Neural optimal feedback control with local learning rules

FLATIRON INSTITUTE Center for Computational Neuroscience

Johannes Friedrich, Siavash Golkar, Shiva Farashahi, Alexander Genkin, Anirvan M. Sengupta, Dmitri B. Chklovskii

Optimal feedback control (OFC)

A prominent framework for how the brain plans and executes proper movements in the face of delayed and noisy sensory feedback [1].

- Integrates feedback and internal model predictions using Kalman Filter [2].
- Generates control actions that optimize behaviorally relevant criteria.
- **Open problem: a biologically plausible neural circuit for OFC.**

Our contribution: Bio-OFC addresses limitations of previous proposals.

	[3]	[4]	[5]	[6]	[7]	Bio-OFC
delayed sensory feedback	✗	✗	✗	✗	✗	✓
control included	✗	✓	✗	✗	✗	✓
noise covariance agnostic	✗	✓	✗	✗	✗	✓
online system identification	✗	✓	✗	✓	✓✗	✓
local learning rules	N/A	✓	N/A	✗	✓	✓
tractable latent size	✗	✓	✗	✓	✓	✓
absence of inner loop	✓	✗	✓	✓	✗	✓
single phase learning/execution	N/A	✗	N/A	✓	✓	✓

Problem formulation

- Estimate $\hat{\mathbf{x}}$ in order to design control $\mathbf{u}$ that minimizes expected cost J.

dynamics: $\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{B}\mathbf{u}_t + \mathbf{v}_t$ where $\mathbf{v}_t \sim \mathcal{N}(0;\mathbf{V})$

observation: $\mathbf{y}_t = \mathbf{C}\mathbf{x}_t + \mathbf{w}_t$ where $\mathbf{w}_t \sim \mathcal{N}(0;\mathbf{W})$

expected cost: $J = \mathbb{E}\left[\sum_{t=0}^{T} c(\mathbf{x}_t, \mathbf{u}_t)\right]$ with instantaneous cost c

control: $\mathbf{u}_t = k(\hat{\mathbf{x}}_t) = \arg\min J$

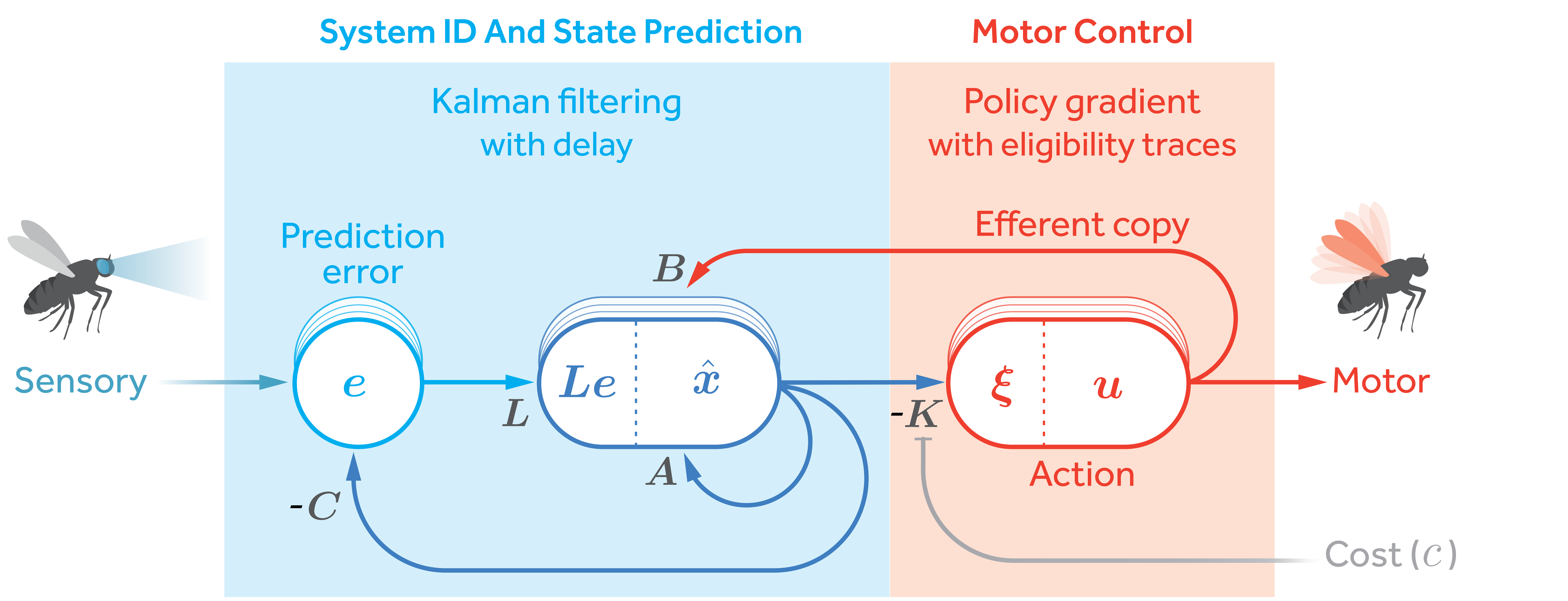
- System identification: Parameters $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ must be learned online.

Kalman estimation and control

- The Kalman filter is a recursive estimator that updates estimate $\hat{\mathbf{x}}_t$ as

$$\hat{\mathbf{x}}_{t+1} = \mathbf{A}\hat{\mathbf{x}}_t + \mathbf{B}\mathbf{u}_t + \mathbf{L}\mathbf{e}_t \quad \text{where} \quad \mathbf{e}_t := \mathbf{y}_t - \mathbf{C}\hat{\mathbf{x}}_t . \tag{1}$$

- It minimizes mean-squared error $\mathbb{E}[\mathbf{e}^\top\mathbf{e}]$. The Kalman gain matrix $\mathbf{L}$ optimally combines the internal model with the noisy observations.
- Delayed feedback: We estimate $\mathbf{x}_t$ before $\mathbf{y}_t$ has been observed.
- The control law simplifies if the cost J is quadratic: $\mathbf{u}_t = -\mathbf{K}\hat{\mathbf{x}}_t$ (2)



Learning Rules

$\Delta A_t \propto L e_t\, \hat{x}_{t-\tau}^\top$

$\Delta B_t \propto L e_t\, u_{t-\tau}^\top$

$\Delta L_t \propto L e_t\, e_{t-\tau}^\top$

$\Delta C_t \propto e_t\, \hat{x}_{t+1-\tau}^\top$

$Z_t = Z_{t-1} + \xi_t \hat{x}_t^\top$

$G_t = m G_{t-1} + c_t Z_t$

$\Delta K_t \propto -G_t$

Neural network representation for OFC

Circuit

- Eqs. (1-2) map onto a network with neural populations representing $\hat{\mathbf{x}}$, $\mathbf{e}$, and $\mathbf{u}$ that are connected by synaptic weights $\mathbf{A}$, $\mathbf{B}$, -$\mathbf{C}$, $\mathbf{L}$, and -$\mathbf{K}$.
- In case delay $\tau > 1$, recomputing the latent state throughout the delay period requires a biologically implausible circuit.

💡 Use direct state estimate updates based on delayed measurement, $\mathbf{e}_t := \mathbf{y}_{t+1-\tau} - \mathbf{C}\hat{\mathbf{x}}_{t+1-\tau}$. This affects performance only mildly.

Learning

- Gains $\mathbf{L}$ and $\mathbf{K}$ are usually each found by solving a Riccati equation, which requires matrix operations difficult to implement in biology.

System identification and Kalman gain

💡 Update $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{L}$ online using steps that minimize $\ell_t := \frac{1}{2}\mathbf{e}_t^\top\mathbf{e}_t$.

- The stochastic gradients are

$$-\nabla_{\mathbf{A}}\ell_t = \mathbf{C}^\top\mathbf{e}_t\hat{\mathbf{x}}_{t-\tau}^\top \quad -\nabla_{\mathbf{B}}\ell_t = \mathbf{C}^\top\mathbf{e}_t\mathbf{u}_{t-\tau}^\top \quad -\nabla_{\mathbf{L}}\ell_t = \mathbf{C}^\top\mathbf{e}_t\mathbf{e}_{t-\tau}^\top \quad -\nabla_{\mathbf{C}}\ell_t = \mathbf{e}_t\hat{\mathbf{x}}_{t+1-\tau}^\top$$

- A Hebbian rule for weights $\mathbf{C}$ 🙂, but non-local rules for $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{L}$ ☹️.

💡 Replace $\mathbf{C}^\top$ with $\mathbf{L}$: Corresponds to left-multiplication of the gradients with a positive definite matrix, thus updates still decrease the objective.

Control

💡 Learn $\mathbf{K}$ via policy gradient [8]

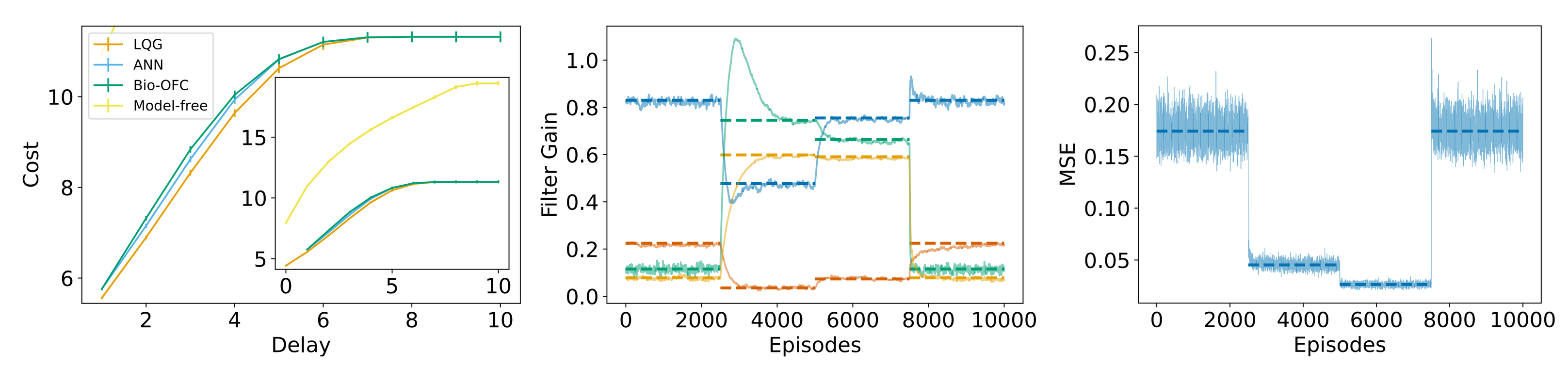
$$\nabla_{\mathbf{K}}J = \mathbb{E}\left[\sum_{t=0}^{T} c_t \sum_{s=0}^{t} \nabla_{\mathbf{K}} \log \pi_{\mathbf{K}}(\mathbf{u}_s \,|\, \hat{\mathbf{x}}_s)\right]$$

Experiments

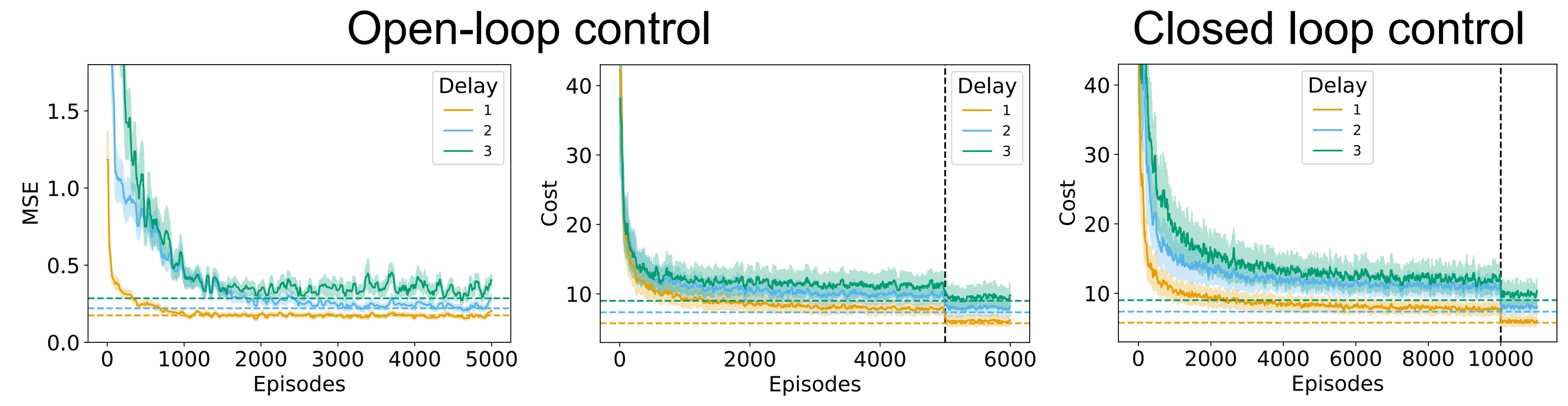
Discrete time double integrator

Bio-OFC converges to optimal values for given delay.

Performance close to LQG — Adaptation to changing noise statistics

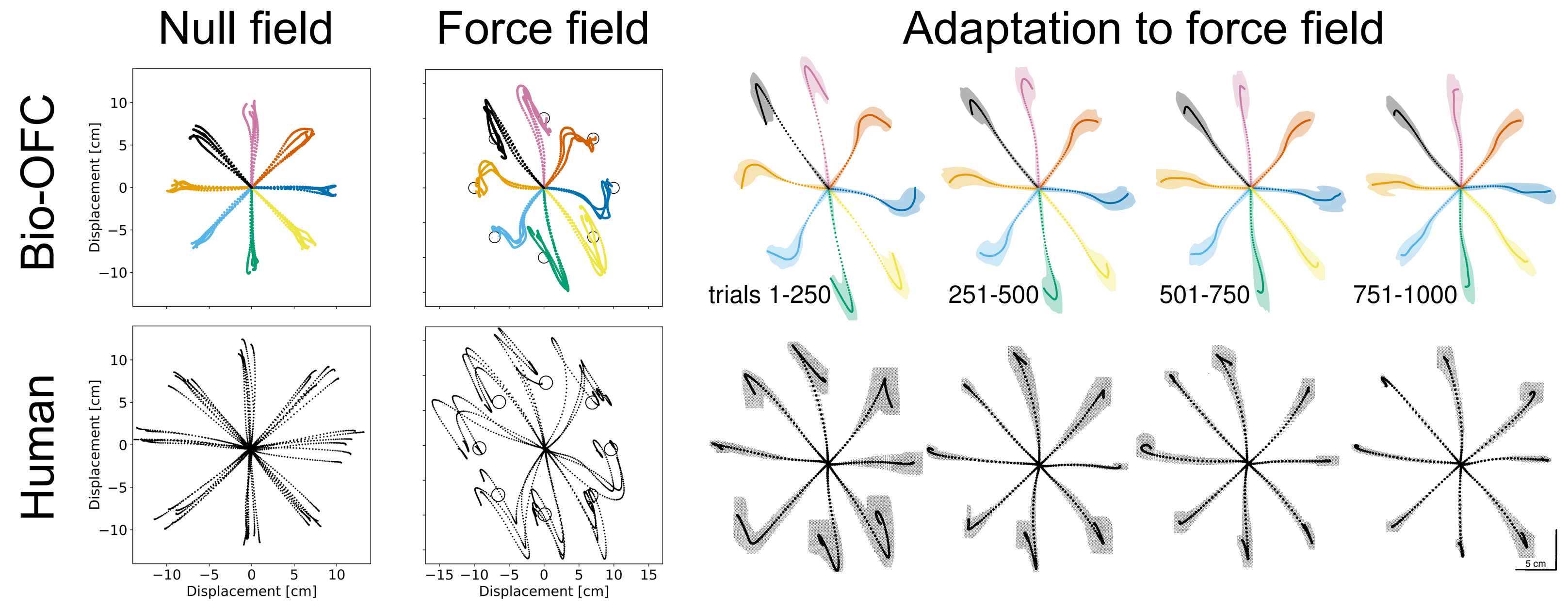


Open-loop control — Closed loop control



Adapting to a force field [9]

Bio-OFC captures the characteristics of human trajectories.



References

[1] E. Todorov and M.I. Jordan (2002). Nat Neurosci, 5(11):1226–1235
[2] R.E. Kalman (1960). J Basic Eng, 82(1):35–45
[3] S. Deneve et al. (2007). J Neurosci, 27(21):5744–5756
[4] R. Linsker (2008). Neural Netw, 21(9):1328–1343
[5] R. Wilson and L. Finkel (2009). In: NIPS 22:2062–2070
[6] A. Kutschireiter et al. (2017). Sci Rep, 7(1):1–13
[7] B. Millidge et al. (2021). arXiv:2102.10021
[8] R.J. Williams (1992). Mach Learn, 8:229-256
[9] R. Shadmehr and F.A. Mussa-Ivaldi (1994). J Neurosci, 14(5):3208-3224
[10] J. Friedrich et al. (2021). In: NeurIPS 34