD is the only currently available analytically derived iterative algorithm that scales efficiently to the full LQMI setting, despite being suboptimal in the general case.

Computational Efficiency with State-of-the-Art Cost Minimization To assess the optimality and computational efficiency of LCO, we consider a family of LQMI problems of varying dimensionality, with $m = p = k = n_{\text{task}}$ and $n_{\text{task}} \in [1, 100]$. For each instance, the matrices A, B, C, and D are randomly generated as described in Appendix A.3.1, with diagonal cost matrices. For each dimension, we generate 10 random instances with trial duration $T = 10$ and initialize both LCO and GD from the corresponding TOD solutions for a fair comparison.

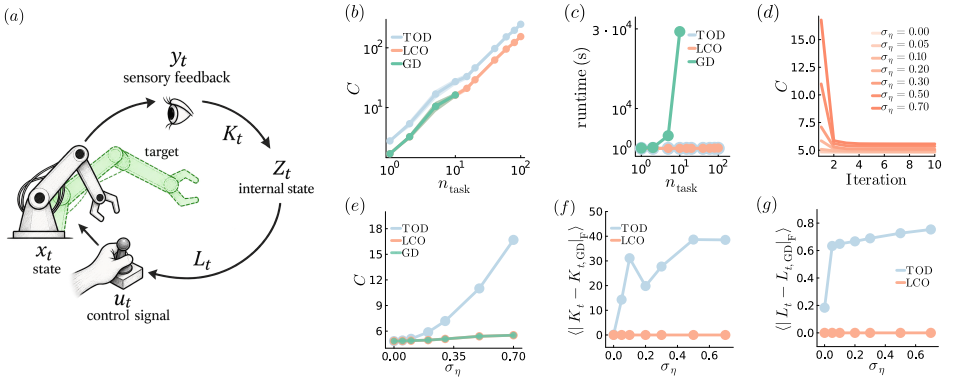

Figure 1: *Computational Efficiency with State-of-the-Art Cost Minimization.* **(a)** Perception–action loops. The mechanical arm illustrates the controlled body or plant, whose state x_t generates noisy sensory feedback y_t that is integrated into an internal state z_t through the filter gain K_t; the control gain L_t then maps z_t to the control signal u_t to guide behavior toward a target state. **(b)** Expected cost C vs. task dimensionality n_{task} for random tasks, for TOD (blue), LCO (red), and GD (green); points are means over 10 random instances, shading denotes SEM – GD is shown only for the dimensions where it was run, since extending it to larger dimensions leads to prohibitive runtimes. **(c)** Mean runtime vs. n_{task} for the same benchmark. **(d)** LCO objective C across iterations for a fixed random problem ($n_{\text{task}} = 5$) and different σ_η, initialized from the corresponding TOD solutions. Each iteration is one control or filter update. **(e)** Expected cost C for the same problem vs. σ_η for TOD, LCO, and GD initialized at the LCO solution; shading denotes SEM over 1000 Monte Carlo trials. **(f)** Time-averaged Frobenius distance between the filter gains found by TOD or LCO and the corresponding GD solution, vs. σ_η. **(g)** Same as (f), for the control gains.

Across dimensions, LCO consistently outperforms TOD in the presence of internal noise (Fig. 1b) and, on the subset of dimensions where GD is computationally feasible, reaches the same final cost with substantially lower runtime (Fig. 1c). This computational advantage is expected to grow further with task dimensionality and trial duration, since GD optimizes over $mp(T-1)+mk(T-2)$ parameters. Additional benchmarks under limited sensory observations further confirm the performance advantage of LCO over TOD and its computational efficiency compared with GD (Appendix A.3.1). In a single random task and varying internal noise levels, the LCO objective decreases monotonically across optimization steps (Fig. 1d), consistent with its theoretical guarantees (Sec. 3). The performance gap between LCO and TOD widens as internal noise increases (Fig. 1e), reflecting the growing suboptimality of TOD in this regime due to its suboptimal filter and control gains (Figs. 1f,g). When initialized at the LCO solution, GD neither further reduces the objective nor changes the solution numerically (Figs. 1e-g), consistent with LCO reaching a stationary point of the cost in Eq. 3 (Sec. 3). To empirically assess the local structure of the solutions found by LCO, we perturb the LCO solution along multiple random directions in L and K spaces, over increasing perturbation amplitudes, and neighboring solutions with lower cost were not found. Additional initialization-sensitivity tests show that LCO also reaches closely comparable low-cost solutions from substantially different initializations (see Appendix A.3.1 for details).

Optimal Control in a Reaching Task We next examine the LCO algorithm in a more empirically grounded motor task: a reaching movement with a psychophysically informed calibration of the noise sources. We consider a simplified three-joint planar arm, whose joints can be interpreted as the shoulder, elbow, and wrist, driven by nine muscle-like actuators. This yields a redundant actuation architecture characteristic of biological and robotic musculoskeletal systems [57], with noisy sensory feedback of the three joint angles. For this task, with $m = 6$, $p = 9$, $k = 3$, and 40 control steps, GD would require optimizing nearly 3×10^3 parameters, making it computationally prohibitive for a systematic analysis across the full internal-noise sweep. Taking the target posture as the origin, the elbow starts 20° away from its target, while the shoulder and wrist are initially at their target positions. The task consists of reaching the target elbow angle within 0.40 s, while keeping the other joints near their targets (Fig. 2a). This setup is motivated by experimental paradigms involving single-joint elbow movements of comparable amplitude [58], while the movement duration is consistent with those reported for short human reaching movements [59].

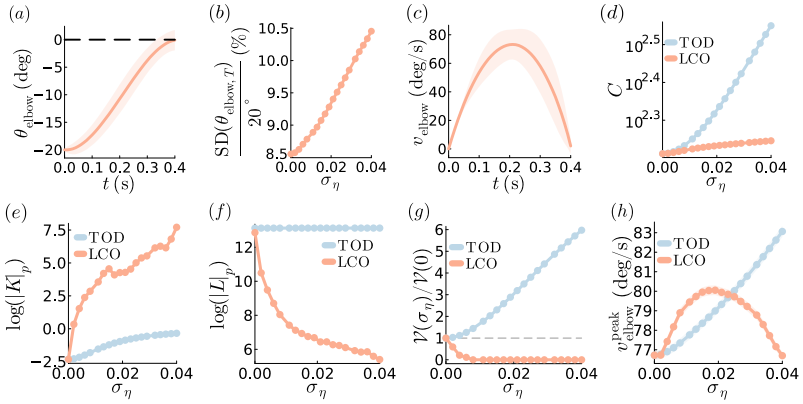

Figure 2: *Optimal Control in a Reaching Task*. **(a)** Mean elbow-angle trajectory for LCO at $\sigma_\eta = 0.019$ during a 20° target-directed movement; shading denotes SD across trials, and the dashed line indicates the target angle. **(b)** Terminal elbow-angle SD, normalized by the 20° movement amplitude and expressed as a percentage, vs. internal noise σ_η for LCO. **(c)** Mean elbow angular-velocity trajectory for LCO at $\sigma_\eta = 0.019$; shading denotes SD across trials. **(d)** Expected cost C vs. σ_η for TOD (blue) and LCO (red), computed analytically. **(e)** Time-averaged log-pseudodeterminant of the filter gains vs. σ_η, computed by summing the natural logarithms of singular values of K_t exceeding 10^{-6}. **(f)** Same as (e), for the control gains L_t. **(g)** Control variance relative to internal-state variance, quantifying the transmission of internal-state variability to the control signal, with $\mathcal{V}(\sigma_\eta) = \frac{\sum_t \mathrm{Tr}[L_t\,\mathrm{Cov}(z_t)L_t^\top]}{\sum_t \mathrm{Tr}[\mathrm{Cov}(z_t)]}$. The plotted quantity is $\mathcal{V}(\sigma_\eta)/\mathcal{V}(\sigma_\eta = 0)$. **(h)** Mean peak absolute elbow angular velocity vs. σ_η; shading denotes SEM across trials. Panels (a)–(c), (g), and (h) are based on 2000 Monte Carlo trials.

Internal noise is varied over $\sigma_\eta \in [0, 0.04]$, keeping the other noise amplitudes fixed. This range was selected to maintain experimentally plausible movement variability while ensuring that internal noise does not dominate the direct contributions to behavioral variability (see Appendix A.3.2 for further details on the model, cost function, and noise calibration). Across the sweep, the terminal elbow-angle SD remains between 8.5% and 10.5% of the movement amplitude (Fig. 2b), broadly consistent with the relative endpoint variability reported for human reaching movements [59]. The resulting trajectories exhibit smooth, approximately bell-shaped angular-velocity profiles (Fig. 2c), with peak speeds ($\sim$ 77–$80^\circ/\mathrm{s}$) of the same order as those reported for human elbow movements [25, 60]. As in the random tasks, LCO achieves lower expected costs than TOD, with the advantage increasing with internal noise (Fig. 2d), reducing the cost by 50% at $\sigma_\eta = 0.04$. As internal computation becomes less reliable due to the effect of η_t, the LCO solution places increasing weight on sensory feedback (Fig. 2e), while rescaling control gains to mitigate the impact of internal fluctuations on motor commands (Fig. 2f), yielding solutions that differ from those of TOD. These differences reflect distinct computational strategies: as internal noise increases, LCO reduces the transmission of internal-state variability to the control signal, whereas TOD exhibits the opposite trend (Fig.2g).

These differences also translate into distinct behavioral predictions. For instance, LCO predicts a non-monotonic dependence of peak elbow angular velocity on internal noise (Fig.2h), whereas TOD predicts a monotonic increase. Complementary analyses of a longer-horizon redundant-arm stabilization task show similar adjustments in optimal control and estimation strategies and a persistent performance advantage of LCO over TOD, including under limited sensory observations (Appendix A.3.3). Although imperfect inference may also contribute to behavioral variability [54], it can induce qualitatively different control and estimation strategies, as well as distinct behavioral predictions compared to internal noise (see Appendix A.3.2 for the case of sensory reliability misestimation). Together, these results show that LCO reveals computational strategies and behavioral predictions not captured by previous analytical approaches, highlighting how internal noise substantially reshapes optimal control strategies even when its direct contribution to movement variability is not dominant.

5 Conclusions

By reformulating stochastic optimal control as a constrained optimization problem, we derive a coordinate-descent iterative algorithm with provable monotonic convergence to a critical point. The resulting method solves optimal control problems under a general noise model with both multiplicative and internal noise, assuming linear control and quadratic costs – namely the LQMI problem. This extends previous analytical approaches, which remained incomplete [18, 33]. Moreover, we provided a formal analytical characterization of the failure of the orthogonality principle in the presence of internal noise, a result that had previously been supported only by numerical evidence. Our algorithm outperforms state-of-the-art gradient-based methods [33] by more than three orders of magnitude in computational efficiency on multidimensional control tasks, while achieving the same final cost. It also yields lower-cost solutions than currently available iterative schemes for the full LQMI setting [18]. This computational advantage is critical for applying stochastic optimal control to real-world problems, particularly in inverse optimal control [11]. When applied to redundant sensorimotor tasks, our method reveals novel internal-noise-dependent control strategies that previous methods could not access, even when internal noise does not dominate behavioral variability.

Overall, our work provides an analytically grounded and computationally efficient solution to stochastic optimal control under multiplicative and internal noise, and shows how such a solution can uncover new computational principles in structured sensorimotor control.

Limitations and Future Work Our work assumes linear dynamics, linear filtering and control, and quadratic costs, which enable closed-form moment dynamics and analytical tractability, but do not capture nonlinear systems. Linear and locally linear approximations are standard and widely used in both neuroscience and robotics [18, 10, 3, 2, 27], supporting the relevance of our novel solutions. Furthermore, our framework can be naturally extended in combination with iterative LQG (iLQG) and differential dynamic programming (DDP) methods [5, 6, 61, 62], which approximate nonlinear control problems through local linearization and quadratic value-function expansions. Integrating our analytical solution into such schemes may enable efficient stochastic optimal control under realistic noise in nonlinear systems.

Acknowledgments and Disclosure of Funding

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

References

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

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

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

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

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

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

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

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

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

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

[11] 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.

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

[13] Jörn Diedrichsen, Reza Shadmehr, and Richard B Ivry. The coordination of movement: optimal feedback control and beyond. *Trends in cognitive sciences*, 14(1):31–39, 2010.

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

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

[16] 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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

[34] Alfonso Renart and Christian K Machens. Variability in neural activity and behavior. *Current opinion in neurobiology*, 25:211–220, 2014.

[35] Mark M Churchland, Byron M Yu, John P Cunningham, Leo P Sugrue, Marlene R Cohen, Greg S Corrado, William T Newsome, Andrew M Clark, Paymon Hosseini, Benjamin B Scott, et al. Stimulus onset quenches neural variability: a widespread cortical phenomenon. *Nature neuroscience*, 13(3):369–378, 2010.

[36] Marlene R Cohen and John HR Maunsell. Attention improves performance primarily by reducing interneuronal correlations. *Nature neuroscience*, 12(12):1594–1600, 2009.

[37] Paul M Bays and Masud Husain. Dynamic shifts of limited working memory resources in human vision. *Science*, 321(5890):851–854, 2008.

[38] Sebastian Schneegans and Paul M Bays. Drift in neural population activity causes working memory to deteriorate over time. *The Journal of Neuroscience*, 38(21):4859–4869, 2018.

[39] Klaudia Grechuta, Laura Ulysse, Belén Rubio Ballester, and Paul FMJ Verschure. Self beyond the body: action-driven and task-relevant purely distal cues modulate performance and body ownership. *Frontiers in human neuroscience*, 13:91, 2019.

[40] Adrián F Amil, Belén Rubio Ballester, Martina Maier, and Paul FMJ Verschure. Chronic use of cannabis might impair sensory error processing in the cerebellum through endocannabinoid dysregulation. *Addictive Behaviors*, 131:107297, 2022.

[41] Chrysanthi Blithikioti, Laia Miquel, Blanca Paniello, Laura Nuno, Antoni Gual, Belen Rubio Ballester, Adrian Fernandez, Ivan Herreros, Paul Verschure, and Mercedes Balcells-Olivero. Chronic cannabis use affects cerebellum dependent visuomotor adaptation. *Journal of Psychiatric Research*, 156:8–15, 2022.

[42] Belén Rubio Ballester, Martina Maier, Rosa María San Segundo Mozo, Victoria Castañeda, Armin Duff, and Paul F MJ Verschure. Counteracting learned non-use in chronic stroke patients with reinforcement-induced movement therapy. *Journal of neuroengineering and rehabilitation*, 13(1):74, 2016.

[43] Jin H Yan, Susan Rountree, Paul Massman, Rachelle Smith Doody, and Hong Li. Alzheimer's disease and mild cognitive impairment deteriorate fine movement control. *Journal of Psychiatric Research*, 42(14):1203–1212, 2008.

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

[45] Vahid Rostami, Thomas Rost, Felix Johannes Schmitt, Sacha Jennifer van Albada, Alexa Riehle, and Martin Paul Nawrot. Spiking attractor model of motor cortex explains modulation of neural and behavioral variability by prior target information. *Nature Communications*, 15(1):6304, 2024.

[46] Emanuel Vassilev Todorov. *Studies of goal directed movements*. PhD thesis, Massachusetts Institute of Technology, 1998.

[47] Immo Schütz, Denise YP Henriques, and Katja Fiehler. No effect of delay on the spatial representation of serial reach targets. *Experimental Brain Research*, 233(4):1225–1235, 2015.

[48] Janis M White, David L Sparks, and Terrence R Stanford. Saccades to remembered target locations: an analysis of systematic and variable errors. *Vision research*, 34(1):79–92, 1994.

[49] Constanze Hesse and Volker H Franz. Grasping remembered objects: Exponential decay of the visual memory. *Vision Research*, 50(24):2642–2650, 2010.

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

[51] RA Kinchla and F Smyzer. A diffusion model of perceptual memory. *Perception & psychophysics*, 2(6):219–229, 1967.

[52] Albert Compte, Nicolas Brunel, Patricia S Goldman-Rakic, and Xiao-Jing Wang. Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cerebral cortex*, 10(9):910–923, 2000.

[53] Barbara Anne Dosher and Zhong-Lin Lu. Perceptual learning reflects external noise filtering and internal noise reduction through channel reweighting. *Proceedings of the National Academy of Sciences*, 95(23):13988–13993, 1998.

[54] Jeffrey M Beck, Wei Ji Ma, Xaq Pitkow, Peter E Latham, and Alexandre Pouget. Not noisy, just wrong: the role of suboptimal inference in behavioral variability. *Neuron*, 74(1):30–39, 2012.

[55] Maryam Fazel, Rong Ge, Sham Kakade, and Mehran Mesbahi. Global convergence of policy gradient methods for the linear quadratic regulator. In *International conference on machine learning*, pages 1467–1476. PMLR, 2018.

[56] Emanuel Todorov et al. Optimal control theory. *Bayesian brain: probabilistic approaches to neural coding*, pages 268–298, 2006.

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

[58] Rodrigo S Maeda, Tyler Cluff, Paul L Gribble, and J Andrew Pruszynski. Feedforward and feedback control share an internal model of the arm's dynamics. *The Journal of Neuroscience*, 38(49):10505–10514, 2018.

[59] Robert J Van Beers, Patrick Haggard, and Daniel M Wolpert. The role of execution noise in movement variability. *Journal of neurophysiology*, 91(2):1050–1063, 2004.

[60] Paul L Gribble and David J Ostry. Compensation for interaction torques during single-and multijoint limb movement. *Journal of neurophysiology*, 82(5):2310–2326, 1999.

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

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

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

[64] Robert J Van Beers, Anne C Sittig, and Jan J Denier van der Gon. The precision of proprioceptive position sense. *Experimental brain research*, 122(4):367–377, 1998.

[65] Chris R Sims, Robert A Jacobs, and David C Knill. Adaptive allocation of vision under competing task demands. *The Journal of Neuroscience*, 31(3):928–943, 2011.

[66] Etienne Burdet, Rieko Osu, David W Franklin, Theodore E Milner, and Mitsuo Kawato. The central nervous system stabilizes unstable dynamics by learning optimal impedance. *Nature*, 414(6862):446–449, 2001.

[67] Frédéric Crevecoeur, Isaac Kurtzer, and Stephen H Scott. Fast corrective responses are evoked by perturbations approaching the natural variability of posture and movement tasks. *Journal of neurophysiology*, 107(10):2821–2832, 2012.

[68] Frédéric Crevecoeur and Stephen H Scott. Priors engaged in long-latency responses to mechanical perturbations suggest a rapid update in state estimation. *PLoS computational biology*, 9(8):e1003177, 2013.

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

[70] Isaac L Kurtzer, J Andrew Pruszynski, and Stephen H Scott. Long-latency reflexes of the human arm reflect an internal model of limb dynamics. *Current Biology*, 18(6):449–453, 2008.

[71] Teige C Bourke, Angela M Coderre, Stephen D Bagg, Sean P Dukelow, Kathleen E Norman, and Stephen H Scott. Impaired corrective responses to postural perturbations of the arm in individuals with subacute stroke. *Journal of NeuroEngineering and Rehabilitation*, 12(1):7, 2015.

[72] Catherine R Lowrey, Teige C Bourke, Stephen D Bagg, Sean P Dukelow, and Stephen H Scott. A postural unloading task to assess fast corrective responses in the upper limb following stroke. *Journal of NeuroEngineering and Rehabilitation*, 16(1):16, 2019.

A Appendix

Code Availability

The implementation of the proposed algorithm, together with the scripts and numerical data supporting the experiments, is publicly available at `https://github.com/francescodamiani/LQMI-Constrained-Optimization`. The repository also provides a regularized version of the algorithm to mitigate numerical instabilities arising from the inversion of singular or ill-conditioned matrices when solving stochastic optimal control problems.

A.1 Supplementary Theoretical Material

A.1.1 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). The influential work of [18] 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, [33] 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 in terms of cost minimization the original approach under internal noise.

The reason the method in [18] still performs optimally when internal noise is absent is that unbiasedness implies the *orthogonality principle* [12, 33], 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 [18] in the zero internal noise regime stems from its implicit reliance on orthogonality, which breaks down otherwise.

In Appendix A.1.6, 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 [33]. Moreover, in Appendix A.1.8, we demonstrate that the orthogonality principle actually leads to the global optimum for the classic LQAG problem.

A.1.2 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 – are given by [63, 12, 18]

$$L_t = -(2R_t + B^\top S_{t+1} B)^{-1} B^\top S_{t+1} A \tag{28}$$

$$S_t = 2Q_t + A^\top S_{t+1}(A + BL_t) \tag{29}$$

$$K_t = A\Sigma_t^e H^\top (H\Sigma_t^e H^\top + \Sigma_\omega)^{-1} \tag{30}$$

$$\Sigma_{t+1}^e = \Sigma_\xi + (A - K_t H)\Sigma_t^e A^\top \, . \tag{31}$$

A detailed derivation can be found in [63], Chapter 12, Sections 4 and 5. We observe that the only differences with the Eqs. in [63] 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.1.8, we prove that the solutions derived in Sec. 3 recover these classical results in the absence of multiplicative and internal noise.

A.1.3 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

$$x_{t+1} = Ax_t + BL_t z_t + \xi_t + \sum\nolimits_i \varepsilon_t^i C_i L_t z_t$$
$$z_{t+1} = M_t z_t + K_t H x_t + \eta_t + K_t \omega_t + K_t \sum\nolimits_i \rho_t^i D_i x_t \tag{32}$$

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

The 2nd-order moments at time $t+1$ can be computed based on those in the previous time step t by using the appropriate averages and interactions between terms in Eqs. 32. The result is

$$\begin{aligned}
S_{t+1}^{xx} &= AS_t^{xx}A^\top + AS_t^{xz}L_t^\top B^\top + BL_t(S_t^{xz})^\top A^\top + BL_t S_t^{zz} L_t^\top B^\top + \Sigma_t^{xx} \\
S_{t+1}^{zz} &= K_t H S_t^{xx} H^\top K_t^\top + K_t H S_t^{xz} M_t^\top + M_t (S_t^{xz})^\top H^\top K_t^\top + M_t S_t^{zz} M_t^\top + \Sigma_t^{zz} \\
S_{t+1}^{xz} &= AS_t^{xx}H^\top K_t^\top + BL_t S_t^{zz} M_t^\top + AS_t^{xz} M_t^\top + BL_t (S_t^{xz})^\top H^\top K_t^\top \ .
\end{aligned} \tag{33}$$

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

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

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

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. 32. 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. 33 by their corresponding non-averaged x and z as

$$\begin{aligned}
\hat{S}_{t+1}^{xx} &= Axx^\top A^\top + Axz^\top L_t^\top B^\top + BL_t zx^\top A^\top + BL_t zz^\top L_t^\top B^\top + \hat{\Sigma}_t^{xx} \\
\hat{S}_{t+1}^{zz} &= K_t H xx^\top H^\top K_t^\top + K_t H xz^\top M_t^\top + M_t zx^\top H^\top K_t^\top + M_t zz^\top M_t^\top + \hat{\Sigma}_t^{zz} \\
\hat{S}_{t+1}^{xz} &= Axx^\top H^\top K_t^\top + BL_t zz^\top M_t^\top + Axz^\top M_t^\top + BL_t zx^\top H^\top K_t^\top \ .
\end{aligned} \tag{34}$$

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

A.1.4 Fixed-Point Equations for Control and Filter

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

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

$$\begin{aligned}
\frac{\partial C_{\mathcal{L}}}{\partial K_t} &= \left[\tilde{\Omega}_{t+1} K_t H + \Gamma_{t+1} A\right] S_t^{xx} H^\top - \left[\tilde{\Omega}_{t+1} M_t + \Gamma_{t+1} BL_t\right] S_t^{zz} H^\top - \tilde{\Omega}_{t+1} K_t H S_t^{xz} H^\top \\
&\quad + \tilde{\Omega}_{t+1} M_t (S_t^{xz})^\top H^\top - \Gamma_{t+1} A S_t^{xz} H^\top + \Gamma_{t+1} BL_t (S_t^{xz})^\top H^\top + \tilde{\Omega}_{t+1} K_t \Sigma_\omega \\
&\quad + \tilde{\Omega}_{t+1} K_t \sum\nolimits_i D_i S_t^{xx} D_i^\top = 0 \ ,
\end{aligned} \tag{36}$$

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

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

and

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

with

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

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.

A.1.5 Consistency of the Cost-to-Go Solution

The cost-to-go obeys the Bellman equation

$$C_t(x,z) = \mathrm{tr}(Q_t xx^\top + L_t^\top R_t L_t zz^\top) + \int dx'dz' C_{t+1}(x',z')p_{x,t+1}(x'|x,z)p_{z,t+1}(z'|x,z) \,, \quad (37)$$

identical to Eq. 19. 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 $\mathbb{E}[z'|x,z] = K_tHx + M_t z$, and 2nd-order moments given by Eqs. 34.

We propose a solution to the Bellman equation of the form

$$C_t(x,z) = \mathrm{tr}(\Lambda_t xx^\top + \Omega_t zz^\top + \Gamma_t xz^\top) + \gamma_t \,, \quad (38)$$

identical to Eq. 20. 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. 18 with the same boundary conditions. In addition we want to show that γ_t follows Eq. 21 with boundary condition $\gamma_T = 0$.

We first note that Eq. 38 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. 38 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. 18 for $t = T$.

Now, assume that Eq. 38 is true for some $t+1$. Let us show that then it is true for t. We insert Eq. 38 for $t+1$ into Eq. 37 and use the expression of the conditional 2nd-order moments in Eqs. 34 to obtain

$$\begin{aligned}
C_t(x,z) &= \mathrm{tr}(Q_t xx^\top + L_t^\top R_t L_t zz^\top) \\
&\quad + \int dx'dz' \left(\mathrm{tr}(\Lambda_{t+1}x'x'^\top + \Omega_{t+1}z'z'^\top + \Gamma_{t+1}x'z'^\top) + \gamma_{t+1}\right) p_{x,t+1}(x'|x,z)p_{z,t+1}(z'|x,z) \\
&= \mathrm{tr}(Q_t xx^\top + L_t^\top R_t L_t zz^\top) \\
&\quad + \mathrm{tr}[\Lambda_{t+1}(Axx^\top A^\top + BL_t zz^\top L_t^\top B^\top + Axz^\top L_t^\top B^\top + BL_t zx^\top A^\top + \hat{\Sigma}_t^{xx})] \\
&\quad + \mathrm{tr}[\Omega_{t+1}(K_tHxx^\top H^\top K_t^\top + M_t zz^\top M_t^\top + K_tHxz^\top M_t^\top + M_t zx^\top H^\top K_t^\top + \hat{\Sigma}_t^{zz})] \\
&\quad + \mathrm{tr}[\Gamma_{t+1}(Axx^\top H^\top K_t^\top + BL_t zz^\top M_t^\top + Axz^\top M_t^\top + BL_t zx^\top H^\top K_t^\top)] \\
&\quad + \gamma_{t+1} \,. \qquad (39)
\end{aligned}$$

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. 38 where the coefficients obey the Lagrange multiplier equations in Eqs. 18 at time t. In addition, γ_t is computed using Eq. 21.

By induction, then we have that Eq. 38 is true for all t and that the coefficients are indeed the Lagrange multipliers defined in Eqs. 18 and Eq. 21.

A.1.6 Breakdown of the Orthogonality Principle With Internal Noise

Theorem 2. *Assume an initial distribution $p_0(x,z)$ such that $S_0^{xz} = S_0^{zz}$. Then the stationarity equations Eqs. 10, 11, and 18 admit a solution satisfying the orthogonality principle $S_t^{xz} = S_t^{zz}$ for all $t = 1, \dots, T$ if and only if internal noise is zero, i.e. $\Sigma_\eta = 0$. Moreover, this solution corresponds to a critical point of the cost in Eq. 3.*

Proof. We first show that (1) assuming OP ($S_t^{xz} = S_t^{zz}$ for $t = 0, \dots, T$) is true, we prove that the satisfaction of the Lagrange equations for the multipliers, Eqs. 18, 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. 33, 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. 18 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 used that Γ_{t+1} is symmetric. Now, summing we have

$$
\begin{aligned}
\Gamma_t + \tilde{\Omega}_t &= 2 L_t^\top R_t L_t + L_t^\top B^\top \tilde{\Lambda}_{t+1} (A + B L_t) + M_t^\top \Gamma_{t+1} (A + B L_t - K_t H - M_t) \\
&\quad + L_t^\top B^\top \Gamma_{t+1}^\top (A + B L_t) + \sum\nolimits_i L_t^\top C_i^\top \tilde{\Lambda}_{t+1} C_i L_t \\
&= L_t^\top \left[2 R_t L_t + B^\top \tilde{\Lambda}_{t+1} (A + B L_t) + B^\top \Gamma_{t+1}^\top (A + B L_t) + \sum\nolimits_i C_i^\top \tilde{\Lambda}_{t+1} C_i L_t \right] \ ,
\end{aligned}
\tag{40}
$$

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.1.4) largely simplifies to

$$
L_t = \bar{E}_t^{-1} \bar{F}_t \ , \tag{41}
$$

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. 40 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. 33. 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.1.4) largely simplifies to

$$
K_t = A \left(S_t^{xx} - S_t^{zz} \right) H^\top \bar{S}_{HH}^{-1} \ , \tag{42}
$$

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:

$$S_{t+1}^{zz} - S_{t+1}^{xz} = \left[-(A - K_t H)(S_t^{xx} - S_t^{zz})H^\top + K_t \Sigma_\omega + K_t \sum\nolimits_i D_i S_t^{xx} D_i^\top\right] K_t^\top + \Sigma_\eta = 0\,, \tag{43}$$

for all t (this expression has been obtained using the 2nd-order moments in Eqs. 33 after several cancellations). In this expression, the bracket equals zero after using Eq. 42. 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.1.7 Connection with Pontryagin's Maximum Principle

The optimal solution to the deterministic Linear Quadratic Regulator (LQR) [56] problem can be derived, instead of using dynamic programming and the Bellman equation, through Pontryagin's Maximum Principle (PMP). In the deterministic LQR setting, PMP provides necessary and sufficient conditions for optimality. In particular, the Riccati recursion characterizing the optimal LQR solution can be derived directly from PMP [56]. Moreover, PMP itself can be obtained from a Lagrangian formulation in which the state dynamics are enforced as constraints via Lagrange multipliers [56]. Thus, the optimal LQR solution may be viewed as arising from constrained optimization of a deterministic dynamical system, with the multipliers playing the role of costate variables.

In the classical LQAG problem with additive noise, it is well known that the separation principle holds [12]: the optimal control depends on the data only through the conditional mean $z_t = \mathbb{E}[x_t \mid \mathcal{I}_t]$, where $\mathcal{I}_t$ denotes the available information. The conditional mean evolves according to a Kalman filter recursion, and the optimal policy is the LQR feedback law applied to z_t [56]. Therefore, the problem reduces to a fully observed LQR problem in the estimate z_t.

In the full LQMI problem, which incorporates multiplicative and internal noise, the separation principle no longer holds and control and estimation become intrinsically coupled [18, 33]. Our algorithm derived in Sec. 3 can be interpreted as an extension of the principle underlying PMP to this more general stochastic setting. The key step is to reformulate the stochastic optimal control problem as a constrained optimization problem in the space of sufficient statistics. Specifically, the second-order moments of the state and state estimate evolve according to a closed deterministic recursion, and the expected cost depends only on these moments. We therefore enforce the moment dynamics through Lagrange multipliers and optimize jointly over the control and filter gains.

The multipliers coincide with the gradient of the cost-to-go with respect to the sufficient statistics, thereby playing the role of costate variables in this lifted formulation. The stationarity conditions with respect to (L_t, K_t) follow from the Lagrangian optimality conditions and combine the immediate effect of the gains on the cost with their indirect effect through the future evolution of the sufficient statistics. Thus, the same constrained-dynamical reasoning used in deterministic optimal control is applied here at the level of second-order moments rather than physical state trajectories.

Connection with LQAG and the Role of the Orthogonality Principle When considering the classical LQAG problem, Riccati equations can be recovered using our algorithm by noting that the orthogonality principle holds (see Appendix A.1.8). Orthogonality implies that the estimation error is uncorrelated with the estimate, which restricts the admissible space of second-order moments. As a consequence, optimization over the full set of second-order moments becomes unnecessary, and the problem reduces to the classical separation between a Kalman filter and an LQR controller. In contrast, in the general LQMI setting, and more specifically when internal noise is present, the orthogonality principle is no longer satisfied by the optimal solution. This has been observed numerically in [33] and is proven analytically in this work. As a result, approaches that implicitly impose orthogonality – such as the dynamic programming formulation in [18], as discussed in Sec. A.1.1 – restrict the optimization to a submanifold of second-order statistics and generally yield suboptimal solutions.

Our formulation makes this constraint explicit and shows that it is only valid in the absence of internal noise.

A.1.8 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.1.2) when both multiplicative and internal noise terms vanish. To illustrate this, we examine the solutions presented in Appendix A.1.6. As empirically validated in [33], the optimal solutions, when internal noise is absent, satisfy the orthogonality principle (OP). Thus, by setting the multiplicative and internal 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.1.6 is given by Eq. 41. When both multiplicative and internal noise terms are turned off, we obtain

$$L_t = -[2R_t + B^\top(\tilde{\Lambda}_{t+1} + \Gamma_{t+1})B]^{-1}[B^\top(\tilde{\Lambda}_{t+1} + \Gamma_{t+1})A] \,, \tag{44}$$

which corresponds to the optimal L_t for the classic LQAG case (see solutions in Sec. A.1.2) if $S_t = (\Gamma_t + \tilde{\Lambda}_t)$. Using Eq.18 and imposing the OP (setting $\Gamma_t = -\tilde{\Omega}_t$ – see Appendix A.1.6) we obtain

$$\Gamma_t + \tilde{\Lambda}_t = 2Q_t + (A + BL_t)^\top(\tilde{\Lambda}_{t+1} + \Gamma_{t+1})A \,. \tag{45}$$

Now we observe, as discussed in Appendix A.1.6, that Γ_t is symmetric and the same holds for $\tilde{\Lambda}_t$ (by definition), therefore we can rewrite Eq. 45 as

$$\Gamma_t + \tilde{\Lambda}_t = 2Q_t + A^\top(\tilde{\Lambda}_{t+1} + \Gamma_{t+1})(A + BL_t) \,. \tag{46}$$

which corresponds to the formula for S_t in Appendix A.1.2, therefore proving the equality between the two optimal solutions.

The optimal Kalman filter derived under the OP in Appendix A.1.6 is given by Eq. 42, corresponding to

$$K_t = A\,(S_t^{xx} - S_t^{zz})\,H^\top[H(S_t^{xx} - S_t^{zz})H^\top + \Sigma_\omega]^{-1} \,, \tag{47}$$

when neither internal nor multiplicative noise is considered. We note that this solution corresponds to the one presented in Appendix A.1.2 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.1.6 correspond to the globally optimal solutions of the classic LQAG problem in the absence of multiplicative and internal noise.

A.2 Pseudocode and Implementation Details

A.2.1 Pseudocode

The pseudocode below implements the algorithm of Sec. 3, referred to as LCO (Lagrangian Constrained Optimization).

Algorithm 1 Lagrangian Constrained Optimization (LCO)

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: Optimized 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. 18 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. 33 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. 18 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. 33 using $L_t^{(k)}$ and $K_t^{(k)}$
  end for
end for
$L^*_{0,\dots,T} \leftarrow L^{(k)}_{0,\dots,T}$; $K^*_{0,\dots,T} \leftarrow K^{(k)}_{0,\dots,T}$

In Algorithm 1, each outer iteration consists of two passes. First, keeping L and K fixed, we compute the multipliers $\Lambda_t, \Omega_t, \Gamma_t$ backward from the terminal conditions, $t = T$ to $t = 0$, using Eqs. 18. We then reset the moments to their initial values and proceed forward for $t = 0, \dots, T-1$. At each time step, we compute L_t from the current moments $S_t^{xx}, S_t^{xz}, S_t^{zz}$ using Eq. 10, overwrite L_t, and use the updated gain to propagate the moments by one step, from t to $t+1$, using Eqs. 33. After this forward sweep, we recompute the multipliers backward through Eqs. 18 using the updated L, reset the moments, and perform the analogous filter sweep. At each time step, we compute K_t using Eq. 11, overwrite K_t, and immediately propagate the moments using Eqs. 33. This completes one optimization iteration.

A.2.2 Implementation Details

Initialization In our experiments, we initialize $(L^{(0)}, K^{(0)})$ with the solutions of [18], which implicitly enforce OP, and then apply the LCO updates. This provides a convenient and robust starting point. However, the additional sensitivity-to-initialization analysis in Appendix A.3.1 shows that LCO reaches closely comparable low-cost solutions also from zero and random initial gains, including initializations far from the TOD solution. Thus, while TOD provides a useful initialization strategy, it is not strictly necessary to obtain a low-cost solution in the tested benchmark.

Algorithms' Hyper-Parameters Here we report the algorithms' hyper-parameters, as selected for the experiments discussed in Sec. 4.

For the random tasks used to evaluate Algorithm 1 – and to compare it with the gradient-based numerical method from [33] (referred to as GD) – we use the parameters listed in Table 1. Note that, in line with [33], 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 random tasks

Algorithm	Description	Value
GD [33]	Number of iterations of the "GradientDescent()" function	10000
TOD [18]	Number of iterations of the estimation-control optimization	500
LCO (Algorithm 1)	Number of iterations of the estimation-control optimization	500

For the Redundant Arm-Control Tasks, first discussed in Sec. 4, we used the hyperparameters listed in Table 2.

Table 2: Hyper-parameters of the algorithms used in the Redundant Arm-Control Tasks

Algorithm	Description	value
TOD [18]	Number of iterations of the estimation-control optimization	500
LCO (Algorithm 1)	Number of iterations of the estimation-control optimization	500

All results were obtained on a MacBook Pro (Apple M1, 16 GB RAM).

A.3 Experimental Details and Additional Experiments

A.3.1 Random Tasks: Model, Parameters, and Additional Analyses

We provide here the full problem definition and describe the additional perturbation experiment on the LCO solution discussed in Sec. 4.

Problem Definition To assess computational efficiency and cost minimization in a controlled setting, we consider a family of random square LQMI problems in which the state, control, and observation dimensions are all set equal, $m = p = k = n_{\text{task}}$. The trial duration is fixed to $T = 10$. For each task dimension, we generate random problem instances by drawing matrices $A, B \in \mathbb{R}^{n_{\text{task}} \times n_{\text{task}}}$ from zero-mean, unit-variance Gaussian ensembles and rescaling them so that their spectral radii are bounded by $\rho_{\max} = 0.9$, ensuring stable linear dynamics. The observation matrix is set to

$$H = I_{n_{\text{task}}}. \tag{48}$$

The multiplicative control-noise matrix is chosen as

$$C = \sigma_\epsilon B, \tag{49}$$

while the multiplicative observation-noise matrix is defined as

$$D = \sigma_\rho D_{\text{tmp}}, \tag{50}$$

where D_{tmp} is obtained by taking the element-wise absolute value of a stable random matrix generated in the same way as A and B.

The additive noise covariances are isotropic,

$$\Sigma_\xi = \sigma_\xi^2 I_{n_{\text{task}}}, \qquad \Sigma_\omega = \sigma_\omega^2 I_{n_{\text{task}}}, \qquad \Sigma_\eta = \sigma_\eta^2 I_{n_{\text{task}}}, \tag{51}$$

and, unless otherwise stated, we fix

$$\sigma_\xi = \sigma_\omega = \sigma_\rho = \sigma_\epsilon = 0.2. \tag{52}$$

For the dimensionality-scaling benchmark of Figs. 1b–c, we set

$$\sigma_\eta = 0.5, \tag{53}$$

whereas in the fixed-instance internal-noise sweep of Figs. 1d–g we vary σ_η while keeping all other problem parameters fixed.

The initial means of the external and internal states are both set to the all-ones vector,

$$\mathbb{E}[x_0] = \mathbb{E}[z_0] = \mathbf{1}_{n_{\text{task}}}, \tag{54}$$

and the corresponding initial covariances are set to zero,

$$\Sigma_{x_0} = 0, \qquad \Sigma_{z_0} = 0. \tag{55}$$

The quadratic cost matrices are chosen as

$$Q_t = I_{n_{\text{task}}}, \qquad R_t = I_{n_{\text{task}}}, \tag{56}$$

for all t, except for the boundary conditions

$$Q_0 = 0, \qquad R_T = 0, \tag{57}$$

so that the initial state is not penalized and no terminal control cost is applied.

For the dimensionality-scaling experiment of Figs.1b–c, we consider

$$n_{\mathrm{task}} \in \{1, 2, 5, 10, 15, 20, 40, 60, 80, 100\}, \tag{58}$$

and generate 10 random instances of the matrices A, B, D for each dimension. TOD and LCO are run on all dimensions, whereas GD is run only on the subset

$$n_{\mathrm{task}} \in \{1, 2, 5, 10\}, \tag{59}$$

for which the numerical optimization remains computationally feasible. In this benchmark, both LCO and GD are initialized from the corresponding TOD solution, ensuring a fair comparison.

For the fixed-instance internal-noise sweep of Fig.1d–g, we fix a single random task with $n_{\mathrm{task}} = 5$, and vary σ_η while keeping the matrices A, B, H, C, D and the additive external-noise covariances fixed. For each value of σ_η, we initialize LCO from the corresponding TOD solution and GD from the corresponding LCO solution. This setup allows us to isolate the effect of internal noise on the solutions, the objective trajectory, and the discrepancy between TOD, LCO, and GD.

Perturbation Analysis To empirically probe the local stability of the LCO solutions (Algorithm 1), we perform a perturbation analysis to assess whether they behave, in practice, as locally stable minima of the expected cost rather than as obvious saddle points. We consider a fixed random control problem, as described above, with $n_{\mathrm{task}} = 5$, compute the corresponding LCO solution, and evaluate how the cost in Eq. 3 changes under small perturbations of the control and filter gains around the reference point $(L^{\mathrm{LCO}}, K^{\mathrm{LCO}})$. Denoting by $C(L, K)$ the expected cost associated with a given sequence of control and filter gains, we define the baseline value

$$C_{\mathrm{LCO}} = C(L^{\mathrm{LCO}}, K^{\mathrm{LCO}}). \tag{60}$$

We probe the neighborhood of the LCO solution along three perturbation families:

1. *L-only perturbations:* only the control gains are perturbed, while the filter gains are kept fixed at K^{LCO};
2. *K-only perturbations:* only the filter gains are perturbed, while the control gains are kept fixed at L^{LCO};
3. *joint perturbations:* control and filter gains are perturbed simultaneously in the full parameter space.

For the L-only and K-only scans, perturbation directions are sampled as Gaussian random tensors with the same shape as the corresponding gains and then normalized to unit Frobenius norm. For the joint scan, we concatenate all entries of L and K into a single vector, sample Gaussian directions in this full parameter space, and normalize them to unit Euclidean norm. For each perturbation family, we sample $N_{\mathrm{dir}} = 10^5$ independent normalized directions. Around each sampled direction, we evaluate the cost at both positive and negative relative displacements from the LCO solution. The perturbation amplitudes are chosen as

$$\epsilon \in \{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}. \tag{61}$$

Here, ϵ controls the relative perturbation magnitude with respect to the norm of the reference solution in the corresponding parameter space. Thus, for an L-only perturbation direction ΔL with $|\Delta L|_F = 1$, we evaluate

$$L^{\pm} = L^{\mathrm{LCO}} \pm \epsilon |L^{\mathrm{LCO}}|_F \Delta L, \qquad K^{\pm} = K^{\mathrm{LCO}}, \tag{62}$$

so that

$$\frac{|L^{\pm} - L^{\mathrm{LCO}}|_F}{|L^{\mathrm{LCO}}|_F} = \epsilon. \tag{63}$$

Similarly, for a K-only perturbation direction ΔK with $|\Delta K|_F = 1$, we evaluate

$$L^{\pm} = L^{\mathrm{LCO}}, \qquad K^{\pm} = K^{\mathrm{LCO}} \pm \epsilon |K^{\mathrm{LCO}}|_F \Delta K. \tag{64}$$

For the joint perturbations, letting

$$x^{\mathrm{LCO}} = \mathrm{vec}(L^{\mathrm{LCO}}, K^{\mathrm{LCO}}) \tag{65}$$

and Δx be a unit-norm direction, we evaluate

$$x^{\pm} = x^{\mathrm{LCO}} \pm \epsilon |x^{\mathrm{LCO}}| \Delta x, \tag{66}$$

and then reshape $x^{\pm}$ back into $(L^{\pm}, K^{\pm})$.

For every perturbed point, we compute the cost increment using Eq. 3

$$\Delta C = C(L^{\pm}, K^{\pm}) - C_{\mathrm{LCO}}. \tag{67}$$

This procedure yields, for each perturbation family and each value of ϵ, a distribution of cost changes over random directions in the local neighborhood of the LCO solution.

In Fig. 3, we show the distributions of ΔC for the three perturbation families and for increasing perturbation amplitudes ϵ. In all cases, the distributions remain concentrated at positive values and shift upward as ϵ increases: all sampled random perturbations around the LCO solution worsen the expected cost. This is consistent with the LCO solution lying in a locally stable region of the objective landscape rather than at an obvious saddle point. Repeating the same analysis with absolute perturbation amplitudes, i.e., without norm-based scaling of the perturbations, confirms the same qualitative behavior.

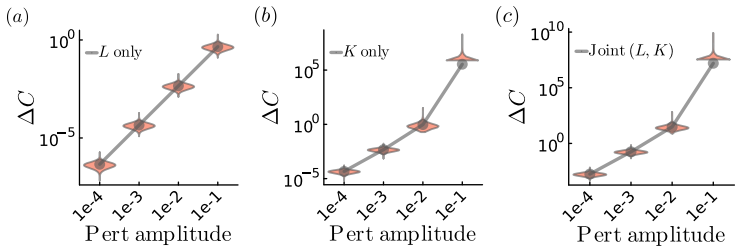

Figure 3: *Local Perturbation Scan around the LCO Solution.* For a fixed random task instance with $n_{\mathrm{task}} = 5$, we take the LCO solution as reference and evaluate the cost increment $\Delta C = C_{\mathrm{pert}} - C_{\mathrm{LCO}}$ under random perturbations of increasing relative amplitude (*Pert Amplitude*). **(a)** ΔC distribution over normalized random directions in the control-gain subspace L. **(b)** Same as (a), for the filter-gain subspace K. **(c)** Same as (a)–(b), for joint perturbations in the full (L, K) parameter space. Violins show the distribution across sampled directions; gray markers and lines denote the median.

Random Benchmarks with Limited Sensory Observations To further evaluate the performance of LCO under limited sensory observations, we repeated the random-task benchmark with a single observed state component. We considered one random task realization for each $m \in \{6, 8, 10\}$, with $p = m$, $k = 1$, and $T = 10$, setting $H = (1, 0, \ldots, 0)$ so that only the first state component was directly observed. All other aspects of the task generation and optimization procedures were retained as described above. We fixed $\sigma_\xi = \sigma_\omega = \sigma_\epsilon = \sigma_\rho = 0.2$ and $\sigma_\eta = 0.5$.

LCO converged reliably in all three dimensions, achieving expected-cost reductions relative to TOD of 35%, 66%, and 60% for $m = 6, 8, 10$, respectively. The resulting costs were also comparable to those obtained with GD, but at substantially lower computational cost. Specifically, LCO required approximately 7.8, 0.71, and 0.89 seconds, respectively, compared with approximately 2810, 6170, and 13000 seconds for GD. The GD runs reached their maximum of 10000 iterations without satisfying the convergence criterion (these recorded runtimes are machine-dependent and may include first-call Julia compilation overhead). These results demonstrate that LCO retains its performance advantage over TOD and achieves costs comparable to direct gradient-based optimization, while remaining computationally efficient even when direct sensory observations are severely limited. These results are further supported by similar analyses performed in the stabilization task, as discussed in Appendix A.3.3.

Sensitivity to Initialization To investigate whether the performance of LCO depends on its initialization, we examined the solutions obtained starting from different initial control and filter gains. We considered one random task realization with $m = p = k = 5$, $T = 10$, and $H = I_5$, using the same task-generation procedure, cost function, and optimization settings described above. The noise amplitudes were fixed at $\sigma_\xi = \sigma_\omega = \sigma_\epsilon = \sigma_\rho = 0.2$ and $\sigma_\eta = 0.5$.

We compared LCO solutions obtained from different initializations: the TOD solution, zero control and filter gains, and 200 random initializations in the joint (L, K) parameter space. To generate the random initializations, we sampled 50 normalized Gaussian directions, considered both signs, and scaled each direction either to the norm of the TOD gains or to ten times that norm. The first condition tested initial gains of comparable magnitude but unrelated orientation and temporal structure, whereas the second tested substantially larger initial gains, deliberately far from the TOD solution. This allowed us to assess whether LCO could recover comparable low-cost solutions without relying on initialization near TOD. After 500 complete LCO passes per initialization, all 202 runs achieved expected costs below TOD, reducing the expected cost by 30.6% when initialized from TOD and by 30.3% when initialized from zero gains. Similarly, across the 200 random initializations, the cost reductions ranged from 30.4% to 30.6%. Moreover, more than 90% of the random runs achieved a final cost within 0.1% of the TOD-initialized LCO solution, and 99% within 0.2%, with a maximum difference of approximately 0.22%. These results were obtained even when the initial costs were several orders of magnitude higher than the TOD cost.

Thus, TOD provides a convenient initialization, but it is not strictly necessary to obtain a low-cost solution in this benchmark: LCO reaches closely comparable solutions from zero, random, and strongly mis-scaled initial gains.

These results complement the local perturbation analysis reported above, where we examined the expected cost under perturbations around the LCO solution in the control-gain, filter-gain, and joint (L, K) parameter spaces. No tested neighboring perturbation produced a lower expected cost, and GD initialized from LCO did not further reduce the objective appreciably. Together, these findings provide empirical evidence that the solutions obtained with LCO behave as stable local minima rather than obvious saddle points, while also demonstrating robustness to substantially different initializations.

A.3.2 Reaching Task: Model, Parameters, and Additional Analyses

In Sec. 4, we apply the LCO algorithm to a reaching task. Here, we provide the full model definition, parameter choices, and additional analyses.

Problem Definition We consider a three-degree-of-freedom (3-DOF) planar arm, whose joints can be interpreted as the shoulder, elbow, and wrist. We use a linear time-invariant (LTI) approximation of the arm dynamics, with a six-dimensional state (three joint angles and their angular velocities), a nine-dimensional control (muscle-like control inputs), providing a redundant actuation architecture inspired by musculoskeletal models [57], and a three-dimensional observation (only joint angles are observed), i.e., $m = 6$, $p = 9$, and $k = 3$. We denote by $\theta_t \in \mathbb{R}^3$ the joint-angle vector and by $\dot{\theta}_t \in \mathbb{R}^3$ the corresponding angular velocities. The deterministic part of the discrete-time dynamics is given by

$$\begin{aligned} \theta_{t+1} &= \theta_t + \Delta t\, \dot{\theta}_t, \\ \dot{\theta}_{t+1} &= (I_3 - \Delta t\, M_{\text{joint}}^{-1} D_{\text{joint}})\dot{\theta}_t + \Delta t\, M_{\text{joint}}^{-1} S u_t, \end{aligned} \tag{68}$$

where I_3 is the 3×3 identity matrix, $\Delta t = 0.01\,\text{s}$ is the time step, and $u_t \in \mathbb{R}^9$ is the control vector. The joint-inertia and damping matrices are

$$M_{\text{joint}} = \text{diag}(1.2, 0.8, 0.5), \qquad D_{\text{joint}} = 2I_3. \tag{69}$$

The muscle-to-joint actuation matrix $S \in \mathbb{R}^{3\times 9}$, which linearly maps muscle-like activations onto joint torques, is

$$S = \begin{pmatrix} 1.2 & -1.0 & 0 & 0.8 & -0.6 & 0 & 0.5 & 0 & 0 \\ 0 & 0 & 1.0 & -0.4 & 0.6 & -0.5 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1.0 & 0 & -0.3 & 0.6 \end{pmatrix}. \tag{70}$$

The structure of S provides a simplified representation of musculoskeletal redundancy, loosely mimicking mono-articular and bi-articular muscle organization. Its sparsity and sign patterns encode which joints each actuator influences and the corresponding torque directions, allowing multiple activation patterns to generate the same joint torques without explicitly modeling individual muscles. Together, the decreasing joint inertias and the structure of S, which includes both single-joint actuation and coupling between adjacent joints, motivate interpreting the three joints as the shoulder, elbow, and wrist, respectively.

Defining the state vector $x_t = (\theta_t^\top, \dot{\theta}_t^\top)^\top \in \mathbb{R}^6$, the system matrices are

$$A = \begin{pmatrix} I_3 & \Delta t\, I_3 \\ 0_{3\times 3} & I_3 - \Delta t\, M_{\text{joint}}^{-1} D_{\text{joint}} \end{pmatrix}, \qquad B = \begin{pmatrix} 0_{3\times 9} \\ \Delta t\, M_{\text{joint}}^{-1} S \end{pmatrix}. \tag{71}$$

The multiplicative control-noise matrix is $C = \sigma_\epsilon B$. Since only joint angles are directly observed, the observation and multiplicative sensory-noise matrices are

$$H = (I_3 \quad 0_{3\times 3}), \qquad D = \sigma_\rho H. \tag{72}$$

The task consists of reaching a target posture in which all three joint angles and angular velocities are zero. The elbow (second component of the vector θ_t) is initially displaced by 20°, while the shoulder and wrist start at their target positions. Taking the target posture as the origin, the initial external and internal states are deterministic:

$$x_0 = z_0 = (0, -\pi/9, 0, 0, 0, 0)^\top, \qquad \Sigma_{x_0} = \Sigma_{z_0} = 0_{6\times 6}. \tag{73}$$

The movement duration is $0.40\,\text{s}$, corresponding to $T = 40$ control steps.

The quadratic cost uses illustrative task penalties chosen to generate a target-directed elbow reach while discouraging shoulder and wrist motion. The running and terminal state-cost matrices are

$$\begin{aligned} Q_t &= \text{diag}(1000, 10^{-5}, 1000, 100, 10^{-5}, 100), \qquad t = 0, \ldots, T-1, \\ Q_T &= \text{diag}(1000, 5000, 1000, 100, 100, 100). \end{aligned} \tag{74}$$

The control cost is

$$R_t = 0.005 I_9, \qquad t = 0, \ldots, T-1, \qquad R_T = 0. \tag{75}$$

These weights impose weak penalties on elbow angle and angular velocity during the movement, while strongly penalizing terminal elbow-angle error and suppressing shoulder and wrist motion throughout. They represent illustrative task penalties rather than costs fitted to human behavior.

The additive process, sensory, and internal noises have isotropic covariances,

$$\Sigma_\xi = \sigma_\xi^2 I_6, \qquad \Sigma_\omega = \sigma_\omega^2 I_3, \qquad \Sigma_\eta = \sigma_\eta^2 I_6. \tag{76}$$

We set the external-noise amplitudes to

$$\sigma_\xi = 0.01, \qquad \sigma_\omega = 0.9^\circ, \qquad \sigma_\epsilon = 0.10, \qquad \sigma_\rho = 0.01, \tag{77}$$

and vary internal noise over $\sigma_\eta \in [0, 0.04]$, using 20 noise levels while keeping all other parameters fixed. The empirical motivation and calibration of these noise amplitudes are discussed below.

Psychophysically Informed Noise Calibration To obtain a biologically plausible noise regime, we combined literature-based estimates, where closely related empirical measurements were available, with a behavioral calibration of the remaining model-specific parameters. The task itself was designed to reproduce basic features of human reaching movements: the 20° elbow displacement is comparable to single-joint movements studied experimentally [60], while the $0.40\,\text{s}$ duration falls within the range reported for short human reaches [59].

The additive sensory-noise amplitude was set to $\sigma_\omega = 0.9^\circ$, within the 0.6°–1.1° range inferred for the precision of shoulder- and elbow-angle information [64]. Since sensory observations in our model consist directly of joint angles, this provides a closely related empirical reference for the additive observation-noise level. The multiplicative motor-noise coefficient was set to $\sigma_\epsilon = 0.10$, representing signal-dependent variability in motor execution. This value is of the same order as the coefficients used by [29] in an optimal-feedback-control model of reaching, which incorporated signal-dependent motor noise and was calibrated to reproduce features of human movement variability. Although their motor-noise parameterization differs from ours, it provides a useful reference for the magnitude of this contribution. For multiplicative sensory noise, we retained a small coefficient, $\sigma_\rho = 0.01$. To the best of our knowledge, no directly transferable psychophysical estimate is available for this model-specific parameter. Nevertheless, multiplicative sensory noise has been calibrated against psychophysical measurements in related optimal-control models. For example, [65] calibrated position- and velocity-dependent visual noise to Weber fractions of 0.05 and 0.08, respectively. Since their sensory modality and noise parameterization differ from ours, we use these estimates as qualitative support for a small multiplicative sensory contribution rather than as a direct numerical calibration.

After fixing the sensory and motor noise parameters, we calibrated the additive process-noise amplitude to $\sigma_\xi = 0.01$, yielding a terminal elbow-angle standard deviation of 8.55% of the 20° movement amplitude at zero internal noise. This is broadly comparable in relative magnitude to the endpoint variability reported for short human reaching movements by [59]. Importantly, ξ_t represents an effective process-disturbance term, which may encompass unpredictable perturbations acting on the plant or environment rather than intrinsic physiological noise alone. Experimental studies involving unstable force fields and externally imposed mechanical perturbations demonstrate the relevance of such disturbances in human motor control [66, 67, 68].

Finally, we selected the internal-noise range $\sigma_\eta \in [0, 0.04]$ according to two criteria: the resulting movement variability should remain within a plausible behavioral regime, and internal noise should influence the optimal control strategy without becoming the dominant direct source of movement variability. Across the full sweep, the terminal elbow-angle SD under LCO increased only from 8.55% to 10.5% of the movement amplitude (Fig. 2b).

To assess the relative contributions of the different noise sources, we performed a complementary leave-one-noise-source-out analysis, removing each source individually while keeping the optimized control and filter gains fixed. These reductions quantify changes in SD and are not additive fractions of the total variance. Across nonzero internal-noise levels, removing process noise reduced terminal elbow-angle SD by 65%–78%, confirming its dominant contribution. Removing internal noise produced reductions of 1%–5%, exceeding those associated with removing additive sensory noise (0.4%–0.8%) or multiplicative motor noise (0.6%–0.7%). Thus, the internal-noise range allows us to investigate substantial changes in optimal control strategies without relying on a regime in which internal noise dominates the overall behavioral variability.

Internal Noise versus Imperfect Inference As suggested by Beck et al. [54], behavioral variability can reflect suboptimal inference arising from incorrect internal models. Here, we illustrate that sensory-model mismatch and stochastic internal fluctuations can lead to qualitatively different behavioral predictions and control strategies. We consider a specific form of imperfect inference, namely the misestimation of sensory reliability, and compare its effects with those induced by internal noise in our calibrated reaching task. We optimized the control and filter gains assuming either higher or lower sensory noise than was actually present, and then evaluated the resulting gains in the true system, keeping internal noise at zero. Both forms of sensory-noise misestimation increased expected cost but produced different effects from internal noise. Sensory-noise overestimation led to an overall decrease in peak elbow velocity, whereas underestimation produced a modest increase. Neither reproduced the pronounced non-monotonic modulation induced by internal noise. The gain profiles also differed: overestimation decreased the time-averaged filter-gain log-pseudodeterminant $\log(|K|_p)$, whereas underestimation increased it, with only minor changes in the corresponding control-gain measure $\log(|L|_p)$ in both cases. In contrast, increasing internal noise led to an overall increase in the filter-gain measure and a substantial decrease in the control-gain measure. For sensory-noise overestimation, the differences in peak-velocity predictions persisted even when endpoint variability was comparable to that produced by internal noise. Thus, within the tested parameter ranges, sensory-model mismatch does not reproduce the combination of behavioral and computational effects induced by internal noise, illustrating how these mechanisms can lead to qualitatively different predictions.

A.3.3 Optimal Control in a Redundant-Arm Stabilization Task

As a complementary numerical test of the LCO algorithm, we consider a stabilization task using the same three-joint planar arm described in Appendix A.3.2 (Fig. 4a). Unlike the reaching task, here the arm starts at a reference posture and must maintain it in the presence of stochastic disturbances. This setting is motivated by experimental paradigms involving postural stabilization and compensation for mechanical perturbations [66, 67]. Although not calibrated to a specific human experiment, the task provides a structured benchmark to further investigate the performance of LCO and the resulting optimal control strategies under different task demands and over a longer time horizon. With $m = 6$, $p = 9$, $k = 3$, and $T = 300$, GD would require optimizing more than 2×10^4 parameters, making a systematic gradient-based analysis computationally expensive.

Problem Definition We retain the same linear arm dynamics, joint-inertia and damping matrices, and muscle-to-joint actuation matrix S defined in Appendix A.3.2. The system matrices A, B, and

H, as well as the multiplicative noise structures $C = \sigma_\epsilon B$ and $D = \sigma_\rho H$, are therefore defined as before. The stabilization task differs from the reaching task in its initial conditions, cost function, noise amplitudes, and time horizon.

The reference posture is taken as the origin, with both the external and internal states initialized at zero:

$$x_0 = z_0 = 0_{6\times 1}, \qquad \Sigma_{x_0} = \Sigma_{z_0} = 0_{6\times 6}. \tag{78}$$

The time step is $\Delta t = 0.01\,\mathrm{s}$, and the horizon comprises $T = 300$ state-time points.

The quadratic state cost penalizes deviations of all three joint angles from the reference posture, together with their angular velocities:

$$Q_t = \mathrm{diag}(1, 1, 1, 10^{-3}, 10^{-3}, 10^{-3}), \qquad t = 0, \ldots, T-1. \tag{79}$$

The control cost is

$$R_t = 0.01 I_9, \qquad t = 0, \ldots, T-2, \qquad R_{T-1} = 0. \tag{80}$$

The additive process, sensory, and internal noises retain the isotropic covariance structure defined for the reaching task. The noise amplitudes and remaining parameters are summarized in Table 3. Unlike the reaching task, these parameters are not psychophysically calibrated, but are chosen to investigate the algorithm and the resulting control strategies under a different noise regime.

Table 3: Parameters of the redundant-arm stabilization task.

Name	Description	Value
Δt	Time step (s)	0.01
T	Number of state-time points	300
σ_ξ	Process-noise amplitude	0.1
σ_ω	Additive sensory-noise amplitude	$0.1\,\mathrm{rad} \approx 5.73^\circ$
σ_ϵ	Multiplicative motor-noise coefficient	0.1
σ_ρ	Multiplicative sensory-noise coefficient	0.1
σ_η	Internal-noise amplitude	$[0, 0.4]$

Main Results In this stabilization task, LCO achieves lower expected costs than TOD for nonzero internal noise, with the gap widening at higher noise levels (Fig. 4b). As internal noise increases, the LCO solutions show an increase in the filter-gain measure and a decrease in the control-gain measure (Figs. 4c,d), consistent with the qualitative trends observed in the reaching task and differing from the gain adjustments obtained with TOD. Importantly, the internal-noise levels considered do not dominate overall behavioral variability, as confirmed by the additional analyses presented below. Beyond these gain adjustments, internal noise induces a marked reorganization of the internal-state dynamics under LCO. In particular, the variability of z_t becomes increasingly concentrated in a low-dimensional subspace, with the first principal component accounting for nearly all the variance at high internal noise (Fig. 4e). This dominant mode lies mainly along task-irrelevant components, namely angular velocities, which are weakly penalized and not directly observed (Fig. 4f). This organization is reminiscent of the minimal-intervention principle in optimal feedback control [10], whereby variability in the physical state is preferentially tolerated along task-irrelevant dimensions. Here, a similar organization emerges in the internal-state representation (see the additional PC analysis in the following paragraph, which further highlights the qualitatively different internal-state dynamics obtained with TOD). Moreover, the readout of this dominant mode decreases with internal noise (Fig. 4g), consistent with the reduced transmission of internal-state variability to motor commands observed in the reaching task. This combination of a dominant internal mode and weak motor readout is qualitatively reminiscent of the oblique regime observed in recurrent neural networks trained on motor tasks [69]. These computational changes are also reflected in behavior, through a modulation of peak angular speed that differs from the predictions of TOD (Fig. 4h). Importantly, these effects arise even though internal noise contributes relatively little to overall physical-state variability compared with process disturbances, as quantified in the additional analyses below.

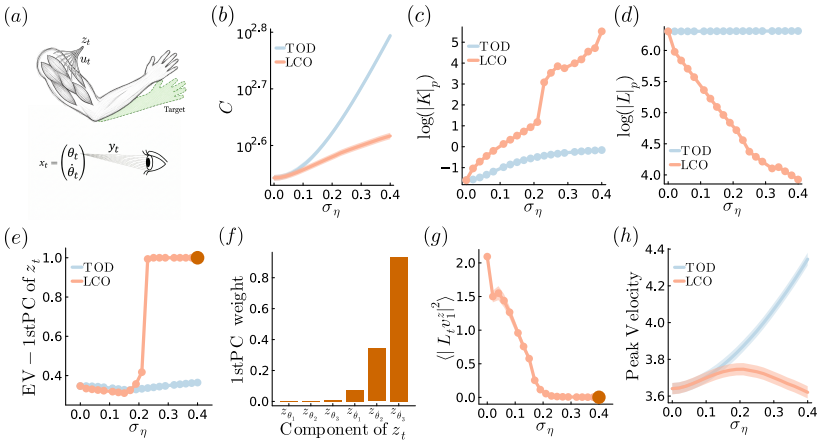

Figure 4: *Optimal Control in a Redundant-Arm Stabilization Task*. **(a)** Schematic illustration of the redundant arm. Upper: redundant three-joint arm and reference posture. Lower: observation model in which only joint angles are directly observed. **(b)** Mean cost C vs. internal noise σ_η for TOD (blue) and LCO (red), estimated over 500 Monte Carlo trials per condition; shading denotes SEM. **(c)** Time-averaged log-pseudodeterminant of the filter gains vs. σ_η. The log-pseudodeterminant is computed as the sum of the logarithms of all singular values of K_t above a numerical tolerance (10^{-12}). **(d)** Same as (c), for the control gains. **(e)** Fraction of variance explained (EV) by the first principal component (PC1) of z_t vs. σ_η. **(f)** Absolute components of PC1 of z_t at $\sigma_\eta = 0.4$ for LCO. **(g)** Time-averaged readout strength of PC1 of z_t for LCO, $\langle |L_t v_1^z|^2 \rangle$, vs. σ_η, where v_1^z denotes the first principal direction of z_t. **(h)** Mean peak angular speed vs. σ_η, computed as the maximum over time of $|\dot{\theta}_t|$ in each trial; shading denotes SEM.

Principal Component Analysis of Internal-State Dynamics We report an additional principal component (PC) analysis of the internal dynamics of z_t at zero and high internal noise σ_η for the solutions obtained with the TOD algorithm [18]. For both noise conditions, we show the fraction of explained variance (EV) associated with each PC and the corresponding PC weights, given by the absolute value of the components of each PC. At zero internal noise, TOD and LCO coincide. In this regime, the optimal solution does not exhibit a clear separation between task-relevant and task-irrelevant components: at least three dimensions are needed to explain most of the variance of z_t (Fig. 5a), and the dominant modes still involve substantial contributions from task-relevant components (Fig. 5b–d).

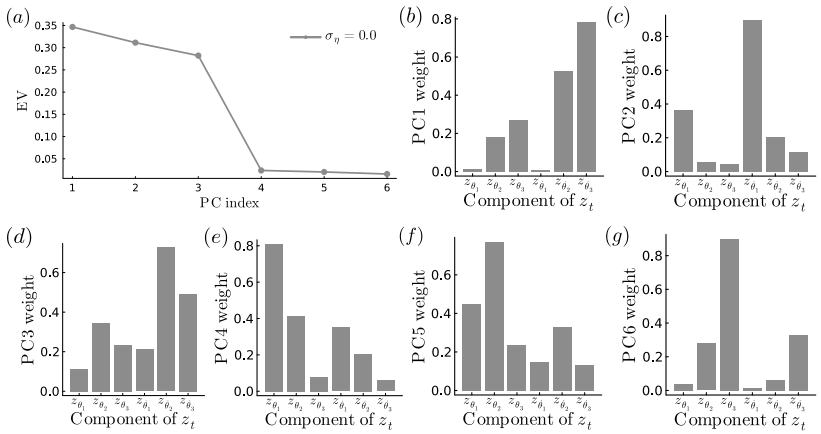

Figure 5: *Principal-Component Structure of the Internal Dynamics at Zero Internal Noise*. **(a)** Fraction of variance explained (EV) by each principal component (PC) of z_t at $\sigma_\eta = 0$. **(b-g)** Absolute components of PCs 1–6 of z_t, respectively. At $\sigma_\eta = 0$, the solutions obtained with TOD and LCO coincide, so only one set of PCs is shown.

At high internal noise, in contrast, LCO exhibits a low-dimensional organization of the internal dynamics that channels variability into task-irrelevant components (Fig. 4e–f). TOD, which is suboptimal in this regime (Fig. 4b), does not achieve the same dimensional compression of z_t (Fig. 6). Specifically, at high internal noise variance, the explained variance (EV) of the first PC in LCO reaches nearly 100% (Fig. 4e), whereas the first PC in TOD accounts for only roughly one third of the total variance (Fig. 6a). Consequently, internal fluctuations are filtered less efficiently, influence behavior more strongly, and impair performance. These results further highlight how LCO uncovers computational strategies for handling internal noise that previous approaches could not reveal.

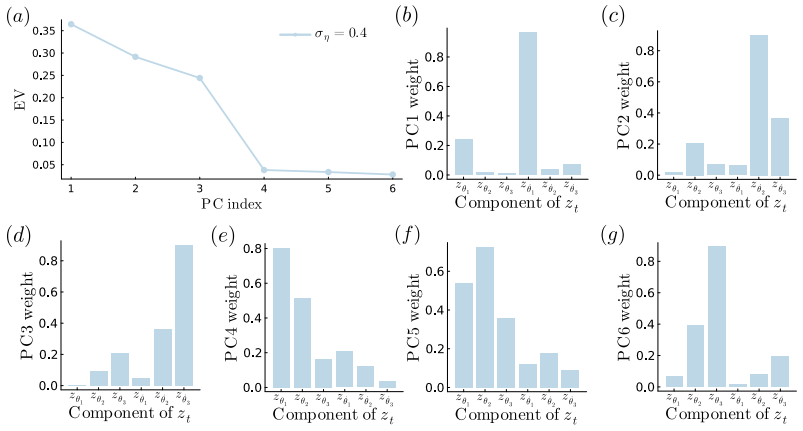

Figure 6: *Principal-Component Structure of the Internal Dynamics at High Internal Noise*. **(a)** Fraction of variance explained (EV) by each principal component (PC) of z_t at $\sigma_\eta = 0.4$. **(b-g)** Absolute components of PCs 1–6 of z_t, respectively, for the solution computed with TOD.

Behavioral Variability and Noise-Source Contributions To further characterize the role of internal noise in the stabilization task, we first quantified its effect on physical-state variability. Unlike the reaching task, the stabilization task was not calibrated to a specific human experiment, but was designed as a structured sensorimotor-control benchmark. Since the arm starts at the reference posture, with $x_0 = z_0 = 0$, the ratio between noise amplitude and mean state is not informative. Instead, we assessed the variability induced in the physical state across the full internal-noise sweep, $\sigma_\eta \in [0, 0.4]$, using 5000 Monte Carlo trials per condition. Between $\sigma_\eta = 0$ and $\sigma_\eta = 0.4$, the standard deviations of the shoulder, elbow, and wrist angles increased by 15.50%, 12.07%, and 10.95%, respectively, while changes in angular-velocity variability were smaller. Thus, increasing internal noise produces measurable but moderate changes in behavior, rather than an abrupt or order-of-magnitude increase in physical-state variability. To quantify the relative contributions of the different noise sources, we performed a leave-one-noise-source-out analysis. At each internal-noise level, we removed each source individually while keeping the LCO control and filter gains optimized for the full model fixed, thereby measuring its contribution within the same closed-loop system. These reductions quantify changes in standard deviation with fixed gains and do not represent additive fractions of the total variance. Removing process noise ξ_t reduced physical-state standard deviations by approximately 81%–98%, depending on the component and internal-noise level, indicating that process disturbances dominate the overall behavioral variability. As discussed above, this noise source can represent unpredictable external mechanical perturbations, as commonly introduced in experimental studies of motor control (e.g., [70, 67, 71, 72]). Removing internal noise reduced joint-angle and angular-velocity standard deviations by at most 1.83% and 1.70%, respectively. By comparison, removing sensory noise produced maximum reductions of only 0.033% and 0.139%, respectively, while multiplicative motor noise had an even smaller effect. Therefore, in this noise regime, internal fluctuations contribute more to physical-state variability than sensory and multiplicative motor noise, but remain secondary to process disturbances. This shows that the substantial effects of internal noise on optimal computational strategies, and the suboptimality of TOD, do not require internal noise to dominate behavioral variability.

The relative contributions of internal and sensory noise can vary across experimental conditions, as suggested by previous studies [24, 44]. We therefore investigated whether the performance advantage of LCO persists when sensory noise contributes more strongly to behavioral variability

than internal noise. To this end, we repeated the stabilization task after increasing the additive sensory-noise amplitude, multiplicative sensory-noise coefficient, and multiplicative motor-noise coefficient tenfold, from 0.1 to 1.0, while keeping process noise fixed at $\sigma_\xi = 0.1$. We considered $\sigma_\eta \in \{0, 0.1, 0.2, 0.3, 0.4\}$ and independently reoptimized TOD and LCO at each internal-noise level. We then repeated the leave-one-noise-source-out analysis using the corresponding optimized gains. Across nonzero internal-noise levels, removing process noise reduced the mean joint-angle standard deviation by approximately 80%, confirming that process disturbances remained the dominant source of physical-state variability. In contrast to the original noise regime, removing additive sensory noise reduced this measure by 1.3%–1.6%, and removing multiplicative sensory noise reduced it by 0.6%, compared with 0.4%–0.5% for internal noise. Thus, sensory noise now contributes more strongly than internal noise to physical-state variability. Nevertheless, LCO continued to outperform TOD, reducing the expected cost by 8.77%, 25.28%, 40.28%, and 52.17% at $\sigma_\eta = 0.1$, 0.2, 0.3, and 0.4, respectively.

Together, these analyses show that the performance advantage of LCO persists across regimes in which internal noise contributes either more or less than sensory noise to behavioral variability. Importantly, these conclusions do not rely on internal noise being the dominant source of physical-state fluctuations. They also complement the reaching-task results, where similar computational strategies and performance advantages emerge under a psychophysically informed noise calibration.

Performance under Limited Observability As a complementary test in a structured sensorimotor setting, we considered a more restrictive observation regime in the stabilization task, where only the elbow angle was directly observed. We retained the original dynamics, costs, time horizon, and noise amplitudes, but reduced the observation dimension from $k = 3$ to $k = 1$, setting $H = (0, 1, 0, 0, 0, 0)$. We considered six internal-noise levels, $\sigma_\eta \in \{0.05, 0.10, 0.15, 0.20, 0.30, 0.40\}$, and compared the solutions obtained with TOD and LCO. LCO converged reliably at every internal-noise level and consistently outperformed TOD, reducing the expected cost by 15.0%, 41.3%, 61.3%, 73.7%, 86.3%, and 91.8%, respectively. Thus, LCO remains numerically stable and substantially outperforms TOD even under limited sensory observations, with its advantage becoming more pronounced as internal noise increases. We did not include GD in this comparison due to the computational cost associated with optimizing the large number of gain parameters.

A.3.4 Robustness to Non-Gaussian Noise

As outlined in Sec. 3, the solutions obtained with the LCO algorithm depend only on the first- and second-order moments of the noise terms. Therefore, no distributional assumptions beyond finite covariance are required, and the method applies to any noise source with well-defined second moments. To verify this empirically, we repeated the Monte Carlo simulations for the redundant arm control task of Appendix A.3.3, using the same parameters except for fixing $\sigma_\eta = 0.1$.

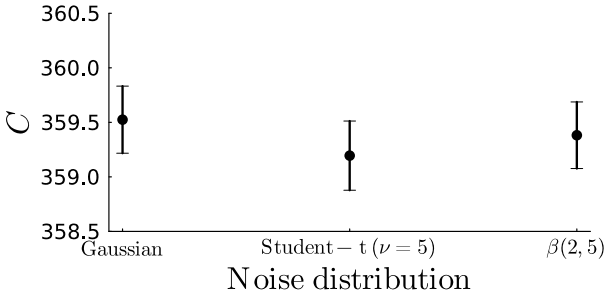

Figure 7: *Robustness to Non-Gaussian Noise*. Mean total cost ($\pm$ standard error of the mean across 100000 Monte-Carlo trials) obtained under three noise distributions with matched variance: Gaussian (baseline), heavy-tailed Student-t ($\nu = 5$), and skewed $\beta(2, 5)$. Despite strong differences in shape, tail behavior, and symmetry, all distributions yield nearly identical expected cost, confirming that—under linear dynamics and quadratic cost—performance depends only on second moments and not on Gaussianity.

We considered three noise distributions for all noise terms, matched in variance but strongly differing in shape: Gaussian, heavy-tailed Student-t noise ($\nu = 5$), and skewed $\beta(2, 5)$ noise rescaled to have zero mean and matched variance. All control, filter, and internal parameters were kept fixed across conditions. Because the dynamics are linear and the cost is quadratic, the expected cost should depend only on second moments and therefore remain invariant across noise distributions. This prediction is confirmed in Fig. 7: the mean total cost is nearly identical for all three distributions, despite their markedly different shapes. This numerical result further supports the theoretical claim that the framework does not require Gaussian noise, and that performance depends solely on the covariance structure of the perturbations.

NeurIPS Paper Checklist

1. **Claims**

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

   Answer: [Yes]

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

   Guidelines:

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

2. **Limitations**

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

   Answer: [Yes]

   Justification: Limitations are discussed in the Conclusions.

   Guidelines:

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

3. **Theory assumptions and proofs**

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

   Answer: [Yes]

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

Guidelines:

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

4. **Experimental result reproducibility**

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

   Answer: [Yes]

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

   Guidelines:

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

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

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

Answer: [Yes]

Justification: The code and instructions required to reproduce the main experimental results are publicly available at `https://github.com/francescodamiani/LQMI-Constrained-Optimization`. Additional methodological details and experimental procedures are provided in the main text and Appendix.

Guidelines:

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

6. **Experimental setting/details**

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

   Answer: [Yes]

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

   Guidelines:

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

7. **Experiment statistical significance**

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

   Answer: [Yes]

   Justification: Where applicable, plots include error bars indicating the standard error of the mean across trials.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).

- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

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

   Answer: [No]

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

   Guidelines:

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

9. **Code of ethics**

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

   Answer: [Yes]

   Justification: This study develops a theoretical framework for addressing stochastic optimal control problems and does not involve applications that raise direct ethical concerns.

   Guidelines:

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

10. **Broader impacts**

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

    Answer: [N/A]

    Justification: Because the work is entirely theoretical, there is no societal impact.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.

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

11. **Safeguards**

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

    Answer: [N/A]

    Justification: The paper poses no such risks.

    Guidelines:

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

12. **Licenses for existing assets**

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

    Answer: [N/A]

    Justification: The paper does not use existing assets.

    Guidelines:

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

- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

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

    Answer: [N/A]

    Justification: The paper does not release new assets.

    Guidelines:

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

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

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

    Answer: [N/A]

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

    Guidelines:

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

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

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

    Answer: [N/A]

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

    Guidelines:

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

16. **Declaration of LLM usage**

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

Answer: [N/A]

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

Guidelines:

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