Improving Optimal Control and Estimation for Realistic Noise Models of the Sensorimotor System

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

Akiyuki Anzai
Department of Brain and Cognitive Sciences
University of Rochester
Rochester, USA
aanzai@ur.rochester.edu

Jan Drugowitsch
Department of Neurobiology
Harvard Medical School
Boston, USA
jan_drugowitsch@hms.harvard.edu

Gregory C. DeAngelis
Department of Brain and Cognitive Sciences
University of Rochester
Rochester, USA
gdeangelis@ur.rochester.edu

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

Abstract

Sustaining perception-action loops is a fundamental brain computation, which can be effectively described by stochastic optimal control theory through optimality principles. When accounting for a realistic noise model of the sensorimotor system, including multiplicative noise in feedback and motor output as well as internal noise in estimation, the mathematical complexity of the problem increases significantly. The standard algorithm in use is the one introduced in the seminal study in [1]. We identify a limitation in the original derivation stemming from the assumption of unbiased estimation and propose an efficient gradient descent optimization that minimizes the cost-to-go, enforcing only the linearity of the control law. To achieve the optimal solution, we propagate sufficient statistics in closed form to evaluate the expected cost, then minimize this cost with respect to the filter and control gains. Our results demonstrate that this approach achieves a lower overall cost than state-of-the-art solutions when internal noise is considered. Deriving the optimal control law in these cases is essential for addressing problems like inverse control inference.

1 Introduction and Formalization of the Problem

The ability to achieve goal-oriented actions stems from the sensorimotor system's exceptional capacity to manage noise and the variety of potential solutions for each problem [2, 3, 4]. However, the computational and algorithmic implementation of a regulatory control—necessary for producing coordinated and complex behavior [5]—remains an open question in systems neuroscience.

NeurIPS 2024 Workshop on Behavioral Machine Learning.

Stochastic optimal control theory serves as a powerful tool for interpreting motor behavior [6, 5, 1, 7], but the success of this approach depends on the mathematical accuracy of the predictions and the assumptions [1].

To solve an optimal control problem, one must identify a state-to-action policy that minimizes a cost function, typically balancing control effort with task objectives [6]. The Linear-Quadratic-Gaussian (LQG) framework allows for analytical solutions in stochastic, partially observable, continuous, non-stationary, high-dimensional systems [8, 9]. However, incorporating a more realistic noise model, including multiplicative and internal noise, is crucial for explaining many observed behaviors, such as smooth velocity profiles and movement corrections [1, 10, 11, 12, 3, 13].

The seminal study in [1] developed an iterative algorithm to solve stochastic optimal control problems, including motor, sensory, and internal noise. However, this algorithm relies on the assumption of unbiased estimators, which is not accurate in this case. We introduce a gradient descent-based algorithm that circumvents this issue, relying solely on the assumption of linear control [14]. Our algorithm is efficient, leveraging recursive cost computation and novel derivative propagation methods. When applied to a sensorimotor task, it outperforms the solution of [1] in the presence of internal noise, reducing the cost by up to 90% when the internal noise accounts for approximately 10% of the total. Applying our algorithm to a reaching task reveals distinct patterns, highlighting the importance of providing the optimal solution to explain behavior in a principled way [14, 15].

Formalization of the Optimal Control Problem We formalize our problem using stochastic optimal control theory [8, 9]. Following [1], we assume that the latent state dynamics (with multiplicative control noise), observations (with multiplicative sensory noise), and state estimation (with internal additive noise) of the full controller-environment system obeys the following equations in discrete time ($t = 1, ..., T-1$)

$$x_{t+1} = Ax_t + Bu_t + \xi_t + \sum_{i=1}^{c} \varepsilon_t^i C_i u_t \tag{1}$$

$$\hat{x}_{t+1} = A\hat{x}_t + Bu_t + K_t(y_t - H\hat{x}_t) + \eta_t \tag{2}$$

$$y_t = Hx_t + \omega_t + \sum_{i=1}^{d} \rho_t^i D_i x_t \ , \tag{3}$$

where $x_t \in \mathbb{R}^m$ is the latent state, $\hat{x}_t \in \mathbb{R}^m$ is the state estimate, $u_t \in \mathbb{R}^p$ is the control signal, $y_t \in \mathbb{R}^k$ is the sensory feedback, at time t, and ξ_t, ε_t, η_t, ω_t and ρ_t are white noise terms described in Appendix A.1, following [1]. The linear dynamics are governed by the passive state-to-state transition matrix $A \in \mathbb{R}^{m\times m}$ and the control matrix $B \in \mathbb{R}^{m\times p}$. By assumption, the control-law is linear in the state estimate,

$$u_t = L_t \hat{x}_t \ , \tag{4}$$

where $L_t \in \mathbb{R}^{p\times m}$ are the control gains, a common result for LQG and the approach of [1]. We furthermore assume a time-dependent, but state-independent Kalman filter $K_t \in \mathbb{R}^{m\times k}$, which is again standard – it is exact for LQG problems with additive noise but becomes approximate under multiplicative noise. Note that, by using the same set of matrices to update the state (Eq. 1) and state estimate (Eq. 2), we implicitly assume that the internal model of the agent already matches the actual latent dynamics of the system. The state estimate arises from a linear Kalman filter integrating and filtering the sensory feedback y_t. Such a feedback is considered to be a noisy linear transformation of the state, as determined by the observation matrix $H \in \mathbb{R}^{k\times m}$. We will assume, as it is also standard, that the initial state x_1 and state estimate $\hat{x}_1$ have the same mean $\mathbb{E}[x_1]$, with covariances respectively given by Σ_{x_1} and $\Sigma_{\hat{x}_1}$. Moreover, they are assumed to be independent at $t = 1$: their joint covariance is $\Sigma_{x_1\hat{x}_1} = 0$. Solving the optimal control problem implies finding the optimal filters K_t and control gains L_t in order to minimize a cost function given by

$$\mathbb{E}[J] = \sum_{t=1}^{T} \mathbb{E}[j_t] = \sum_{t=1}^{T} \mathbb{E}\left[x_t^\intercal Q_t x_t + u_t^\intercal R_t u_t\right] \ , \tag{5}$$

where T is the duration of the task, j_t is the cost per step and J the total cost in a trial. We define the expectation as $\mathbb{E}[f(\cdot)] = \int dx_{2,...,T}\ d\hat{x}_{2,...,T} f(\cdot)\ p(x_{2,...,T}, \hat{x}_{2,...,T})$, where p is the joint density of latent and estimation variables. As the control u_t is a linear function of $\hat{x}_t$, the cost is a quadratic

function of state and state estimate. It includes a control cost, determined by the symmetric positive definite matrix $R_t \in \mathbb{R}^{p\times p}$, $R_t > 0$, and a state cost, determined by $Q_t \in \mathbb{R}^{m\times m}$. Again, Q_t is symmetric and positive definite, $Q_t > 0$, and modulates the cost of the state being far from a chosen target. In Appendix A.2 we provide more details on the LQG problem and on the state-of-the-art solution for the multiplicative noise scenario, expanding on what said before.

2 A Novel Gradient Descent-Based Algorithm for Optimal Control Problems

2.1 How Internal Noise Affects Optimality

Here, we briefly address the challenges posed by internal noise in the algorithm from [1]. As noted in Section 1, [1] uses the unbiased estimator condition, $\mathbb{E}[x_t|\hat{x}_t] = \hat{x}_t$, to derive the optimal control. However, this condition never holds. In Appendix A.3, we show numerically how this condition fails, with increasing discrepancies as internal noise grows. To heuristically prove this, we consider a one-dimensional problem with a partially observable stochastic process x_t. If $\mathbb{E}[x_{t-1}|\hat{x}_{t-1}] = \hat{x}_{t-1}$ holds at time $t-1$, and a large positive internal noise fluctuation $\eta_{t-1} \gg 1$ occurs, then at time t the state x_t has on average changed little relative to x_{t-1}, but the state estimate has increased significantly due to the fluctuation, so that $\hat{x}_t \gg \hat{x}_{t-1}$. Then, at time t, the averaged x_t conditioned on $\hat{x}_t$ cannot be equal to $\hat{x}_t$, breaking the unbiasedness condition (Fig.1). Notably, the same issue arises with large sensory noise fluctuations, even without internal noise. Furthermore, this effect is more pronounced with large fluctuations in the state estimate, although a smaller bias persists even with minor fluctuations.

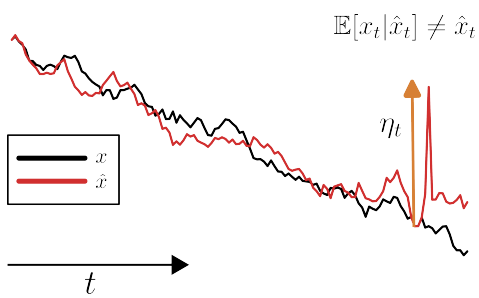

Figure 1: *A toy example to show the estimation bias*. The black line, x_t, represents the behavior of a partially observable stochastic process. The red line, $\hat{x}_t$ stands for the time evolution of a state estimate, built upon the available observations, biased by random internal fluctuations due to the noise term η_t (orange arrow).

In Appendix A.3.1, we demonstrate (in 1D for simplicity) that without internal noise, the unbiasedness condition implies the orthogonality principle, $\mathbb{E}[x_t\hat{x}_t] = \mathbb{E}[\hat{x}_t^2]$, which holds for an optimal Kalman filter (see Fig. 5a). Thus, [1] finds optimal solutions in the absence of internal noise by implicitly imposing this principle. However, when internal noise is present, this principle is no longer optimal (see Fig. 5a), causing the algorithm to underperform in minimizing cost.

2.2 A Novel Algorithm for Optimal Control Problems

We propose an alternative method to solve the optimal control problem outlined in Section 1, following [14]. We compute the expected total accumulated cost, $\mathbb{E}[J]$, while averaging over all stochastic terms appearing in Eqs. 1-3, as a function of L_t and K_t. With L_t and K_t fixed, $\mathbb{E}[J]$ becomes the objective function for a standard gradient descent algorithm used to minimize it. In Appendix A.4, we show how to compute the afore-mentioned objective function in closed form by moment propagation, and discuss how to minimize it with respect to $L_{1,\dots,T-1}$ and $K_{1,\dots,T-2}$. We also derive the computationally cheaper analytical counterpart of this algorithm for a more constrained definition of the problem in Appendix A.5. This results in an algorithm that iteratively alternates between forward moment propagation and backward optimization of control and estimation until convergence. We stress that our algorithm derives the optimal solutions without relying on the incorrect unbiasedness condition. Here and in the following we call GD the numerical algorithm

introduced in Appendix A.4, FPOMP ('Fixed Point Optimization with Moments Propagation') the analytically-derived one, proposed in Appendix A.5, and TOD the algorithm of [1].

We also note that our framework is exact: no approximations are needed to compute the expectation of the cost function, apart from the linearity of control. This is due to the linearity of control and estimation processes (in both the state and state estimate), allowing the first two moments of x and $\hat{x}$ to be propagated in closed form. These moments serve as sufficient statistics for computing the expected cost, independent of the distributions of x and $\hat{x}$.

In Appendix A.8 we extend our approach to switching linear dynamics, to make the linear dynamics assumption less restrictive.

3 Experiments

We first apply our algorithm and compare it with the solutions of [1] in a 1D example (see Appendix A.9.1) to show the qualitative and quantitative differences. Second, we apply our numerical approach to a sensorimotor task, a single-joint reaching movement, with a four-dimensional state, one-dimensional control and sensory feedback ($m = 4$, $p = k = 1$) and multiplicative and internal noise, equivalent to the task implemented in [1].

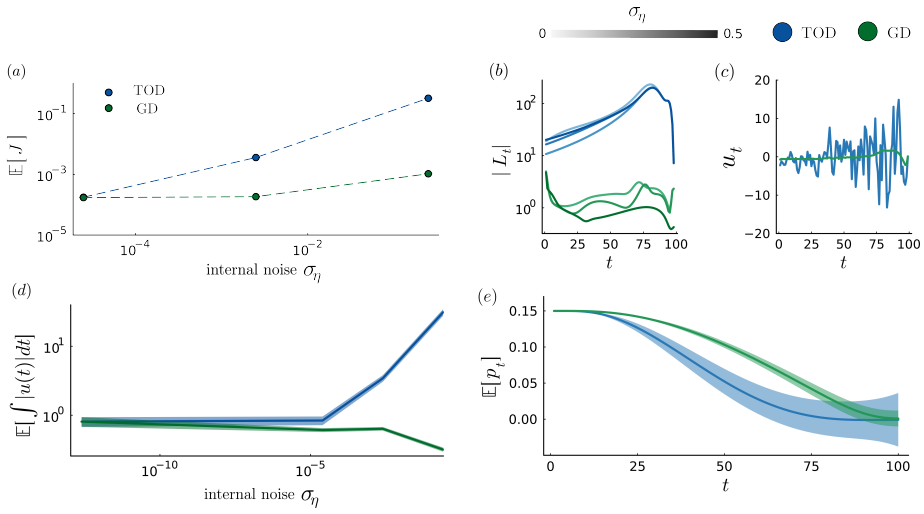

Figure 2: *Reaching task*. **(a)** Expected accumulated cost $\mathbb{E}[J]$ (average of Eq. 5 over $50k$ trials) as a function of internal noise, with error bars (mean $\pm$ 1SEM from Monte Carlo simulations, error bars not visible as too small), for TOD and GD algorithms. **(b)** Magnitude of the control gain vector as a function of time for TOD and GD solutions. **(c)** Control signal $u(t)$ in a sample trial for the two algorithms for $\sigma_\eta = 0.05$. **(d)** Amount of control (mean integral of the absolute control signal for the two algorithms, averaged over $50k$ trials) as a function of σ_η. **(e)** Mean position over time for the two solutions, averaged over $50k$ trials.

The definition of the task is the same as in [1], with the only difference that [1] did not assume any internal noise, whereas we do. All parameters of the problem can be found in Appendix A.9.8. The solutions found by the GD algorithm lead to a lower expected accumulated cost $\mathbb{E}[J]$, with a performance gap increasing with the level of internal noise, σ_η (Fig. 2a). This is achieved by lowering the control gains as σ_η increases (Fig. 2b), leading to a significantly smoother control signal on individual trials (Fig. 2c) and an overall reduction in control effort (Fig. 2d). From these predictions, two additional behavioral features emerge: movements become slower than with TOD solutions, and trial-to-trial variability is significantly reduced (Fig. 2e). The performance improvement of the GD algorithm is around 90% when the relative importance of internal noise is $\approx 10\%$.

In Appendices A.9.1, A.9.4 and A.9.7, we show that the FPOPM algorithm introduced in Appendix A.5 and explicitly derived in Appendices A.7.1 and A.7.2, matches the solution of the numerical GD algorithm. For the multi-dimensional case, we empirically validate it in the simplified case without

multiplicative noise, as discussed in Appendix A.7.2. In Appendix A.9.9 we show how our algorithm scales to higher-dimensional problems.

4 Conclusion

We propose a gradient descent-based algorithm for solving stochastic optimal control problems that incorporate a realistic noise model of the sensorimotor system. By including control and signal-dependent noise, along with internal noise in the estimation process, our approach extends, as outlined in [1], the classic LQG framework to capture more realistic scenarios, albeit with reduced mathematical tractability. Developing models that can generate and explain complex behaviors is essential both in behavioral science and AI, particularly in creating real-time controllers. Solving control problems with such models provides quantitative insights into the computational mechanisms underlying behavior.

The key study in [1] derived an algorithm to determine the optimal feedback control law and non-adaptive filters within this extended noise model. This solution is currently used in inverse optimal control to interpret human behavior, with impacts that reach beyond theoretical aspects [15, 16, 17, 18, 19, 20, 21]. Unfortunately, the algorithm's derivation has a limitation due to its assumption of unbiased estimators. This results in suboptimal solutions when internal noise is introduced, and before algorithmic convergence, as discussed in Section 2, due to the breakdown of the orthogonality principle.

Under the assumption that control is linearly dependent on the current state estimate, we develop a revised algorithm that optimizes both control and estimation using gradient descent [14]. This approach calculates the expected cost by propagating the sufficient statistics, followed by minimizing this cost with respect to the filter and control gains. From the analytical point of view, this results in an algorithm that achieves convergence by iteratively alternating two main steps: forward propagation of moments and backward optimization of control and estimation. We demonstrate that our approach results in lower costs, thus outperforming the current state-of-the-art solution in the presence of internal noise (and at fixed filter gains, regardless of the presence of internal noise, see Appendix A.9.6), providing mathematical and heuristic explanations for this enhanced performance. Optimal filtering of internal fluctuations is achieved through an intertwined modulation of control and filter gains, thereby enhancing adaptability, as discussed in Appendix A.9.3: when optimized for high levels of internal noise, the system generalizes well to other noise levels.

We apply our algorithm to a reaching task, producing novel behavioral predictions that distinguish our solution from that in [1]. As internal noise rises, control gains decrease, resulting in smoother control signals within each trial. Consequently, movements become slower, with reduced variability observed across trials.

Overall, our algorithm enhances the applicability of optimal feedback control, with notable benefits for inverse optimal control applications [15].

Limitations and Future Work A limitation of our work is the assumption of state-independent filter gains, which is suboptimal for multiplicative noise. Future directions include exploring more realistic cost functions and biologically plausible learning rules. We also only rely on [1] for exponential convergence guarantees. Next, we plan to extend the FPOMP algorithm to the case with multiplicative noise and integrate it into the inverse optimal control framework to test behavioral data predictions.

Acknowledgments and Disclosure of Funding

This work was supported by: NEI-NIH R01 EY016178; Grant PRE2021-097778, funded by MICIU/AEI/10.13039/501100011033 and by “ESF+”; "Project PID2023-146524NB-I00 financed by MCIN/AEI/10.13039/501100011033/ ERDF, EU, the Spanish State Research Agency and the European Union, and ICREA Academia.

References

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

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

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

[4] Emmanuel Guigon, Pierre Baraduc, and Michel Desmurget. Computational motor control: redundancy and invariance. *Journal of neurophysiology*, 97(1):331–347, 2007.

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

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

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

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

[9] Robert F Stengel. *Optimal control and estimation*. Courier Corporation, 1994.

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

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

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

[13] Joseph Y Nashed, Frédéric Crevecoeur, and Stephen H Scott. Influence of the behavioral goal and environmental obstacles on rapid feedback responses. *Journal of neurophysiology*, 108(4):999–1009, 2012.

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

[15] 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.

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

[17] 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.

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

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

[20] 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.

[21] 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.

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

[23] 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.

[24] Kelvin E Jones, Antonia F de C Hamilton, and Daniel M Wolpert. Sources of signal-dependent noise during isometric force production. *Journal of neurophysiology*, 88(3):1533–1544, 2002.

[25] 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.

[26] 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.

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

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

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

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

[31] Andrew M Gordon, Goran Westling, Kelly J Cole, and Roland S Johansson. Memory representations underlying motor commands used during manipulation of common and novel objects. *Journal of neurophysiology*, 69(6):1789–1796, 1993.

[32] Philippe Vindras and Paolo Viviani. Frames of reference and control parameters in visuomanual pointing. *Journal of Experimental Psychology: Human Perception and Performance*, 24(2):569, 1998.

[33] Sanjoy K Mitter. Filtering and stochastic control: A historical perspective. *IEEE Control Systems Magazine*, 16(3):67–76, 1996.

[34] Tryphon T Georgiou and Anders Lindquist. The separation principle in stochastic control, redux. *IEEE Transactions on Automatic Control*, 58(10):2481–2494, 2013.

[35] D Peter Joseph and T Julius Tou. On linear control theory. *Transactions of the American Institute of Electrical Engineers, Part II: Applications and Industry*, 80(4):193–196, 1961.

[36] Henk Van de Water and Jan Willems. The certainty equivalence property in stochastic control theory. *IEEE Transactions on Automatic Control*, 26(5):1080–1087, 1981.

[37] Kaare Brandt Petersen, Michael Syskind Pedersen, et al. The matrix cookbook. *Technical University of Denmark*, 7(15):510, 2008.

[38] Philip Becker-Ehmck, Jan Peters, and Patrick Van Der Smagt. Switching linear dynamics for variational bayes filtering. In *International conference on machine learning*, pages 553–562. PMLR, 2019.

A Appendix

A.1 A Realistic Noise Model of the Sensorimotor System

Following [1], we consider the system dynamics to be perturbed by additive and multiplicative noise sources, affecting both the latent state x_t and sensory feedback y_t. The stochasticity on the state dynamics is induced by the additive Gaussian noise $\xi_t \in \mathbb{R}^m$, with zero mean and covariance $\Omega_\xi \geq 0$, and the control-dependent multiplicative noise term $C(u_t) = \sum_{i=1}^{c} \varepsilon_t^i C_i u_t$. Following the notation in [1], C_i are constant scaling matrices, $C_i \in \mathbb{R}^{m \times p}$, multiplying the i-th component ε_t^i of the random vector $\varepsilon_t \in \mathbb{R}^c$, where ε_t is a zero mean Gaussian noise term with identity covariance $\Omega_\varepsilon = \mathbb{I}$. Such a term takes into account the fact that noise at the motor output level increases linearly with the control signal [1, 22, 6, 23, 24]. In our case, this dependence is linear in the state estimate $\hat{x}_t$, given that u_t is given by Eq. 4. Similarly, at the sensory feedback level, there is additive Gaussian noise $\omega_t \in \mathbb{R}^k$, with zero mean and covariance $\Omega_\omega \geq 0$, and the signal-dependent term $\sum_{i=1}^{d} \rho_t^i D_i x_t$. Again, $\rho_t \in \mathbb{R}^d$ is a zero mean Gaussian noise term with covariance $\Omega_\rho = \mathbb{I}$, and $D_i \in \mathbb{R}^{k \times m}$ are the scaling matrices controlling the linear dependence between noise strength and sensory feedback signal. Such a relationship has an experimental counterpart when considering vision as the main sensory modality, since sensory noise linearly scales with visual eccentricity [1, 25, 26, 27]. Here we extend the same linear dependence to other possible perceptual modes, such as proprioception [11]. More generally, multiplicative noise sources are required to mimic experimentally observed properties of reaching movements, such as stereotyped bell-shaped profiles [1, 10, 11], and speed-accuracy trade-off [1, 22, 28, 23]. Crucially, we assume state estimate computations , Eq. 2, to be perturbed by zero-mean additive Gaussian noise, $\eta_t \in \mathbb{R}^m$, with covariance $\Omega_\eta \geq 0$. This term links experimental evidence at the neural [2, 29, 30, 3] and behavioral [1, 31, 32] level. Importantly, η_t cannot be directly filtered by the gains K_t, leading to possible control-estimation inter-dependencies, as we see in Sections 2 and 3 — a fact that has been overlooked by previous approaches.

A.2 LQG Problem and State-of-the-Art Solutions for the Multiplicative Noise Scenario

When only additive noise at state and feedback levels is considered in the system of Eqs. 1-3 (that is when setting $C_i = 0, i = 1, ..., c$, $D_i = 0, i = 1, ..., d$ and $\Omega_\eta = 0$), the problem reduces to the Linear-Quadratic-Gaussian (LQG) problem [8]. It is known to have non-approximate analytical solutions [8], which are for completeness provided in Appendix A.2.1. Its analytical tractability stems from the mathematical independence between control and estimation optimizations [1], known as the separation principle [33, 34, 35], which is closely related to the concept of certainty equivalence [36]. Indeed, Eqs. 6 and 8 demonstrate that L_t and K_t can be solved for independently. The only dependence between control and estimation arises through the state estimate $\hat{x}_t$. Once multiplicative control and signal-dependent noise is present, such independence breaks down, causing control and estimation to be closely intertwined [1]. For this scenario, [1] proposes an algorithm that alternates between optimizing control and estimation while assuming linear and non-adaptive (i.e, state-independent) filters for the estimator. As for the classic LQG problem, the control is optimized iteratively in a backward-in-time fashion, while keeping the filters K_t fixed. The solution is derived by using the method of dynamic programming, writing down the Bellman equation for the optimal cost-to-go [1]. In this derivation, multiplicative noise causes the optimal u_t to depend on the latent state x_t and the condition $\mathbb{E}[x_t|\hat{x}_t] = \hat{x}_t$ is then used to ensure that the actual control signal would depend only on the state estimate $\hat{x}_t$ (which is the only available information at time t, given that the system is only partially observable by the definition of the problem). From there, the optimal filters K_t are found at fixed L_t, again by minimizing the cost-to-go. Taken together, these two optimization steps lead to an iterative algorithm that is supposed to provide the optimal solution to the control problem [1].

However, the assumption of "unbiasedness" $\mathbb{E}[x_t|\hat{x}_t] = \hat{x}_t$ that is used to derive the optimal control law in [1] does not hold (see Section 2).

A.2.1 LQG Solutions

The optimal control and filter gains, $L_{1,...,T-1}$ and $K_{1,...,T-2}$, for the LQG problem, that is the control problem described by Eqs. 1-2 without multiplicative ($C_i = 0$, $i = 1, ..., c$, $D_i = 0$, $i = 1, ..., d$) and

internal ($\Omega_\eta = 0$) noise are given by

$$L_t = -(R_t + B^\intercal S_{t+1} B)^{-1} B^\intercal S_{t+1} A \tag{6}$$

$$S_t = Q_t + A^\intercal S_{t+1}(A - BL_t) \tag{7}$$

$$K_t = A\Sigma_t^e H^\intercal (H\Sigma_t^e H^\intercal + \Omega_\omega)^{-1} \tag{8}$$

$$\Sigma_{t+1}^e = \Omega_\xi + (A - K_t H)\Sigma_t^e A^\intercal \ . \tag{9}$$

A.3 The Condition of Unbiasedness and the Role of Internal Noise

As mentioned in Sections 1 and 2, the condition $\mathbb{E}[x_t|\hat{x}_t] = \hat{x}_t$ used in [1] to derive the optimal control law is violated. We demonstrate this concept numerically, considering a $1D$ example, with the same system dynamics as the one of Appendix A.9.1 (Table 2). We set $\sigma_\omega = \sigma_\xi = \sigma_\epsilon = \sigma_\rho = 0.5$ and varied $\sigma_\eta = 0.0, 0.3, 0.6$. The initial condition are $x_0 = \hat{x}_0 = 1$. We considered $T = 10$ and selected $t = 8$ to compute $\mathbb{E}[x_t|\hat{x}_t]$ as a function of $\hat{x}_t$.

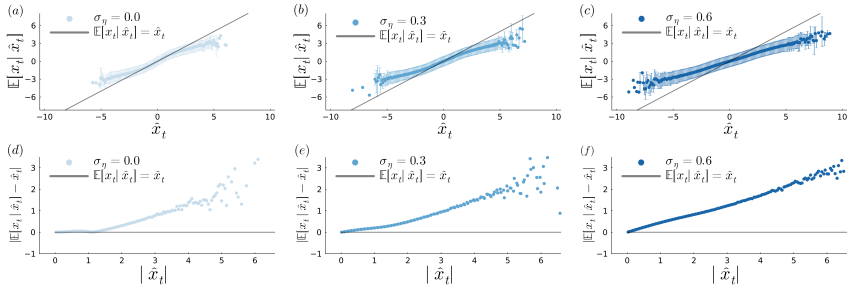

Figure 3: *The invalidity of the unbiasedness condition*. Here we plot $\mathbb{E}[x_t|\hat{x}_t]$, for a given value of t, as a function of the values $\hat{x}_t$ for different levels of internal noise σ_η, for the algorithm in [1]. The conditional expectation $\mathbb{E}[x_t|\hat{x}_t]$ is computed through Monte Carlo (MC) simulations. **(a)** $\mathbb{E}[x_t|\hat{x}_t]$ as a function of $\hat{x}_t$ for $\sigma_\eta = 0$ (dots with error bars given by the std of our MC estimate). The gray dotted line stands for the bisector, where $\mathbb{E}[x_t|\hat{x}_t] = \hat{x}_t$. Note that for big values of $\hat{x}_t$ the incorrectness of the condition of unbiasedness is visible. **(b)** Same as (a), but for $\sigma_\eta = 0.3$. With internal noise the deviation from $\mathbb{E}[x_t|\hat{x}_t] = \hat{x}_t$ is more salient. **(c)** Same as (a) and (b), but for $\sigma_\eta = 0.6$. As the internal noise level increases the condition of unbiasedness is more and more incorrect. **(d-f)** Absolute value of the distance between $\mathbb{E}[x_t|\hat{x}_t]$ and $\hat{x}_t$ as a function of $|\hat{x}_t|$ for $\sigma_\eta = 0.0, 0.3, 0.6$. The gray dotted lines represent $\mathbb{E}[x_t|\hat{x}_t] = \hat{x}_t$.

Note that the choice of t is arbitrary. We collected the list of $x_{t=8}$ and $\hat{x}_{t=8}$ over $5 \cdot 10^7$ trials. We then binned the data for $\hat{x}_t$ using $\delta\hat{x} = 0.1$ for the size of the bins. To obtain $\mathbb{E}[x_t|\hat{x}_t]$ we then computed the mean of all the x_t falling in the same bin, and we used the std for the error bars.

We find that, without internal noise, the violation of unbiasedness is present, but it is only clearly visible for large $\hat{x}_t$, because close to $\hat{x}_t = 0$ the bias will be small (the state estimate experiences small fluctuations if we constrain to small values of $\hat{x}_t$ – this is also why we chose in our example of Fig. 1 a large fluctuation, but the same bias, albeit small, happen with small sensory or internal noise perturbations). We also find that, when considering internal noise, the bias sensibly increases because the internal fluctuations are not filtered at all by the gains K_t.

It is relevant to observe that the filter gains cannot directly reduce the internal noise. The only way to filter internal fluctuations passes through the control optimization, given that the control signal u_t is a linear function of the state estimate $\hat{x}_t$ (Eq. 4). As we see in Section 3, the optimal L_t and K_t have to be interdependent when internal noise is turned on. In these circumstances, the separation principle [8] does not hold anymore, regardless of the multiplicative nature of the noise. On the contrary, in the algorithm proposed in [1], such a dependence arises only when multiplicative noise is considered, as we show in Appendix A.9.7.

Additionally, we note here that the condition $\mathbb{E}[x_t|\hat{x}_t] = \hat{x}_t$ is closely related to the orthogonality principle, stating $\Omega_t = \mathbb{E}[\hat{x}_t\hat{x}_t^\intercal] - \mathbb{E}[x_t\hat{x}_t^\intercal] = 0, \forall t = 1, ..., T$ [8]. As we show in Appendix A.3.1, the latter can be derived from the former. As a confirmation, the optimal estimator found in [1] satisfies this condition in the absence of internal noise. Therefore, for zero internal noise, $\Omega_\eta = 0$, the

control-estimation optimization can be run with two distinct objective functions: K_t can be optimized by imposing $\Omega_t = 0, \forall t = 1, ..., T$ (see Appendix A.3.1), while L_t comes from the minimization of the cost function $\mathbb{E}[J]$ (Eq. 19). It can be shown (see Appendix A.3.1) that when internal noise is non-zero, $\Omega_\eta > 0$, a solution satisfying $\Omega_t = 0, \forall t = 1, ..., T$ is not guaranteed to exist and, in any case, is not optimal (see Fig. 5a).

Separation Principle, Orthogonality Principle, Unbiasedness: A Brief Digression Unbiasedness, orthogonality and separation principle are related but distinct concepts. Here we briefly elaborate on the differences and point in commons between them.

The separation principle comes from the formulation of the classic LQG problem: the optimal solutions for control and estimator are mathematically independent, allowing for separate optimization of the two. With multiplicative noise, this independence breaks down [1]. We argue that this happens already with additive internal noise, contrary to what said in [1].

The orthogonality principle (in 1D) states that $\mathbb{E}[x_t \hat{x}_t] = \mathbb{E}[\hat{x}_t^2]$. This condition holds for an optimal Kalman filter, without internal noise (Appendix A.3.1 and Fig. 5a). Then, internal fluctuations break the mathematical independence between control and estimation, invalidating at the same time the orthogonality principle.

However, these two concepts are separated: e.g., if we have no internal noise, but non zero multiplicative noise, the orthogonality principle would still hold, but the mathematical independence between control and estimation would break.

In Appendix A.3.1, we show that without internal noise, the orthogonality principle is satisfied when K_t follows Eq. 15. Empirical validation of the optimality of this comes from finding the same K_t with GD, TOD, and Eq. 15 (without internal noise).

The condition of unbiasedness (which never holds, as previously discussed) implies that $\mathbb{E}[x_t|\hat{x}_t] = \hat{x}_t$. In Appendix A.3.1 we show that if $\mathbb{E}[x_t|\hat{x}_t] = \hat{x}_t$ holds, then also the orthogonality principle does. This explains the optimality of the approach in [1] in the absence of internal noise: not because the unbiasedness condition is true, but because the orthogonality condition holds.

A.3.1 The Relationship Between Internal Noise and the Orthogonality Principle

We demonstrate here that from the condition $\mathbb{E}[x_t|\hat{x}_t] = \hat{x}_t$, used in [1] to derive the optimal control law, we can derive the orthogonality principle. For simplicity, we consider here the one-dimensional case in which $m = p = k = 1$ and we set $c = d = 1$, as done in [14].
The orthogonality principle states that [8]

$$\Omega_t \equiv \mathbb{E}[\hat{x}_t(x_t - \hat{x}_t)] = \mathbb{E}[\hat{x}_t^2] - \mathbb{E}[x_t \hat{x}_t] = 0. \tag{10}$$

Taking

$$\mathbb{E}[x_t|\hat{x}_t] = \hat{x}_t \tag{11}$$

and multiplying by $\hat{x}_t$ on both sides and then taking the expectation over $\hat{x}_t$ we obtain

$$\mathbb{E}[\hat{x}_t^2] = \mathbb{E}[x_t \hat{x}_t] \tag{12}$$

corresponding to Eq. 10.

We also show here that, in the absence of internal noise, the optimal filter gains K_t can be found by imposing the orthogonality principle, without the need to minimize the cost function. Using Eq. 27,

$$\mathbb{S}_{t+1} = M_t \mathbb{S} M_t^\intercal + G_t \tag{13}$$

for the update of the non-central moments we get

$$\begin{aligned} \Omega_{t+1} = &(K_t^2 H^2 + K_t^2 D^2 - A K_t H)\mathbb{E}[x_t^2] + \\ &+ (A^2 + K_t^2 H^2 + A B L_t - 2 A K_t H - B L_t K_t H)\Omega_t + \\ &+ (A K_t H - K_t^2 H^2)\mathbb{E}[x_t \hat{x}_t] + K_t^2 \Omega_\omega + \Omega_\eta \end{aligned} \tag{14}$$

If we use $\Omega_{t=1} = 0$, as in [1] due to the initial conditions, we can solve the equation $\Omega_t = 0$, $\forall t = 1, ..., T$, obtaining an equation for K_t,

$$K_t = \frac{A H \Gamma_t \pm \sqrt{A^2 H^2 \Gamma_t^2 - 4(H^2 \Gamma_t + D^2 \mathbb{E}[x_t^2] + \Omega_\omega)\Omega_\eta}}{2(H^2 \Gamma_t + D^2 \mathbb{E}[x_t^2] + \Omega_\omega)} \tag{15}$$

with

$$\Gamma_t = \mathbb{E}[x_t^2] - \mathbb{E}[x_t \hat{x}_t]. \tag{16}$$

For $\Omega_\eta = 0$, Eq. 15 simplifies to

$$K_t = \frac{AH\Gamma_t \pm AH\Gamma_t}{2(H^2\Gamma_t + D^2\mathbb{E}[x_t^2] + \Omega_\omega)}. \tag{17}$$

Observing that the solution $K_t = 0$ would correspond to an open-loop strategy, sub-optimal (sensory information would not be integrated) for a stochastic partially observable system as the one we are considering, we get for the optimal filter gains

$$K_t^* = \frac{AH\Gamma_t}{H^2\Gamma_t + D^2\mathbb{E}[x_t^2] + \Omega_\omega}. \tag{18}$$

It can be proven that the solution of [1] for $\Omega_\eta = 0$ aligns with Eq. 18 in one dimension. We observe that Eq. 18 can be used in place of Eq. 39 in Algorithm 2 to optimize the filter gains. For $\Omega_\eta = 0$ this would lead to the optimal solution. Therefore, in the absence of internal noise, the optimization of control and estimation can be performed with two separate objective functions, one imposing the orthogonality principle for the optimal estimator and the other minimizing the cost function for the optimal controller, regardless of the multiplicative nature of the noise. This could also be relevant for more biologically plausible scenarios [5]. When $\Omega_\eta > 0$, it can be shown that the existence of a real solution for Eq. 15 depends on the choice of the initial conditions. Moreover, we show in Appendix A.9.1 that the optimal solutions found by the numerical gradient descent do not satisfy $\Omega_t = 0$ for $\Omega_\eta > 0$. Therefore, it turns out that the orthogonality principle holds for an optimal Kalman filter only for $\Omega_\eta = 0$. Consequently, the algorithm derived in [1] uses the assumption of unbiased estimation, which should imply the orthogonality principle, even in situations where the orthogonality principle no longer applies to the optimal estimator.

A.4 Minimization of Expected Cost Through Numerical Gradient Descent (GD)

By taking the expected value of Eq. 5 and using Eq. 4 we obtain

$$\begin{aligned}\mathbb{E}[J] = \sum_{t=1}^{T} \mathbb{E}[j_t] = \sum_{t=1}^{T} (&\mathbb{E}[x_t]^\intercal Q_t \mathbb{E}[x_t] + \mathbb{E}[\hat{x}_t]^\intercal L_t^\intercal R_t L_t \mathbb{E}[\hat{x}_t] + \\ &+ Tr[Q_t \Sigma_{x_t}] + Tr[L_t^\intercal R_t L_t \Sigma_{\hat{x}_t}]),\end{aligned} \tag{19}$$

where $Tr[\cdot]$ stands for the trace operation, Σ_{x_t} is the covariance matrix of the latent state x_t and $\Sigma_{\hat{x}_t}$ is the covariance of the state estimate at t. Note that $\mathbb{E}[x_t]$, $\mathbb{E}[\hat{x}_t]$, Σ_{x_t} and $\Sigma_{\hat{x}_t}$ will implicitly depend on $L_{1,\dots,t-1}$ and $K_{1,\dots,t-1}$. From Eqs. 1-3 we can derive the update equations to propagate the first and second-order moments $\mathbb{E}[x_t]$ $\mathbb{E}[\hat{x}_t]$, Σ_{x_t} and $\Sigma_{\hat{x}_t}$ in a closed-form manner, in order to compute the total expected cost $\mathbb{E}[J]$ at fixed $L_{1,\dots,T-1}$ and $K_{1,\dots,T-2}$. Here and in the following we set $c = d = 1$ for simplicity (and without loss of generality, given the case with $c, d > 1$ would follow the same exact procedure). To rewrite our results in a more compact form, we define, as in [14]

$$\mu_t = \begin{pmatrix} \mathbb{E}[x_t] \\ \mathbb{E}[\hat{x}_t] \end{pmatrix}, \tag{20}$$

$$\Sigma_t = \begin{pmatrix} \Sigma_{x_t} & \Sigma_{x_t,\hat{x}_t} \\ \Sigma_{\hat{x}_t,x_t} & \Sigma_{\hat{x}_t} \end{pmatrix}, \tag{21}$$

$$M_t = \begin{pmatrix} A & BL_t \\ K_t H & A + BL_t - K_t H \end{pmatrix} \tag{22}$$

and

$$G_t = \begin{pmatrix} CL_t\mathbb{E}[\hat{x}_t\hat{x}_t^\intercal]L_t^\intercal C^\intercal + \Omega_\xi & 0 \\ 0 & K_t D\mathbb{E}[x_t x_t^\intercal]D^\intercal K_t^\intercal + K_t\Omega_\omega K_t^\intercal + \Omega_\eta \end{pmatrix}, \tag{23}$$

where $\Sigma_{x_t} = \mathbb{E}[x_t x_t^\intercal] - \mathbb{E}[x_t]\mathbb{E}[x_t]^\intercal$, $\Sigma_{\hat{x}_t} = \mathbb{E}[\hat{x}_t\hat{x}_t^\intercal] - \mathbb{E}[\hat{x}_t]\mathbb{E}[\hat{x}_t]^\intercal$, $\Sigma_{x_t,\hat{x}_t} = \mathbb{E}[x_t\hat{x}_t^\intercal] - \mathbb{E}[x_t]\mathbb{E}[\hat{x}_t]^\intercal$ and $\Sigma_{\hat{x}_t,x_t} = \Sigma_{x_t,\hat{x}_t}^\intercal$. We have defined μ_t as a column vector whose components are $m-$dimensional vectors. Similarly, Σ_t, M_t and G_t are block matrices, whose elements are $m \times m$ matrices. We then have

$$\mu_{t+1} = M_t \mu_t \tag{24}$$

and

$$\Sigma_{t+1} = M_t \Sigma_t M_t^\intercal + G_t \ . \tag{25}$$

In the presence of multiplicative noise, the covariance propagation, Eq. 25, furthermore requires us to derive update equations for the non-central moments that appear in G_t. We can proceed in a similar way by defining

$$\mathbb{S}_t = \begin{pmatrix} \mathbb{E}[x_t x_t^\intercal] & \mathbb{E}[x_t \hat{x}_t^\intercal] \\ \mathbb{E}[\hat{x}_t x_t^\intercal] & \mathbb{E}[\hat{x}_t \hat{x}_t^\intercal] \end{pmatrix} . \tag{26}$$

We then have

$$\mathbb{S}_{t+1} = M_t \mathbb{S}_t M_t^\intercal + G_t \ , \tag{27}$$

where G_t is now given by

$$G_t = \begin{pmatrix} C L_t \mathbb{S}_t^{22} L_t^\intercal C^\intercal + \Omega_\xi & 0 \\ 0 & K_t D \mathbb{S}_t^{11} D^\intercal K_t^\intercal + K_t \Omega_\omega K_t^\intercal + \Omega_\eta \end{pmatrix} , \tag{28}$$

and $\mathbb{S}_t^{ij}$ stands for the (i, j)-th block element of the block matrix $\mathbb{S}_t$.

We assume, as in [1], that state and state estimate are initially uncorrelated, $\Sigma_{x_t, \hat{x}_t} = \Sigma_{\hat{x}_t, x_t} = 0$. As a result, given the initial conditions for μ_1 and Σ_1, we can compute the expected accumulated cost $\mathbb{E}[J]$ at fixed $L_{1,\dots,T-1}$ and $K_{1,\dots,T-2}$, by using Eqs. 24-25 and 27 together with Eq. 19. The pseudo-code for the algorithm to compute the expected cost $\mathbb{E}[J]$ and therefore to implement the numerical gradient descent is provided (with the details of the implementation) in Appendix A.6, Algorithm 1. To find the optimal control and filter gains we would then use $\mathbb{E}[J]$, given by Algorithm 1, as the objective function of a numerical gradient descent procedure. The analytical counterpart is discussed in Appendix A.5.

We stress here that framework is exact: no approximations are needed to compute the expectation of the cost function and therefore to find the optimal solutions.

Indeed, Eqs. 24-25 hold true even if the whole process is not Gaussian (the distributions of x_t and $\hat{x}_t$ are not Gaussian due to the multiplicative noise in Eqs. 1-3). This is because control (Eq. 4) and estimation processes are linear (in state and state estimate), which enables the propagation of the first two moments of x and $\hat{x}$ in closed form, regardless of the nature of the distribution. If this is violated, higher order terms would appear in the propagation of the expected cost and an approximation of Gaussianity might be needed. Note also that the noise terms are assumed to be uncorrelated one with respect to the other and with zero mean [1].

Moreover, given that the cost function is quadratic in u_t and x_t, the first two moments of $(x, \hat{x})$ serve as sufficient statistics, allowing us to compute the expected cost without requiring higher-order moments.

Derivation of Closed-Form Equations for Moments Propagation Eqs. 24-27 can be derived from Eqs. 1-4 by taking the expected value of $x_t, \hat{x}_t, x_t \hat{x}_t^\intercal, x_t x_t^\intercal, \hat{x}_t \hat{x}_t^\intercal$ over the joint distribution of state, state estimate and sensory feedback, without assuming any form for the underlying distribution. Indeed, we only need to compute the first two moments of the joint variable $(x_t, \hat{x}_t)$.

By taking the expected value of Eqs. 1-2 we obtain

$$\mathbb{E}[x_{t+1}] = A\mathbb{E}[x_t] + BL_t \mathbb{E}[\hat{x}_t] \tag{29}$$

$$\mathbb{E}[\hat{x}_{t+1}] = K_t H \mathbb{E}[x_t] + (A + BL_t - K_t H)\mathbb{E}[\hat{x}_t] \ , \tag{30}$$

which correspond to Eq. 24. Similarly, we compute the second non-central moments of the joint variable $(x, \hat{x})$, resulting in

$$\begin{aligned} \mathbb{E}[x_{t+1} x_{t+1}^\intercal] &= A\mathbb{E}[x_t x_t^\intercal] A^\intercal + BL_t \mathbb{E}[\hat{x}_t \hat{x}_t^\intercal] L_t^\intercal B^\intercal + \\ &+ A\mathbb{E}[x_t \hat{x}_t^\intercal] L_t^\intercal B^\intercal + BL_t \mathbb{E}[\hat{x}_t x_t^\intercal] A^\intercal + CL_t \mathbb{E}[\hat{x}_t \hat{x}_t^\intercal] L_t^\intercal C^\intercal + \Omega_\xi \end{aligned} \tag{31}$$

$$\begin{aligned} \mathbb{E}[\hat{x}_{t+1} \hat{x}_{t+1}^\intercal] &= K_t H \mathbb{E}[x_t x_t^\intercal] H^\intercal K_t^\intercal + (A + BL_t - K_t H)\mathbb{E}[\hat{x}_t \hat{x}_t^\intercal](A + BL_t - K_t H)^\intercal + \\ &+ K_t H \mathbb{E}[x_t \hat{x}_t^\intercal](A + BL_t - K_t H)^\intercal + (A + BL_t - K_t H)\mathbb{E}[\hat{x}_t x_t^\intercal] H^\intercal K_t^\intercal + \\ &+ K_t D \mathbb{E}[x_t x_t^\intercal] D^\intercal K_t^\intercal + K_t \Omega_\omega K_t^\intercal + \Omega_\eta \end{aligned} \tag{32}$$

$$\begin{aligned} \mathbb{E}[\hat{x}_{t+1} x_{t+1}^\intercal] &= K_t H \mathbb{E}[x_t x_t^\intercal] A^\intercal + (A + BL_t - K_t H)\mathbb{E}[\hat{x}_t \hat{x}_t^\intercal] L_t^\intercal B^\intercal + \\ &+ K_t H \mathbb{E}[x_t \hat{x}_t^\intercal] L_t^\intercal B^\intercal + (A + BL_t - K_t H)\mathbb{E}[\hat{x}_t x_t^\intercal] A^\intercal \end{aligned} \tag{33}$$

$$\mathbb{E}[x_{t+1} \hat{x}_{t+1}^\intercal] = \mathbb{E}[\hat{x}_{t+1} x_{t+1}^\intercal]^\intercal \ . \tag{34}$$

Given the definitions of $\mu_t, \Sigma_t, \mathbb{S}_t$ and of the block matrices M_t, G_t one can rewrite the equations above in the form of Eqs. 24-27.

A.5 An Analytical Approach: FPOMP Algorithm

For complex and realistic tasks, finding the optimal control and filter gains through numerical gradient descent as described in the previous Section (Appendix A.4) can become computationally expensive. As $L_{1,...,T-1} \in \mathbb{R}^{p\times m}$ and $K_{1,...,T-2} \in \mathbb{R}^{m\times k}$, we would have a total of $mp(T-1) + mk(T-2)$ parameters to optimize, which can be large for a problem with a high-dimensional state.

We propose here an analytically-derived algorithm, where we alternate between finding the optimal (i.e., cost-minimizing) L_t and K_t, denoted L_t^* and K_t^*, for fixed state and state estimate moments, μ_t and Σ_t, and re-computing these moments in light of the updated L_t's and K_t's. We refer to this method as the 'Fixed Point Optimization with Moments Propagation' (FPOMP) algorithm. We follow the mathematical derivation presented in [14]. We can compute the expected cost per step at time $t+i$, $i = 0, ..., T-t$, conditioned to the moments Σ_t and μ_t as

$$\begin{aligned}\mathbb{E}[j_{t+i}|\mu_t, \Sigma_t] = \mathbb{E}[x_{t+i}|\mu_t, \Sigma_t]^\intercal Q_{t+i}\mathbb{E}[x_{t+i}|\mu_t, \Sigma_t] + \\ + \mathbb{E}[\hat{x}_{t+i}|\mu_t, \Sigma_t]^\intercal L_{t+i}^\intercal R_{t+i} L_{t+i}\mathbb{E}[\hat{x}_{t+i}|\mu_t, \Sigma_t] + \\ + Tr[Q_{t+i}\Sigma_{x_{t+i}|\mu_t,\Sigma_t}] + Tr[L_{t+i}^\intercal R_{t+i} L_{t+i}\Sigma_{\hat{x}_{t+i}|\mu_t,\Sigma_t}]\ ,\end{aligned} \tag{35}$$

where $\mathbb{E}[x_{t+i}|\mu_t, \Sigma_t]$, $\mathbb{E}[x_{t+i}|\mu_t, \Sigma_t]$, $\Sigma_{x_{t+i}|\mu_t,\Sigma_t}$ and $\Sigma_{\hat{x}_{t+i}|\mu_t,\Sigma_t}$ are computed by propagating the moments μ_t and Σ_t (Eqs. 24-25) until $\tilde{t} = t+i$. Indeed, μ_t and Σ_t are the only necessary information to propagate the expected cost, at fixed control and filter gains, and thus they are sufficient statistics.

We then set to zero the derivatives of the expected cost Eq. 19, excluding the constant terms, to derive L_t^* and K_t^*

$$\frac{\partial}{\partial L_t}\sum_{i=0}^{T-t}\mathbb{E}[j_{t+i}|\mu_t, \Sigma_t] = \sum_{i=0}^{T-t}\frac{\partial}{\partial L_t}\mathbb{E}[j_{t+i}|\mu_t, \Sigma_t] = 0 \tag{36}$$

$$\frac{\partial}{\partial K_t}\sum_{i=1}^{T-t}\mathbb{E}[j_{t+i}|\mu_t, \Sigma_t] = \sum_{i=1}^{T-t}\frac{\partial}{\partial K_t}\mathbb{E}[j_{t+i}|\mu_t, \Sigma_t] = 0\ . \tag{37}$$

As shown in Appendices A.7.1 and A.7.2, solving Eqs. 36-37 leads to a backward algorithm to compute L_t^* and K_t^*,

$$L_t^* = f(\mu_t, \Sigma_t, L_{t+1,...,T-1}^*, K_{t+1,...,T-2}^*) \tag{38}$$

$$K_t^* = g(\mu_t, \Sigma_t, L_{t+1,...,T-1}^*, K_{t+1,...,T-2}^*)\ , \tag{39}$$

with $t = 1, ...T-1$ for L_t^* and $t = 1, ...T-2$ for K_t^*. From this we can build an algorithm that, starting from an initial guess for $L_{1,...,T-1}^*$ and $K_{1,...,T-2}^*$, iteratively computes all the moments $\mu_{1,...,T-1}$ and $\Sigma_{1,...,T-1}$ (Eqs. 24-25) at fixed $L_{1,...,T-1}^*$ and $K_{1,...,T-2}^*$. Given those moments, $L_{1,...,T-1}^*$ and $K_{1,...,T-2}^*$ are updated by using Eqs. 38-39, and so on, until convergence is attained. The pseudo-code for this analytical gradient descent algorithm can be found in Appendix A.6, Algorithm 2. In such a way, we eliminate the numerical optimization procedure, making the algorithm suitable for realistic optimal control problems. In Appendix A.7.1, we explicitly solve Eqs. 36-37 for the one-dimensional case, while in Appendix A.7.2 we extend the approach to a multi-dimensional scenario, considering, for the sake of simplicity, the classic LQG problem (but including internal noise), to prove the generalizability of Algorithm 2.

A.6 Pseudo-Codes

Algorithm 1 Propagation of the expected - GD algorithm

Input: μ_1, Σ_1, $\mathbb{S}_1$ (initial conditions), $L_{1,\dots,T-1}$, $K_{1,\dots,T-2}$, and the system parameters (A, B, H, $C_{i=1,\dots,c}$, $D_{i=1,\dots,d}$, Ω_ξ, Ω_ω, Ω_η).
Output: $\mathbb{E}[J]$
Algorithm steps:
$\mathbb{E}[J] = 0$
$\mu_{old} = \mu_1$
$\Sigma_{old} = \Sigma_1$
$\mathbb{S}_{old} = \mathbb{S}_1$
for each iteration $t = 1, 2, \dots, T$ **do**
  $E[J] \leftarrow \mathbb{E}[J] + \mathbb{E}[j_t]$, (Eq. 19)
  Update M_t and G_t
  $\Sigma_{new} = M_t \Sigma_{old} M_t^\intercal + G_t$
  $\mu_{new} = M_t \mu_{old}$
  $\mathbb{S}_{new} = M_t \mathbb{S}_{old} M_t^\intercal + G_t$
  $\Sigma_{old} \leftarrow \Sigma_{new}$
  $\mu_{old} \leftarrow \mu_{new}$
  $\mathbb{S}_{old} \leftarrow \mathbb{S}_{new}$
end for

Algorithm 2 FPOMP algorithm

Input: μ_1, Σ_1, $\mathbb{S}_1$ (initial conditions), $L^{(1)}_{1,\dots,T-1}$, $K^{(1)}_{1,\dots,T-2}$ (initial guesses for L^* and K^*), and the system parameters (A, B, H, $C_{i=1,\dots,c}$, $D_{i=1,\dots,d}$, Ω_ξ, Ω_ω, Ω_η).
Output: $L^*_{1,\dots,T-1}$, $K^*_{1,\dots,T-2}$ (optimal control and filter gains)
Algorithm steps:
for each iteration $k = 2, \dots,$ optimization steps **do**
  $\mu_{1,\dots,T-1}, \Sigma_{1,\dots,T-1} \leftarrow$ Eqs. 24 and 25 using $L^{(k-1)}_{1,\dots,T-1}$ and $K^{(k-1)}_{1,\dots,T-2}$
  for each iteration $i = 1, \dots, T-1$ **do**
    $t \leftarrow T - i$
    $L^{(k)}_t \leftarrow f(\mu_t, \Sigma_t, L^{(k)}_{t+1,\dots,T-1}, K^{(k-1)}_{t+1,\dots,T-2})$
    $K^{(k)}_t \leftarrow g(\mu_t, \Sigma_t, L^{(k-1)}_{t+1,\dots,T-1}, K^{(k)}_{t+1,\dots,T-2})$
  end for
end for
$L^*_{1,\dots,T-1} \leftarrow L^{(k)}_{1,\dots,T-1}$
$K^*_{1,\dots,T-2} \leftarrow K^{(k)}_{1,\dots,T-2}$

In the pseudo-code, $L^{(k)}_t$ and $K^{(k)}_t$ stand for, respectively, the control and filter gains at time t and at optimization step k.

The hyper-parameters of all the used algorithms are provided in Table 1. For the GD algorithm (Appendix A.4) we minimize the expected accumulated cost $\mathbb{E}[J]$, computed through Algorithm 1, using the function "GradientDescent()" in the "Optim.jl" Julia package.

Table 1: Hyper-parameters of the used algorithms

Algorithm	Description	value
GD	Number of iterations of the "GradientDescent()" function	100000
FPOMP	Number of iterations of the control-estimation optimization	1000
TOD	Number of iterations of the control-estimation optimization	1000

A.7 Derivations of the FPOMP Algorithm

We derive here the explicit form of Eqs. 38-39 to implement Algorithm 2, following [14].

A.7.1 One-Dimensional Case

In the one-dimensional case we have $m = p = k = 1$. Additionally, to simplify the notation, we set $c = d = 1$. We start by defining

$$\vec{F}_t = \begin{pmatrix} F_{t,1} \\ F_{t,2} \\ F_{t,3} \end{pmatrix} = \begin{pmatrix} A^2 \\ (B^2 + C^2)L_t^2 \\ 2ABL_t \end{pmatrix} \tag{40}$$

$$\vec{G}_t = \begin{pmatrix} G_{t,1} \\ G_{t,2} \\ G_{t,3} \end{pmatrix} = \begin{pmatrix} K_t^2(H^2 + D^2) \\ (A + BL_t)^2 + K_t^2 H^2 - 2AK_tH - 2BL_tK_tH \\ 2BL_tK_tH + 2AK_tH - 2K_t^2H^2 \end{pmatrix} \tag{41}$$

$$\vec{H}_t = \begin{pmatrix} H_{t,1} \\ H_{t,2} \\ H_{t,3} \end{pmatrix} = \begin{pmatrix} AK_tH \\ ABL_t + B^2L_t^2 - BL_tK_tH \\ A^2 + ABL_t - AK_tH + BL_tK_tH \end{pmatrix} . \tag{42}$$

In one dimension, we can then rewrite Eq. 27 as

$$\mathbb{E}[x_{t+1}^2] = F_{t,1}\mathbb{E}[x_t^2] + F_{t,2}\mathbb{E}[\hat{x}_t^2] + F_{t,3}\mathbb{E}[x_t\hat{x}_t] + \Omega_\xi \tag{43}$$

$$\mathbb{E}[\hat{x}_{t+1}^2] = G_{t,1}\mathbb{E}[x_t^2] + G_{t,2}\mathbb{E}[\hat{x}_t^2] + G_{t,3}\mathbb{E}[x_t\hat{x}_t] + K_t^2\Omega_\omega + \Omega_\eta \tag{44}$$

$$\mathbb{E}[x_{t+1}\hat{x}_{t+1}] = H_{t,1}\mathbb{E}[x_t^2] + H_{t,2}\mathbb{E}[\hat{x}_t^2] + H_{t,3}\mathbb{E}[x_t\hat{x}_t] \, . \tag{45}$$

The derivatives of the non-central moments with respect to L_t and K_t obey the following equations, for $i = 1, ..., T$,

$$\frac{\partial \mathbb{E}[x_{t+i}^2]}{\partial L_t} = a_{t+i-1,1}L_t + b_{t+i-1,1} \tag{46}$$

$$\frac{\partial \mathbb{E}[\hat{x}_{t+i}^2]}{\partial L_t} = a_{t+i-1,2}L_t + b_{t+i-1,2} \tag{47}$$

$$\frac{\partial \mathbb{E}[x_{t+i}\hat{x}_{t+i}]}{\partial L_t} = a_{t+i-1,3}L_t + b_{t+i-1,3} \tag{48}$$

and

$$\frac{\partial \mathbb{E}[x_{t+i}^2]}{\partial K_t} = \alpha_{t+i-1,1}K_t + \beta_{t+i-1,1} \tag{49}$$

$$\frac{\partial \mathbb{E}[\hat{x}_{t+i}^2]}{\partial K_t} = \alpha_{t+i-1,2}K_t + \beta_{t+i-1,2} \tag{50}$$

$$\frac{\partial \mathbb{E}[x_{t+i}\hat{x}_{t+i}]}{\partial K_t} = \alpha_{t+i-1,3}K_t + \beta_{t+i-1,3} \, , \tag{51}$$

with $\vec{a}$, $\vec{b}$, $\vec{\alpha}$ and $\vec{\beta}$ given by the following recursive equations

$$\vec{a}_{t+1} = \begin{pmatrix} a_{t+1,1} \\ a_{t+1,2} \\ a_{t+1,3} \end{pmatrix} = \begin{pmatrix} \vec{F}_{t+1} \cdot \vec{a}_t \\ \vec{G}_{t+1} \cdot \vec{a}_t \\ \vec{H}_{t+1} \cdot \vec{a}_t \end{pmatrix} \tag{52}$$

$$\vec{b}_{t+1} = \begin{pmatrix} b_{t+1,1} \\ b_{t+1,2} \\ b_{t+1,3} \end{pmatrix} = \begin{pmatrix} \vec{F}_{t+1} \cdot \vec{b}_t \\ \vec{G}_{t+1} \cdot \vec{b}_t \\ \vec{H}_{t+1} \cdot \vec{b}_t \end{pmatrix} \tag{53}$$

$$\vec{\alpha}_{t+1} = \begin{pmatrix} \alpha_{t+1,1} \\ \alpha_{t+1,2} \\ \alpha_{t+1,3} \end{pmatrix} = \begin{pmatrix} \vec{F}_{t+1} \cdot \vec{\alpha}_t \\ \vec{G}_{t+1} \cdot \vec{\alpha}_t \\ \vec{H}_{t+1} \cdot \vec{\alpha}_t \end{pmatrix} \tag{54}$$

$$\vec{\beta}_{t+1} = \begin{pmatrix} \beta_{t+1,1} \\ \beta_{t+1,2} \\ \beta_{t+1,3} \end{pmatrix} = \begin{pmatrix} \vec{F}_{t+1} \cdot \vec{\beta}_t \\ \vec{G}_{t+1} \cdot \vec{\beta}_t \\ \vec{H}_{t+1} \cdot \vec{\beta}_t \end{pmatrix} . \tag{55}$$

The initial conditions for Eqs. 52- 55 are

$$\vec{a}_t = \begin{pmatrix} 2(B^2+C^2)\mathbb{E}[\hat{x}_t^2] \\ 2B^2\mathbb{E}[\hat{x}_t^2] \\ 2B^2\mathbb{E}[\hat{x}_t^2] \end{pmatrix} \tag{56}$$

$$\vec{b}_t = \begin{pmatrix} 2AB\mathbb{E}[x_t\hat{x}_t] \\ 2AB\mathbb{E}[\hat{x}_t^2] - 2BK_tH(\mathbb{E}[\hat{x}_t^2] - \mathbb{E}[x_t\hat{x}_t]) \\ AB(\mathbb{E}[\hat{x}_t^2] + \mathbb{E}[x_t\hat{x}_t]) - BK_tH(\mathbb{E}[\hat{x}_t^2] - \mathbb{E}[x_t\hat{x}_t]) \end{pmatrix} \tag{57}$$

$$\vec{\alpha}_t = \begin{pmatrix} 0 \\ 2H^2(\mathbb{E}[x_t^2] + \mathbb{E}[\hat{x}_t^2] - 2\mathbb{E}[x_t\hat{x}_t]) + 2\Omega_\omega + 2D^2\mathbb{E}[x_t^2] \\ 0 \end{pmatrix} \tag{58}$$

$$\vec{\beta}_t = \begin{pmatrix} 0 \\ -2H(A+BL_t)(\mathbb{E}[\hat{x}_t^2] - \mathbb{E}[x_t\hat{x}_t]) \\ AH(\mathbb{E}[x_t^2] - \mathbb{E}[x_t\hat{x}_t]) - BL_tH(\mathbb{E}[\hat{x}_t^2] - \mathbb{E}[x_t\hat{x}_t]) \end{pmatrix} . \tag{59}$$

By observing that the expected accumulated cost, Eq. 19 (adapted to the one-dimensional case), will be a function of $\mathbb{E}[x_t^2]$ and $\mathbb{E}[\hat{x}_t^2]$, for $t = 1, ..., T-t$, and by using Eqs. 46-51, we can rewrite Eqs. 36-37 as

$$\begin{aligned} \frac{\partial}{\partial L_t}\sum_{i=0}^{T-t}\mathbb{E}[j_{t+i}|\mu_t,\Sigma_t] = 2R_t\mathbb{E}[\hat{x}_t^2]L_t + \\ + \sum_{i=1}^{T-t}[(Q_{t+i}a_{t+i-1,1} + R_{t+i}L_{t+i}^2a_{t+i-1,2})L_t + \\ + (Q_{t+i}b_{t+i-1,1} + R_{t+i}L_{t+i}^2b_{t+i-1,2})] = 0 \end{aligned} \tag{60}$$

and

$$\begin{aligned} \frac{\partial}{\partial K_t}\sum_{i=0}^{T-t}\mathbb{E}[j_{t+i}|\mu_t,\Sigma_t] = \sum_{i=1}^{T-t}[(Q_{t+i}\alpha_{t+i-1,1} + R_{t+i}L_{t+i}^2\alpha_{t+i-1,2})K_t + \\ + (Q_{t+i}\beta_{t+i-1,1} + R_{t+i}L_{t+i}^2\beta_{t+i-1,2})] = 0 \ . \end{aligned} \tag{61}$$

Therefore, from Eqs. 60-61, we have the following instantiations of Eqs. 38-39 for the optimal control and filter gains at time t, L_t^* and K_t^*,

$$L_t^* = -\frac{L_t^{num}}{L_t^{den}} \tag{62}$$

$$K_t^* = -\frac{K_t^{num}}{K_t^{den}} \ , \tag{63}$$

with

$$L_t^{num} = \sum_{i=1}^{T-t}\left(Q_{t+i}b_{t+i-1,1} + R_{t+i}L_{t+i}^2b_{t+i-1,2}\right) , \tag{64}$$

$$\begin{aligned} L_t^{den} = 2R_t\mathbb{E}[\hat{x}_t^2] + \\ + \sum_{i=1}^{T-t}(Q_{t+i}a_{t+i-1,1} + R_{t+i}L_{t+i}^2a_{t+i-1,2}) \end{aligned} \tag{65}$$

and

$$K_t^{num} = \sum_{i=1}^{T-t}\left(Q_{t+i}\beta_{t+i-1,1} + R_{t+i}L_{t+i}^2\beta_{t+i-1,2}\right) , \tag{66}$$

$$K_t^{den} = \sum_{i=1}^{T-t}\left(Q_{t+i}\alpha_{t+i-1,1} + R_{t+i}L_{t+i}^2\alpha_{t+i-1,2}\right) . \tag{67}$$

We can then use Eqs. 62-63, to implement Algorithm 2 and extract $L_{1,...,T-1}^*$, and $K_{1,...,T-2}^*$, for the one-dimensional problem.

A.7.2 Multi-Dimensional Case

For the multi-dimensional case, we derive Eqs. 38-39 for the classic LQG problem ($C_i = 0$ for $i =, .., c$ and $D_i = 0$ for $i =, .., d$) in the presence of internal noise ($\Omega_\eta \geq 0$) [14].

As a title of example, we derive here Eq. 38 for the optimal L_t^* (to be used in Algorithm 2), but the approach would be the same for the optimal filter gains K_t^*. The extension to the more general scenario including the multiplicative sources of noise would follow the same method. As outlined in Appendix A.4, Eq. 19, the expected cost per step is given by

$$\begin{aligned}\mathbb{E}[j_{t+i}] = \mathbb{E}[x_{t+i}]^\intercal Q_{t+i}\mathbb{E}[x_{t+i}] + \mathbb{E}[\hat{x}_{t+i}]^\intercal L_t^\intercal R_t L_{t+i}\mathbb{E}[\hat{x}_{t+i}] + \\ + Tr[Q_{t+i}\Sigma_{x_{t+i}}] + Tr[L_{t+i}^\intercal R_{t+i} L_{t+i}\Sigma_{\hat{x}_{t+i}}],\end{aligned} \tag{68}$$

for $i = 0, ..., T - t$.

When computing $\mathbb{E}[j_{t+i}|\mu_t, \Sigma_t]$ to write down Eq. 36 (with $C_i = 0$, $i = 1, ..., c$ and $D_i = 0$, $i = 1, ..., d$), and derive Eq. 38, the coefficients multiplying $\mathbb{E}[\hat{x}_t]\mathbb{E}[\hat{x}_t]^\intercal$ coming from the term $\mathbb{E}[x_{t+i}]^\intercal Q_{t+i}\mathbb{E}[x_{t+i}]$ in Eq. 68 will be the same as the ones multiplying $\Sigma_{\hat{x}_t}$ and coming from the term $Tr[Q_{t+i}\Sigma_{x_{t+i}}]$. The same holds for the coefficients multiplying respectively $\mathbb{E}[x_t]\mathbb{E}[\hat{x}_t]^\intercal$ and $\Sigma_{x_t,\hat{x}_t}$.

Similarly, we can group together the coefficients coming from the other two factors $\mathbb{E}[\hat{x}_{t+i}]^\intercal L_t^\intercal R_t L_{t+i}\mathbb{E}[\hat{x}_{t+i}]$ and $Tr[L_{t+i}^\intercal R_{t+i} L_{t+i}^\intercal \Sigma_{\hat{x}_{t+i}}]$ in Eq. 68.

We now note that the terms dependent on L_t appearing in $\mathbb{E}[j_{t+i}|\mu_t, \Sigma_t]$ will show a dependence on the afore-mentioned moments $\mathbb{E}[\hat{x}_t]\mathbb{E}[\hat{x}_t]^\intercal$, $\Sigma_{\hat{x}_t}$, $\mathbb{E}[x_t]\mathbb{E}[\hat{x}_t]^\intercal$ and $\Sigma_{x_t,\hat{x}_t}$. More specifically, the quadratic factors in L_t will only depend on $\mathbb{E}[\hat{x}_t]\mathbb{E}[\hat{x}_t]^\intercal$ and $\Sigma_{\hat{x}_t}$. Taken together, these observations lead to the following form for Eq. 36,

$$\mathcal{J}_t L_t^* \mathbb{E}[\hat{x}_t\hat{x}_t^\intercal] + \mathcal{S}_t\mathbb{E}[x_t\hat{x}_t^\intercal] + \mathcal{P}_t\mathbb{E}[\hat{x}_t\hat{x}_t^\intercal] = 0, \tag{69}$$

where we have used $\Sigma_{\hat{x}_t} + \mathbb{E}[\hat{x}_t]\mathbb{E}[\hat{x}_t]^\intercal = \mathbb{E}[\hat{x}_t\hat{x}_t^\intercal]$ and $\Sigma_{x_t,\hat{x}_t} + \mathbb{E}[x_t]\mathbb{E}[\hat{x}_t]^\intercal = \mathbb{E}[x_t\hat{x}_t^\intercal]$.

Therefore, to find the optimal control gains L_t^* from Eq. 69, we only need to compute the coefficients $\mathcal{J}_t$, $\mathcal{S}_t$ and $\mathcal{P}_t$, similar to what we have done for the one-dimensional case in Appendix A.7.1. As before, we can compute the coefficients $\mathcal{J}_t$, $\mathcal{S}_t$ and $\mathcal{P}_t$ by only looking at the first two terms appearing in Eq. 68, that is $\mathbb{E}[x_{t+i}]^\intercal Q_{t+i}\mathbb{E}[x_{t+i}]$ and $\mathbb{E}[\hat{x}_{t+i}]^\intercal L_t^\intercal R_t L_{t+i}\mathbb{E}[\hat{x}_{t+i}]$. By using ([37])

$$\frac{\partial \vec{v}^\intercal X \vec{w}}{\partial X} = \vec{v}\vec{w}^\intercal, \tag{70}$$

$$\frac{\partial \vec{v}^\intercal X^\intercal \vec{w}}{\partial X} = \vec{w}\vec{v}^\intercal, \tag{71}$$

$$\frac{\partial}{\partial X}(\vec{v}^\intercal X^\intercal N X \vec{v}) = 2NX\vec{v}\vec{v}^\intercal, \tag{72}$$

where $\vec{v}$ and $\vec{w}$ are vectors and N is a symmetric matrix, we obtain

$$\mathcal{J}_t = 2R_t + 2\sum_{i=1}^{T-t}\left[V_{t+i-1}^\intercal(Q_{t+i} + L_{t+i}^\intercal R_{t+i}L_{t+i})V_{t+i-1}\right] \tag{73}$$

$$\mathcal{S}_t = 2\sum_{i=1}^{T-t}\left\{V_{t+i-1}^\intercal\left[Q_{t+i}\left(\mu^{t+i}_{L_t=0,(\mathbb{I},0)}\right)_1 + L_{t+i}^\intercal R_{t+i}L_{t+i}\left(\mu^{t+i}_{L_t=0,(\mathbb{I},0)}\right)_2\right]\right\} \tag{74}$$

$$\mathcal{P}_t = 2\sum_{i=1}^{T-t}\left\{V_{t+i-1}^\intercal\left[Q_{t+i}\left(\mu^{t+i}_{L_t=0,(0,\mathbb{I})}\right)_1 + L_{t+i}^\intercal R_{t+i}L_{t+i}\left(\mu^{t+i}_{L_t=0,(0,\mathbb{I})}\right)_2\right]\right\} \tag{75}$$

with V_{t+i} given by

$$V_{t+i} = \prod_{j=1}^{i}(A + BL_{t+j})B \tag{76}$$

for $i = 1, ..., T - t$, and

$$V_t = B\ . \tag{77}$$

In Eqs. 74-75, $\left(\mu^{t+i}_{L_t=0,(\cdot,\cdot)}\right)$, is a vector whose elements are $m \times m$ matrices:

$$\mu^{t+i}_{L_t=0,(\cdot,\cdot)} = \begin{pmatrix} \left(\mu^{t+i}_{L_t=0,(\cdot,\cdot)}\right)_1 \\ \left(\mu^{t+i}_{L_t=0,(\cdot,\cdot)}\right)_2 \end{pmatrix} \tag{78}$$

The subscript $(\cdot,\cdot)$ indicates the initial condition ($i = 0$) for the evolution of $\mu^{t+i}_{L_t=0,(\cdot,\cdot)}$, with $\mathbb{I}$ denoting the $m \times m$ identity matrix and 0 being an $m \times m$ matrix whose elements are all zeros, e.g.,

$$\mu^{t}_{L_t=0,(\mathbb{I},0)} = \begin{pmatrix} \mathbb{I} \\ 0 \end{pmatrix}. \tag{79}$$

$\mu^{t+i}_{L_t=0,(\cdot,\cdot)}$ is updated through the following equations

$$\mu^{t+i}_{L_t=0,(\cdot,\cdot)} = \begin{cases} \tilde{M}_t \mu^{t}_{L_t=0,(\cdot,\cdot)}, & \text{for } i = 1 \\ M_{t+i-1} \mu^{t+i-1}_{L_t=0,(\cdot,\cdot)}, & \text{for } i = 2, ..., T-t \end{cases} \tag{80}$$

with M_t given by Eq. 22 and $\tilde{M}_t$ having the same form as the block matrix M_t, but with $L_t = 0$,

$$\tilde{M}_t = \begin{pmatrix} A & 0 \\ K_t H & A - K_t H \end{pmatrix}. \tag{81}$$

From Eq. 69 we can then write for Eq. 38

$$L^*_t = -\mathcal{J}_t^{-1} \left(\mathcal{S}_t \mathbb{E}[x_t \hat{x}_t^\intercal] + \mathcal{P}_t \mathbb{E}[\hat{x}_t \hat{x}_t^\intercal]\right) \mathbb{E}[\hat{x}_t \hat{x}_t^\intercal]^\dagger \tag{82}$$

where $\cdot^\dagger$ denotes the pseudoinverse operation. Note that $\mathcal{J}_t$ is a symmetric $p \times p$ matrix with $det[\mathcal{J}_t] > 0$ and therefore invertible. Due to the initial conditions for $\Sigma_{\hat{x}_1}$ and $\mathbb{E}[\hat{x}_1]$, the symmetric matrix $\mathbb{E}[\hat{x}_t \hat{x}_t^\intercal]$ could have a null determinant: for this reason we use the pseudoinverse operation. This consideration is relevant only for an initial transient: after a certain time $\tilde{t} > 0$ (depending on the dynamics parameters) we would have $det[\mathbb{E}[\hat{x}_{\tilde{t}} \hat{x}_{\tilde{t}}^\intercal]] > 0$ and $\mathbb{E}[\hat{x}_t \hat{x}_t^\intercal]^\dagger = \mathbb{E}[\hat{x}_t \hat{x}_t^\intercal]^{-1}$, due to the properties of the pseudoinverse. With Eq. 82 we can implement Algorithm 2 to find the optimal control gains $L^*_{1,...,T-1}$. The derivation of Eq. 39 for the optimal filter gains K^*_t would follow the same procedure. To extend the presented approach to the case with multiplicative noise, we need to propagate the terms depending on C_i, $i = 1, ..., c$ and D_i, $i = 1, ..., d$ coming from the factors $Tr[Q_{t+i} \Sigma_{x_{t+i}}]$ and $Tr[L^\intercal_{t+i} R_{t+i} L_{t+i} \Sigma_{\hat{x}_{t+i}}]$ in Eq. 68, similarly to what we have done with the other terms in Eq. 80, but using ([37])

$$\frac{\partial}{\partial X} Tr[\tilde{A} X \tilde{B}] = \tilde{A}^\intercal \tilde{B}^\intercal, \tag{83}$$

$$\frac{\partial}{\partial X} Tr[\tilde{A} X^\intercal \tilde{B}] = \tilde{B} \tilde{A}, \tag{84}$$

$$\frac{\partial}{\partial X} Tr[\tilde{A} X \tilde{B} X^\intercal \tilde{C}] = \tilde{A}^\intercal \tilde{C}^\intercal X \tilde{B}^\intercal + \tilde{C} \tilde{A} X \tilde{B} \tag{85}$$

where $\tilde{A}$, $\tilde{B}$ and $\tilde{C}$ are matrices. We observe that, even when considering multiplicative noise, Eq. 82 will still be valid: only the matrices $\mathcal{J}_t$, $\mathcal{S}_t$ and $\mathcal{P}_t$ will change, including now also the terms depending on C_i, $i = 1, ..., c$, and D_i, $i = 1, ..., d$.

From the form of Eq. 82, we can see why, mathematically, the control gains decrease when the internal noise level is increased: the factor $\mathbb{E}[\hat{x}_t \hat{x}_t^\intercal]$ will get bigger and bigger as Ω_η gets larger.

A.8 Switching Linear Dynamics

We discuss here how to extend our approach to switching linear dynamics. One of the underlying assumptions in this work and in [1], as outlined in Section 1, is that the agent has complete knowledge of the updating rules of the latent dynamical system. By using the same set of matrices to update the state and the state estimate, we implicitly assume that all uncertainty in the estimation process arises solely from noise sources: the problem of inferring the matrices A and B goes beyond the objectives of this approach. For this reason, to extend our work to the more general and realistic case of switching linear dynamics (SLD), we can consider a matrix A depending on the time step t, A_t.

A complete formulation of SLD might require adding another variable, a discrete switch variable s_t regulating the way the matrices A_t vary with time and context (see [38]). Given that in our case the agent has access to the updating rules of the dynamical system, we can omit s_t (the agent does not have to infer s_t and A_t) and directly consider the case in which we have a predetermined set of matrices $A_{1,\dots T-1}$. The same applies to the matrix B, that can be replaced by $B_{1,\dots T-1}$. Note that to preserve linearity we assume A_t and B_t to be independent on x and $\hat{x}$. We consider here the multidimensional case to be as general as possible. To extend the GD algorithm we only need to modify the block matrix M_t that we use to update the moments $\Sigma_t, \mathbb{S}_t, \mu_t$ and eventually propagate the expected cost $\mathbb{E}[J]$ through Eq. 19. Indeed, once we can compute the expected cost at fixed control and filter gains, $L_{1,\dots,T-1}$, and $K_{1,\dots T-2}$, we can use Algorithm 1 to define the objective function to be minimized through Gradient Descent with respect to L_t and K_t. To update the block matrix M_t we have to substitute A and B respectively with $A_{1,\dots T-1}$ and $B_{1,\dots T-1}$ in Eq. 22. To handle the potentially high computational costs of performing a numerical Gradient Descent, we introduced the analytical counterpart of the GD algorithm, the FPOMP algorithm. For the one-dimensional case, it supports all the noise sources mentioned in Section 1 and Appendix A.1 (additive, multiplicative and internal). We extended this algorithm to the multi-dimensional case for additive and internal noise in Appendix A.7.2 for the sake of simplicity, leaving the more general version for future work (Appendix A.7.2 outlines how this can be done). Here, we extend the afore-mentioned FPOMP algorithm (for both one-dimensional and multi-dimensional cases) to switching linear dynamics, following a similar procedure to that of the numerical algorithm. For the one-dimensional case, we replace A and B respectively with A_t and B_t in Eqs. 40-42,and 56-59. For the multi-dimensional case we have to substitute A with A_{t+j} and B with B_{t+j} in Eq. 76 and B with B_t in Eq. 77. Finally, as previously done, we replace A with A_t in Eq. 81 for $\tilde{M}_t$. With these changes, we can implement Algorithm 2 for the case with switching linear dynamics.

A.9 Experiments

A.9.1 One-Dimensional Case

We consider a one-dimensional ($m = p = k = 1$) reaching task, in which all noise sources are present: additive, control and signal-dependent, and internal. We show that, for non-zero internal noise, $\Omega_\eta > 0$, our GD and FPOMP algorithms (see also Appendix A.9.4) outperform the widely used algorithm of [1] (Fig. 6a). Parameters of the system are listed in Table 2 in Appendix A.9.5 (note that here we call $\sigma_\eta = \sqrt{\Omega_\eta}$).

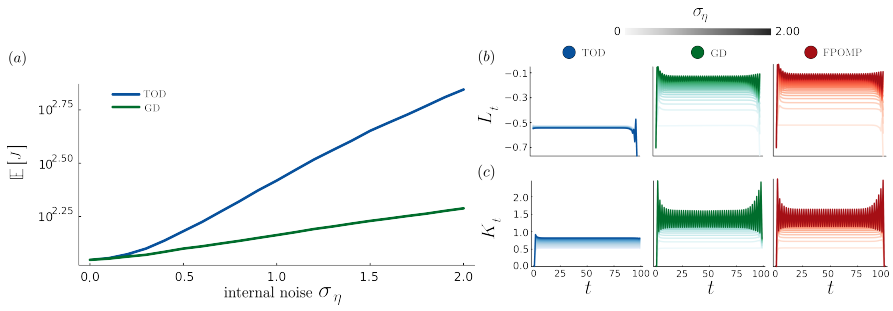

Figure 4: *Enhanced performance with internal noise*. **(a)** $\mathbb{E}[J]$, computed by averaging the quantity from Eq. 5 over $50k$ trials, as a function of σ_η, with error bars (SEM - not visible as too small), for TOD and GD algorithms. **(b-c)** Optimal control and filter gains (we also display the solutions of FPOMP algorithm to show the convergence to the same solution).

The observed enhanced optimality is due to the different modulation of $L_{1,\dots,T-1}$ and $K_{1,\dots,T-2}$ while varying σ_η (Fig. 4b-c). Importantly, our solution leads to control gains that decrease with an increase in internal noise, while the TOD solution does not show a strong dependence on the internal noise magnitude (Fig. 4b). For a further discussion, see the next Section.

A.9.2 One-Dimensional Case - The Validity of Orthogonality Principle in the Presence of Internal Noise

In the presence of internal noise the optimal solution no longer is the one that minimizes $\Omega_t = \mathbb{E}[\hat{x}_t\hat{x}_t^\intercal] - \mathbb{E}[x_t\hat{x}_t^\intercal]$ (Fig. 5a): the orthogonality principle, which would require $\Omega_t = 0, \forall t = 1, ..., T$, does not hold anymore for $\sigma_\eta > 0$. In contrast, the optimal strategy seems to favour lower values for $\Gamma_t = \mathbb{E}[x_t x_t^\intercal] - \mathbb{E}[x_t\hat{x}_t^\intercal]$, as shown in Fig. 5b. This allows the system to filter out the internal fluctuations that affect the estimation process. In such a way, those fluctuations correlate less with the dynamics of the latent state x. This mechanism results in a slightly (but significantly) larger absolute estimation error $|e_t| = \sqrt{\mathbb{E}[(x_t - \hat{x}_t)^2]} = \sqrt{\Omega_t + \Gamma_t}$, for the GD solutions (Fig. 5c), that seems to help in decorrelating the internal noise from the state evolution. This "decorrelation mechanism" is achieved through an intertwined modulation of control and filter gains. Such a relationship emerges even in the absence of multiplicative noise: as discussed in Appendix A.3, the presence of internal noise breaks the separation principle, making L_t and K_t interdependent.

In Appendix A.9.3, we further investigate this dependence through an eigenvector decomposition of the dynamics, showing also how our optimal solution leads to a good generalizability to other levels of internal noise. In Appendix A.9.4 we show that the FPOMP algorithm matches the solutions of the numerical GD, leading to the same performance.

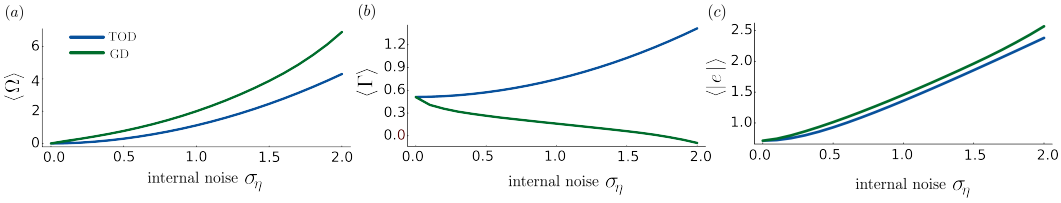

Figure 5: *Filtering out the internal fluctuations*. **(a)** Ω_t, averaged over time (we indicate the time average with $\langle\cdot\rangle$), as a function of σ_η for TOD and GD algorithms. **(b)** $\langle\Gamma\rangle$ as a function of σ_η. **(c)** $\langle e\rangle$ as a function of σ_η. The error bars ($\pm$ 1SEM from Monte Carlo simulations) are not visible as too small. The system parameters are the same as the 1D problem discussed in the previous Section.

A.9.3 One-Dimensional Case - Eigenvector Analysis and Adaptability of the Optimal Solution

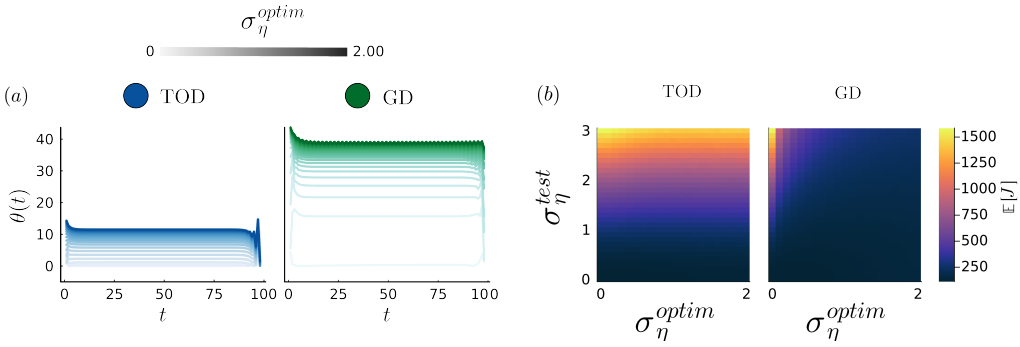

Figure 6: *An intertwined modulation of control and filter gains*. **(a)** Angles θ_t between the two eigenvectors of the matrix M_t, at different levels of internal noise σ_η^{optim}, for TOD and GD algorithms. **(b)** "Adaptability" of the two solutions; the solution found by the GD algorithm (right panel) generalizes better than the one by TOD (left panel) when optimized for a certain level of internal noise, σ_η^{optim}, and tested on another one, σ_η^{test}: for larger σ_η^{optim}, the generalization property improves thanks to due modulation of θ_t.

We consider the matrix M_t, defined in Eq. 22, that regulates the dynamics of the mean and covariance of x and $\hat{x}$, and therefore of Ω_t and Γ_t. The optimal solution arises from the adjustment of the angle θ between the two eigenvectors (we are considering here a one-dimensional case) of M_t: as

σ_η increases, θ also increases, thanks to the joined modulation of L_t and K_t with σ_η (Fig. 6a). Increasing θ allows the system to filter the internal fluctuations (see next paragraph for the details), and to better generalize to other levels of σ_η (Fig. 6b).

Eigenvector Decomposition - Mathematical Details In one dimension we can write the update equations for

$$\Gamma_t = \mathbb{E}[x_t^2] - \mathbb{E}[x_t \hat{x}_t] \tag{86}$$

$$\Omega_t = \mathbb{E}[\hat{x}_t^2] - \mathbb{E}[x_t \hat{x}_t] \tag{87}$$

as

$$\begin{pmatrix} \Gamma_{t+1} \\ \Omega_{t+1} \end{pmatrix} = \mathcal{M}_t \begin{pmatrix} \Gamma_t \\ \Omega_t \end{pmatrix} + \begin{pmatrix} \Omega_\xi + C^2 L_t^2 \mathbb{E}[\hat{x}_t^2] \\ \Omega_\eta + K_t^2 \Omega_\omega + K_t^2 D^2 \mathbb{E}[x_t^2] \end{pmatrix} \tag{88}$$

where

$$\mathcal{M}_t = (A - K_t H) \begin{pmatrix} A & -BL_t \\ -K_t H & A + BL_t - K_t H \end{pmatrix}. \tag{89}$$

The eigenvectors of $\mathcal{M}_t$ are given by

$$\vec{w}_1 = \begin{pmatrix} -1 \\ 1 \end{pmatrix} \tag{90}$$

$$\vec{w}_2 = \begin{pmatrix} BL_t/K_t H \\ 1 \end{pmatrix}. \tag{91}$$

Note that the angles θ_t between these two eigenvectors are the same as the angles between the eigenvectors of the matrix M_t. Indeed, the eigenvectors of M_t are given by

$$\vec{v}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \tag{92}$$

$$\vec{v}_2 = \begin{pmatrix} -BL_t/K_t H \\ 1 \end{pmatrix}. \tag{93}$$

A parity operation (along the x-axis) maps ones into the others, preserving the angles. Therefore, the results presented in Fig. 6 hold also for the matrix $\mathcal{M}_t$. By looking at the modulation of $\vec{w}_2$, while changing σ_η, in the plane $(\Gamma_t - \Omega_t)$, we can provide a heuristic interpretation of the different solutions found by TOD and GD algorithms. As σ_η increases, the angle between $\vec{w}_1$ and $\vec{w}_2$ gets larger for both algorithms. However, such a modulation is much more pronounced in GD solution, as we can see in Fig. 6a. Moreover, if only additive noise were considered, we would not observe any modulation of θ_t with σ_η in TOD solution (for a confirmation of this see Appendix A.9.7: without multiplicative noise, there is no modulation of the control gains with σ_η in TOD derivation).

As we can see in Fig. 7, the joined modulation of L_t and K_t causes $\vec{w}_2$ to get closer and closer to the y-axis in GD solution (green line). This configuration results in a more effective filtering of the internal fluctuations, decoupling them from the dynamics of the latent state. Indeed, these fluctuations are taking place on Ω_t (see Eq. 88). This result is in line with what we observed in the previous Sections regarding the relationship between internal noise and the separation principle: in order to filter the internal fluctuations affecting the estimation process, the optimal solution involves an intertwined optimization of control and filter gains, regardless of the presence of multiplicative noise. Moreover, it also aligns with the observed decrease of L_t with σ_η (see Figs. 4 and 2): lowering the control gain allows $\vec{w}_2$ to be closer to the y-axis. Thus, this consideration of eigenvectors can qualitatively explain the trends observed in Fig. 5 for $\langle \Gamma \rangle$ as a function of σ_η for TOD and GD solutions.

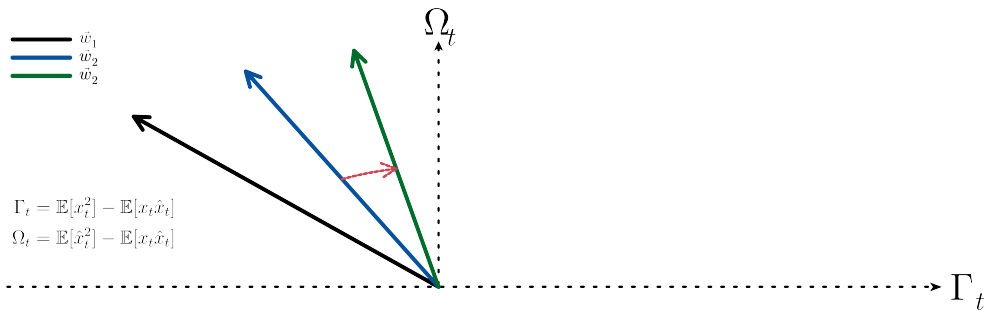

Figure 7: *Eigenvector decomposition of the dynamics*. We show here a qualitative representation of the eigenvectors of the matrix $\mathcal{M}_t$ in the plane (Γ_t, Ω_t). The black arrow represents the "shared" eigenvector $\vec{w}_1$, while the blue (green) arrow represents $\vec{w}_2$ for TOD (GD) solution. Note that, as can be seen in Fig. 4, the optimal L_t are negative, while the optimal K_t are positive.

A.9.4 One-Dimensional Case - FPOMP

We show that the FPOMP and GD algorithms yield the same performance in the one-dimensional problem introduced in Section 3, confirming their equivalence as discussed in Section 3.

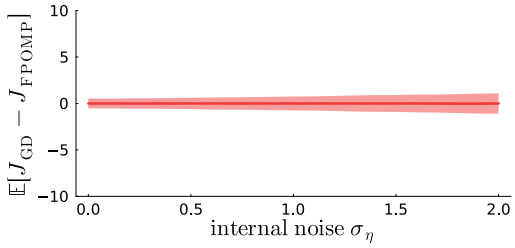

Figure 8: *Accumulated cost difference*. Difference of $\mathbb{E}[J]$ for GD and FPOMP solutions (computed by averaging the quantity from Eq. 5 over $50k$ trials), as a function of σ_η, with error bars (mean $\pm$ 1SEM from Monte Carlo simulations).

A.9.5 One-Dimensional Case - Parameters

We set $c = d = 1$.

Table 2: Parameters of the one-dimensional problem

Name	Description	value
A	Linear map for the system dynamics	1.0
B	Scaling of the control signal	1.0
C	Scaling matrix for control-dependent noise	0.5
D	Scaling for signal-dependent noise in the sensory feedback	0.5
H	Observation matrix	1
R_t	Control-dependent cost at each $t < T$	1
Q_t	Task-related cost at each time $t < T$	1
Q_T	Task-related cost at time $t = T$	20
T	time steps	100
$\mathbb{E}[\hat{x}_1] = \mathbb{E}[x_1]$	Initial condition for the mean state and state estimate	1.0
Σ_{x_1}	Initial covariance of the state	0.0
$\Sigma_{\hat{x}_1}$	Initial covariance of the state estimate	0.0
Ω_ξ	Covariance matrix of the additive Gaussian noise ξ_t	0.5^2
Ω_ω	Covariance matrix of the additive Gaussian noise ω_t	0.5^2
σ_η	Standard deviation of the additive internal Gaussian noise η_t	$\{0.0 : 0.1 : 2.0\}$

A.9.6 One-Dimensional Case - Suboptimality of TOD Solution at Fixed Filters

Here we briefly show that, even when there is no internal noise but convergence of the algorithm has not been achieved, the solution proposed in [1] does not provide the optimal control law. We demonstrate this in a one-dimensional example, using the same parameters shown in Appendix A.9 (but the result is valid in general), while only varying the scaling matrix for the multiplicative sensory noise D and keeping $\sigma_\eta = 0$. We fix the filter gains at the suboptimal constant value $K_{1,\dots,T-2} = A = 1.0$, and optimize the vector $L_{1,\dots,T-1}$ using TOD and GD algorithms. In Fig. 9 we can see how TOD control law leads to a higher expected accumulated cost $\mathbb{E}[J]$. This is because control optimization implies already the optimality of the estimator, by using the unbiasedness condition (that implies the orthogonality principle, as commented in Appendix A.3.1).

We note that a similar performance difference between the two algorithms would be observed if we fixed the control gains and optimized the filter gains.

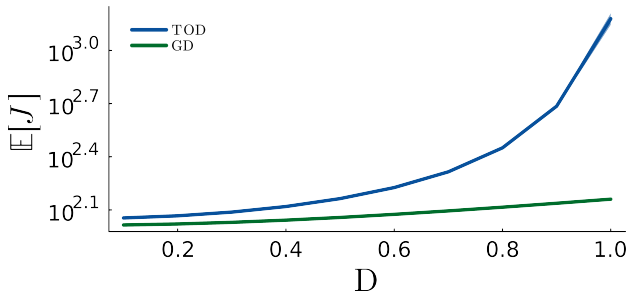

Figure 9: *Enhanced performance when optimizing control at fixed filter gains and zero internal noise*. We plot the expected accumulated cost $\mathbb{E}[J]$, computed by averaging the quantity from Eq. 5 over $50k$ trials, as a function of the scaling matrix D, with error bars (mean $\pm$ 1SEM from Monte Carlo simulations, error bars not visible as too small), for the two algorithms TOD and GD.

A.9.7 LQG with Internal Noise: A Simplified Problem to Validate FPOMP

We consider here the same problem as the one described in Section 3, but without any multiplicative sources of noise (see parameters in Appendix A.9.8, Table 4). This serves as a validation of the FPOMP algorithm derived in Appendix A.7.2. In Fig. 10a, we see how using the FPOMP algorithm to optimize the control gains L_t (at fixed filter gains K_t given by the TOD solution) leads to increasingly better performance (orange dashed line) as we increase the level of internal noise.

However, this solution does not correspond to the optimal one, given that the estimator is still optimized through the TOD algorithm. Indeed, when using the numeric GD to get the optimal L_t and K_t, we obtain a lower accumulated cost (green dashed line).

The qualitative features found for the sensorimotor task (Section 3, Fig. 2b) are confirmed here: control magnitude decreases as internal noise increases (Fig. 10b). We can also observe how TOD solution does not enforce any modulation of the control with respect to the internal noise level if only additive noise is considered. In contrast, in the FPOMP algorithm such a modulation takes place (Fig. 10b), leading to a lower accumulated cost (Fig. 10b).

An interesting feature of our algorithm is that, being completely "analytical", it could help improving the numerical solutions. Indeed, due to a possibly shallow landscape (vanishing gradient) in the parameters space (by parameters we mean now the coefficients of $L_{1,\dots,T-1}$ and $K_{1,\dots,T-2}$) and due to potentially limited computation time, the GD optimization could stop in the proximity of the global optimum, without fully reaching it. In Fig. 10c, we show how, taking the GD solution for the optimal filter gains and re-optimizing the control gains L_t using the FPOMP algorithm we get a small but significant performance boost, with small differences in the final L_t vector (we show, as an example the first component of this vector in Fig. 10d). This also confirms that our algorithm finds the optimal solutions. The extensions to the estimator optimization and to the multiplicative case are discussed in Appendix A.7.2.

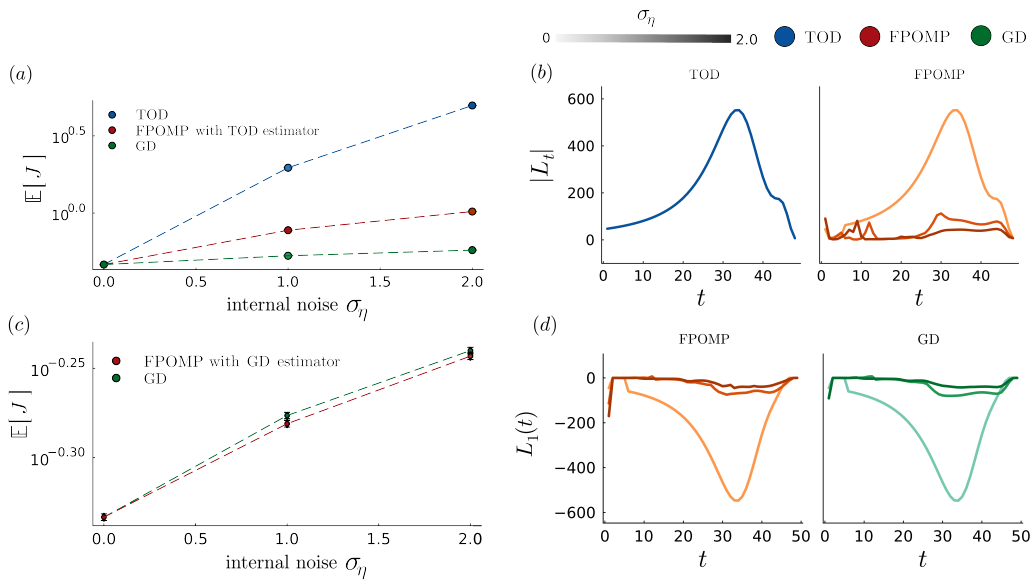

Figure 10: *Application of the FPOMP algorithm.* **(a)** Expected accumulated cost $\mathbb{E}[J]$, computed by averaging the quantity from Eq. 5 over $50k$ trials, as a function of the internal noise level σ_η, with error bars (mean $\pm$ 1SEM from Monte Carlo simulations, error bars not visible as too small) and for TOD, GD and FPOMP (at fixed filters given by TOD solution). **(b)** Magnitude of the control gain vector as a function of time for TOD and FPOMP (at fixed filters given by TOD solution) solutions. **(c)** Same as **(a)**, but comparing GD and FPOMP (now at fixed filters given by GD solution). **(d)** First component of the vector L_t for the solution given by GD and FPOMP (now at fixed filters given by GD solution).

A.9.8 Multi-Dimensional Case - Parameters

For the sensorimotor task described in Section 3 and the task for the LQG problem with additive internal noise (Appendix A.9.7), the discrete-time dynamics is the same as in [1],

$$p(t+\Delta t) = p(t) + \dot{p}(t)\Delta t \tag{94}$$
$$\dot{p}(t+\Delta t) = \dot{p}(t) + f(t)\Delta t/m \tag{95}$$
$$f(t+\Delta t) = f(t)(1-\Delta t/\tau_2) + g(t)\Delta t/\tau_2 \tag{96}$$
$$g(t+\Delta t) = g(t)(1-\Delta t/\tau_1) + u(t)(1+\sigma_\varepsilon \varepsilon_t)\Delta t/\tau_1 \tag{97}$$

We have therefore the following system parameters (with $c = d = 1$)

$$A = \begin{pmatrix} 1 & \Delta t & 0 & 0 \\ 0 & 1 & \Delta t/m & 0 \\ 0 & 0 & 1-\Delta t/\tau_2 & \Delta t/\tau_2 \\ 0 & 0 & 0 & 1-\Delta t/\tau_1 \end{pmatrix} \tag{98}$$

$$B = (0 \quad 0 \quad 0 \quad \Delta t/\tau_1)^\intercal \tag{99}$$

$$C = (0 \quad 0 \quad 0 \quad \sigma_\varepsilon \Delta t/\tau_1)^\intercal \tag{100}$$

$$H = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{101}$$

$$D = \begin{pmatrix} \sigma_\rho & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{102}$$

$$Q_{1,\dots,T-1} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{103}$$

$$Q_T = \vec{p}\vec{p}^\intercal + \vec{v}\vec{v}^\intercal + \vec{f}\vec{f}^\intercal \tag{104}$$

$$R_{1,\dots,T-1} = \frac{r}{T-1} \tag{105}$$

$$R_T = 0 \tag{106}$$

$$\vec{p} = (1 \quad 0 \quad 0 \quad 0) \tag{107}$$

$$\vec{v} = (0 \quad w_v \quad 0 \quad 0) \tag{108}$$

$$\vec{f} = (0 \quad 0 \quad w_v \quad 0) \tag{109}$$

$$\Omega_\xi = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \sigma_\xi^2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{110}$$

$$\Omega_\omega = \sigma_\omega^2 \tag{111}$$

$$\Omega_\eta = \begin{pmatrix} \sigma_\eta^2 & 0 & 0 & 0 \\ 0 & \sigma_{\eta_v}^2 & 0 & 0 \\ 0 & 0 & \sigma_{\eta_f}^2 & 0 \\ 0 & 0 & 0 & \sigma_{\eta_c}^2 \end{pmatrix} \tag{112}$$

with the initial conditions given by

$$\mathbb{E}[x_1] = (z \quad 0 \quad 0 \quad 0)^\intercal \tag{113}$$

$$\mathbb{E}[\hat{x}_1] = \mathbb{E}[x_1] \tag{114}$$

$$\Sigma_{x_1} = \begin{pmatrix} \sigma_z^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{115}$$

$$\Sigma_{\hat{x}_1} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{116}$$

For the sensorimotor task we have (std = standard deviation)

Table 3: Parameters of the sensorimotor task

Name	Description	value
Δt	time-step (s)	0.010
m	mass of the hand (Kg), modelled as a point mass	1
τ_1	time constant of the second order low pass filter	0.04
τ_2	time constant of the second order low pass filter	0.04
r	Control-dependent cost at each $t < T$	$1e^{-5}$
w_v	Task-related cost at time $t = T$ for the velocity	0.2
w_f	Task-related cost at time $t = T$ for the force	0.01
T	time steps	100
z	Target position	0.15
σ_z	Target position standard deviation	0.0
σ_ξ	std of the additive Gaussian noise ξ_t	0.0
σ_ω	std of the additive Gaussian noise ω_t	0.0
σ_ε	std of the control-dependent noise ε_t	0.5
σ_ρ	std of the signal-dependent noise ρ	0.5
σ_η	std of the additive internal noise η_t for the position estimate	$\{0.0, 0.005, 0.05, 0.5\}$
σ_{η_v}	std of the additive internal noise η_t acting on the velocity estimate	0
σ_{η_f}	std of the additive internal noise η_t for the force estimate	0
σ_{η_g}	std of the additive internal noise η_t for the estimate of g	0

while for the LQG problem with internal noise we consider

Table 4: Parameters of the LQG problem with internal noise

Name	Description	value
Δt	time-step (s)	0.010
m	mass of the hand (Kg), modelled as a point mass	1
τ_1	time constant of the second order low pass filter	0.04
τ_2	time constant of the second order low pass filter	0.04
r	Control-dependent cost at each $t < T$	$1e^{-5}$
w_v	Task-related cost at time $t = T$ for the velocity	0.2
w_f	Task-related cost at time $t = T$ for the force	0.01
T	time steps	50
z	Target position	0.15
σ_z	Target position standard deviation	0.0
σ_ξ	std of the additive Gaussian noise ξ_t	0.5
σ_ω	std of the additive Gaussian noise ω_t	0.5
σ_ε	std of the control-dependent noise ε_t	0.0
σ_ρ	std of the signal-dependent noise ρ_t	0.0
σ_η	std of the additive internal noise η_t for the position estimate	$\{0.0, 1.0, 2.0\}$
σ_{η_v}	std of the additive internal noise η_t acting on the velocity estimate	0
σ_{η_f}	std of the additive internal noise η_t for the force estimate	0
σ_{η_g}	std of the additive internal noise η_t for the estimate of g	0

Note that in both the multi-dimensional problems the initial condition for the state x_1 is the actual target position: in such a way the control signal u_t aims at minimizing the distance from $x_t = 0$.

A.9.9 Application to Higher-Dimensional Problems

We demonstrate how our algorithm scales to high-dimensional problems, building on the discussion in the final paragraph of Section 3. We implement a high-dimensional task to show the generalizability of the GD algorithm. The same results would apply to its analytical counterpart, the FPOMP algorithm, as discussed in Appendix A.5, and Appendices A.9.4, A.9.7. In this scenario, we set the dimensions of the state, control, and observation to $m = 10$, $p = 4$, and $k = 10$, respectively. Note that this significantly increases the dimensionality compared to the problem in Section 3 (for the multi-dimensional case).

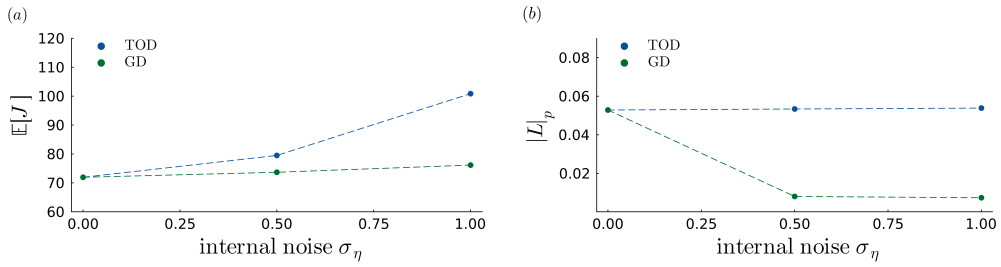

Figure 11: *High-dimensional task.* **(a)** Expected accumulated cost as a function of σ_η for TOD (blue dots) and GD (green dots) algorithms. We see that even in this high-dimensional task, GD solutions outperform the ones from [1]. To compute the expected cost, we used Algorithm 1 (but the results are confirmed by Monte Carlo simulations). **(b)** Pseudo-determinant of the control gains L (averaged over time), denoted as $|L|_p$ as a function of σ_η for TOD (blue dots) and GD (green dots) algorithms.

The system matrices A, B, and D are random matrices with elements drawn from a standard normal distribution (mean zero, standard deviation one), while C is defined as $C = \sigma_\varepsilon B$. The matrix H is the identity matrix, and the time horizon is set to $T = 10$. All elements of the state and state estimate vectors are initialized to one. We used $\sigma_\xi = \sigma_\omega = \sigma_\rho = \sigma_\varepsilon = 0.5$ and varied σ_η across values of 0.0,

0.5, and 1.0. The matrices defining the quadratic cost functions, Q and R, are identity matrices at each time step. All the findings from Section 3 are confirmed in this high-dimensional setting (Fig. 11). The GD algorithm continues to outperform the solutions in [1], with performance improving as internal noise increases, and the control gain magnitude decreases as internal fluctuations grow. In fact, as internal noise increases, the optimal strategy involves reducing control over the system. To quantify control magnitude, we compute the pseudo-determinant of $L_{1,\cdots,T-1}$ and average it over time. The pseudo-determinant, a generalization of the determinant for non-square matrices, provides a measure of the volume scaling induced by the control gains.