Optimal Control with Adaptive Internal Dynamics Models

Djordje Mitrovic, Stefan Klanke, Sethu Vijayakumar – *School of Informatics, The University of Edinburgh, UK*

1 – Motivation

- Optimal feedback control (OFC) is a plausible movement generation strategy in goal reaching tasks for **biological systems** → attractive for **anthropomorphic manipulators**
- OFC yields minimal-cost trajectory with implicit resolution of kinematic and dynamic redundancies **plus** a feedback control law which corrects errors *only* if they adversely affect the task performance.
- Systems with non-linear dynamics and non-quadratic costs: OFC law can only be found locally and iteratively, e.g., using **iLQG** (Todorov & Li, 2005). However, **iLQG** relies on analytic form of system dynamics: → **often unknown, difficult to estimate, subject to changes**.
- Our approach: Combine **iLQG** framework with a **learned forward dynamics** model, based on Locally Weighted Projection Regression (**LWPR** – Vijayakumar, D'Souza & Schaal, 2005) → **iLQG–LD**
- Learned model is **adaptive**: Can compensate for complex dynamic perturbations in an online fashion
- Learned model is **efficient**: Derivatives are easy to compute (**iLQG** involves linearization)

2 – Background: OFC and iLQG

Notation we use here:

$\mathbf{x}(t)$	state of a plant (joint angles $\mathbf{q}$ and velocities $\dot{\mathbf{q}}$)
$\mathbf{u}(t)$	control signal applied at time t (torques)
$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u})$	forward dynamics (this is what we **learn**)

Problem statement:
Given initial state $\mathbf{x}_0$ at time $t = 0$, seek optimal control sequence $\mathbf{u}(t)$ such that "final" state $\mathbf{x}(T) = \mathbf{x}^*$.

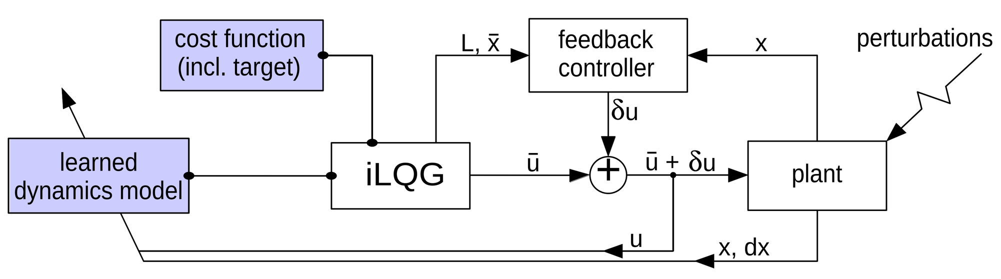
Cost function:
Final cost $h(\mathbf{x}(T))$ plus accumulated cost $c(t, \mathbf{x}, \mathbf{u})$ of sending a control signal $\mathbf{u}$ at time t in state $\mathbf{x}$ → "Error" in final state plus used "energy"

Weighted cost for time-discrete system, target in joint space:

$$v = w_p \|\mathbf{q}_K - \mathbf{q}_{tar}\|^2 + w_v \|\dot{\mathbf{q}}_K\|^2 + w_e \sum_{k=0}^{K} \|\mathbf{u}_k\|^2 \Delta t$$

Basic idea of iLQG:
Simulate initial control sequence, linearize dynamics $\mathbf{f}(\cdot)$ around resulting trajectory, solve local **LQG** problem for *deviations* $\delta\mathbf{u}_k$ and $\delta\mathbf{x}_k$ analytically. Update control sequence and repeat procedure until convergence.

3 – Incorporate learned dynamics

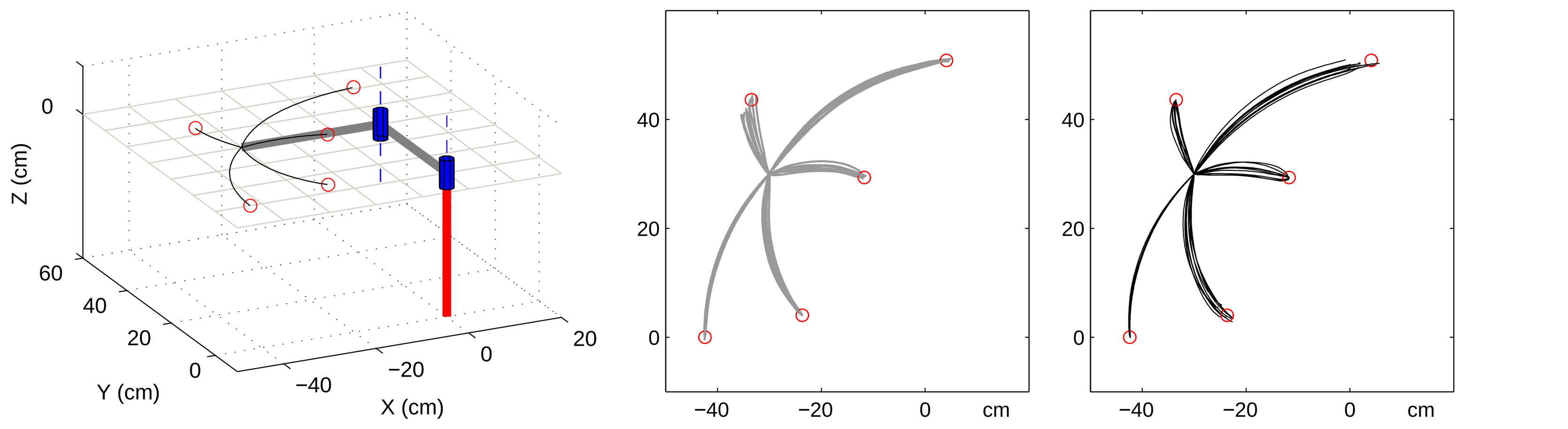


Learned model (**LWPR**) is **used** in simulated trials, and **trained online** during final trials on the real plant.

- Learned model can be pre-trained using motor babbling or a coarse analytic model
- **LWPR** is fully localised algorithm, local models learn independently → **incremental** training **without interference** problems
- **LWPR** features built-in dimensionality reduction by using Partial Least Squares within the local models
- Range of validity of local models ("receptive fields") can be automatically adjusted during training
- Calculating derivatives of learned model is **much faster** than using, e.g., finite differences of a model based on Newton-Euler recursions → **performance gain** especially for large number of DOF

4 – Basic 2-DOF example

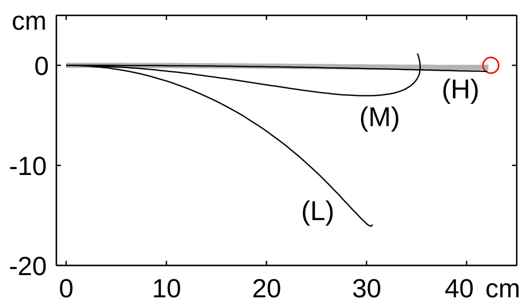
First investigation: Reaching performance evaluated on a simple 2-DOF planar arm, simulated using the MATLAB Robotics Toolbox.



Left: Simulated arm and reaching targets. Middle: Trajectories of **iLQG** (analytic). Right: **iLQG–LD** (learned).

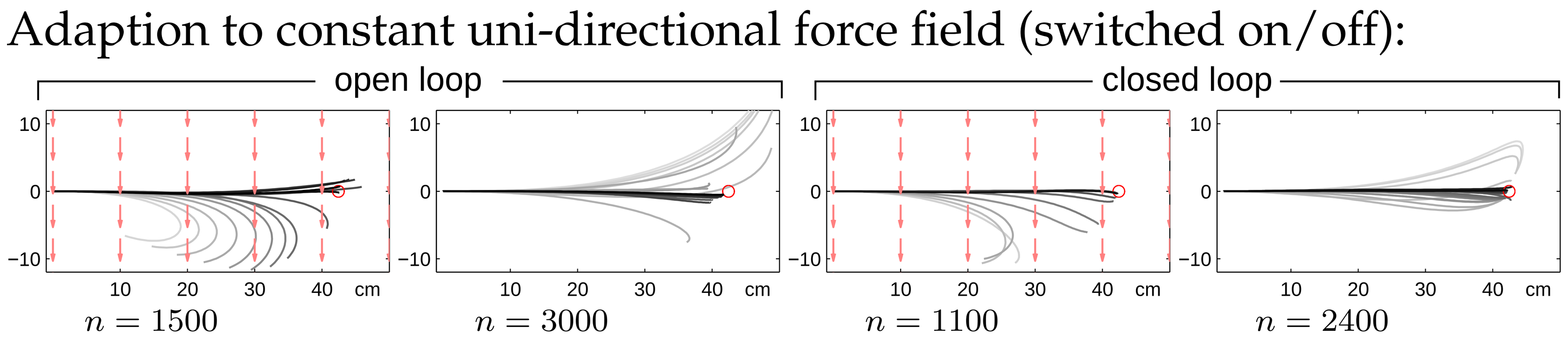
Naturally, reaching performance depends on quality of learned model:

iLQG–LD	(L)	(M)	(H)	iLQG
Train. points	111	146	276	–
Prediction error (nMSE)	0.80	0.50	0.001	–
Iterations	19	17	5	4
Cost	2777	1810	192	192
Eucl. target distance (cm)	19.50	7.20	0.40	0.01

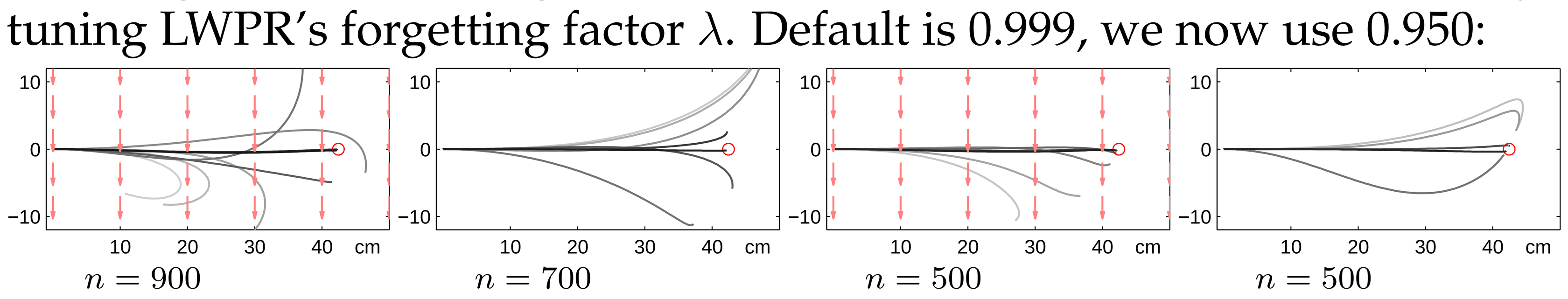


5 – Adapting to nonstationary dynamics

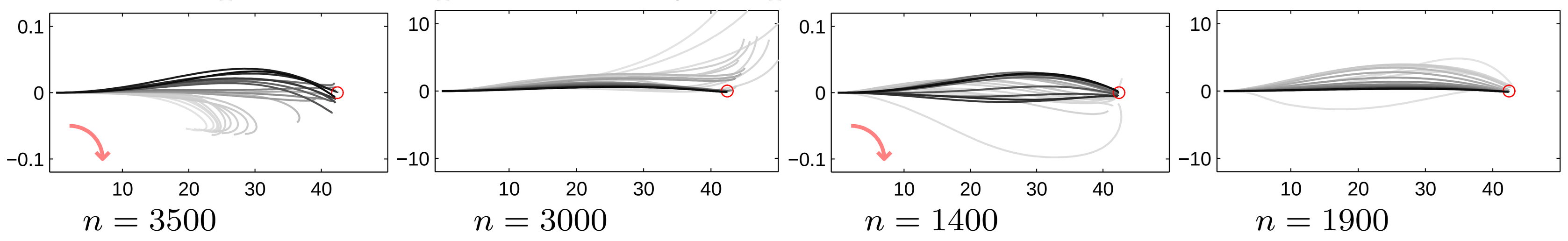
Simulate systematic perturbances by virtual force fields: Analytic models cannot account for this, but the learned model can.
Adaption to constant uni-directional force field (switched on/off):



Learning and un-learning the influence force field can be accelerated by tuning LWPR's forgetting factor λ. Default is 0.999, we now use 0.950:

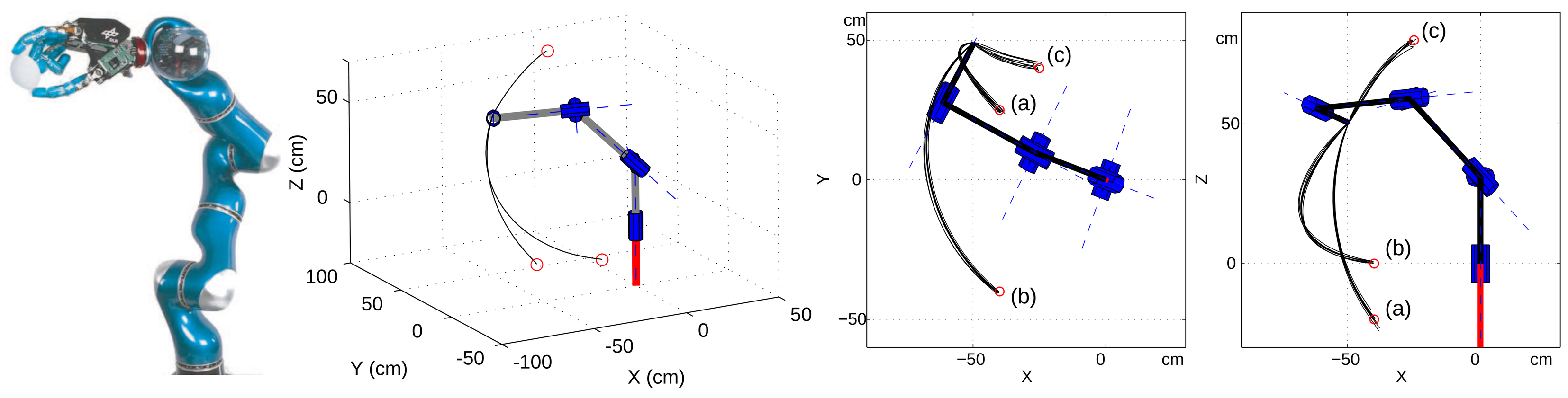


More complex example: Velocity-dependent force field.



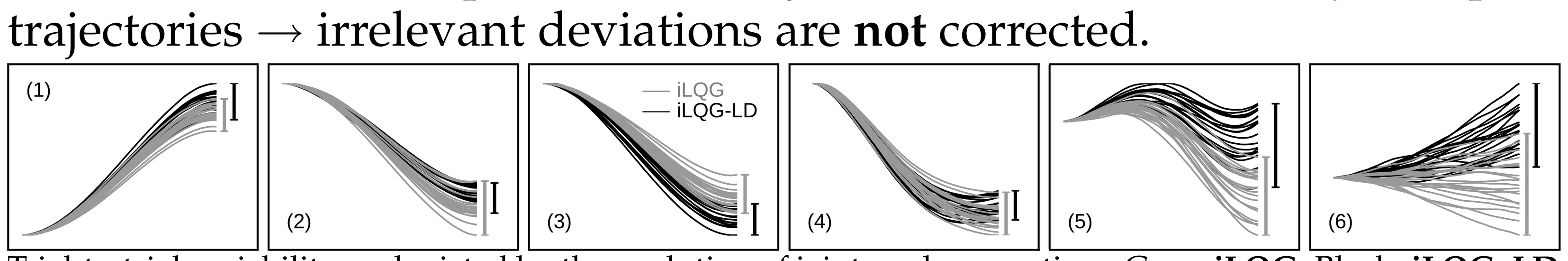
6 – Scaling iLQG–LD: 6-DOF

Faithful model of DLR robot arm (w/o last joint): Reaching 3 targets



Left to right: Real DLR arm, toolbox simulation and targets, resulting trajectories in xy−view and xz-view.

Modified cost function only includes end-effector **position** (through forward kinematics) → **iLQG** resolves redundancy implicitly.
Simulated control-dependent noise yields lots of variance in joint-space trajectories → irrelevant deviations are **not** corrected.



Trial-to-trial variability as depicted by the evolution of joint angles over time. Grey: **iLQG**. Black: **iLQG–LD**.

Compare cost and accuracy: **iLQG–LD** (left) and **iLQG** (right)

target	running cost	pos. error (cm)	running cost	pos. error (cm)
(a)	**18.32** ± 0.55	**1.92** ± 1.03	**18.50** ± 0.13	**2.63** ± 1.63
(b)	**18.65** ± 1.61	**0.53** ± 0.20	**18.77** ± 0.25	**1.32** ± 0.69
(c)	**12.18** ± 0.03	**2.00** ± 1.02	**12.92** ± 0.04	**1.75** ± 1.30

References

Todorov & Li. *A generalized iterative LQG method for locally-optimal feedback control of constrained nonlinear stochastic systems.* Proceedings of the American Control Conference, 2005.

Vijayakumar, D'Souza & Schaal. *Incremental online learning in high dimensions.* Neural Computation 17, pp. 2602–2634, 2005.

Acknowledgments: This research is supported partially by the EC project SENSOPAC.

Contact Author: Djordje Mitrovic (d.mitrovic@ed.ac.uk) for further details on this work.