By *including the output model* in the problem formulation, the controller can *systematically reduce uncertainty* where it hurts the objective.

universität freiburg

A dual-control effect preserving formulation for nonlinear output-feedback stochastic model predictive control with constraints

Florian Messerer, Katrin Baumgärtner, Moritz Diehl
University of Freiburg

Contribution

Dual control can be formulated as an output-feedback stochastic optimal control problem, which is however intractable. We derive a tractable approximation in which the dual control effect is preserved.

(Intractable) output-feedback stochastic OCP

▶ Stochastic nonlinear system with policy $\pi(\cdot)$

$$
\begin{aligned}
x_0^\pi(\xi) &= p_0(\xi_0),\ u_0^\pi(\xi) = \bar{u}_0, && \\
x_{k+1}^\pi(\xi) &= f_k(x_k^\pi(\xi), u_k^\pi(\xi), w_k), && k = 0, \dots, N-1,\\
y_k^\pi(\xi) &= g_k(x_k^\pi(\xi), v_k), && k = 1, \dots, N-1,\\
I_k^\pi(\xi) &= (I_{k-1}^\pi(\xi), u_{k-1}^\pi(\xi), y_k^\pi(\xi)), && k = 1, \dots, N-1,\\
u_k^\pi(\xi) &= \pi_k(I_k^\pi(\xi)), && k = 1, \dots, N-1.
\end{aligned}
$$

▶ Optimal policy π^* defined by

cost: $J^\pi(\xi) = \sum_{k=0}^{N-1} l_k(x_k^\pi(\xi), u_k^\pi(\xi)) + l_N(x_N^\pi(\xi))$,

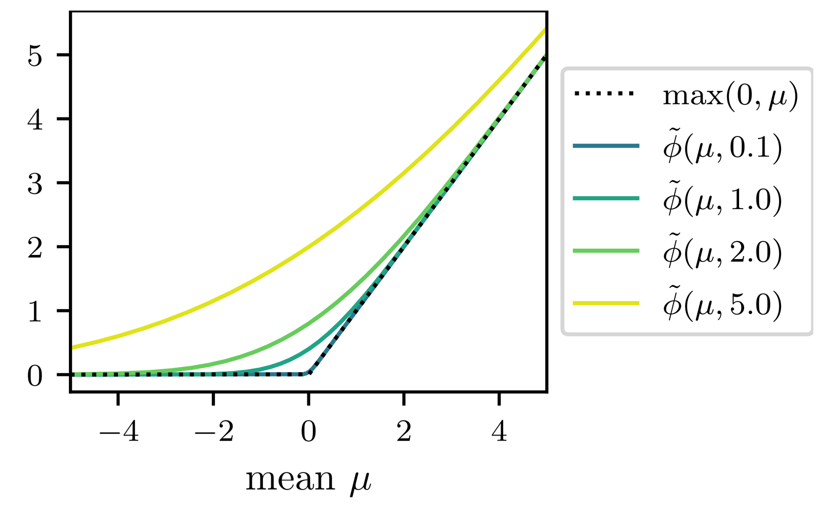
constraints: $h_k(x_k^\pi(\xi), u_k^\pi(\xi)) \leq 0,\ h_N(x_N^\pi(\xi)) \leq 0$,

$$\min_{\pi(\cdot)} \quad \mathbb{E}_\xi \left\{ J^\pi(\xi) + \sum_{i=0}^{n_h} \rho_i \max(0, h_i^\pi(\xi)) \right\}$$

Constraint penalization

The expectation operator has a smoothing effect on the constraint violation penalty.

$$\tilde{\phi}(\mu, \sigma) := \mathbb{E}_{\eta \sim \mathcal{N}(\mu, \sigma^2)} \{\max(0, \eta)\}$$



Uncertain linear system around nonlinear nominal trajectory

$$
\begin{aligned}
\bar{x}_{k+1} &= f_k(\bar{x}_k, \bar{u}_k, 0),\ \bar{y}_k = g_k(\bar{x}_k, 0)\\
x_{k+1} - \bar{x}_{k+1} &\approx A_k(x_k - \bar{x}_k) + B_k(u_k - \bar{u}_k) + \Gamma_{k+1} w_k,\\
y_k - \bar{y}_k &\approx C_k(x_k - \bar{x}_k) + D_k v_k,
\end{aligned}
$$

▶ Summarize information in Kalman estimate $x_k \sim \mathcal{N}(\hat{x}_k, \hat{P}_k)$

▶ linear feedback κ based on the state *estimate*: $u_k = \bar{u}_k + K_k(\hat{x}_k - \bar{x}_k)$

$\Rightarrow$ linear system $\tilde{x}_k^\kappa(\xi) = \begin{bmatrix} x_k^\kappa(\xi) - \bar{x}_k \\ \hat{x}_k^\kappa(\xi) - x_k^\kappa(\xi) \end{bmatrix}$ with *predictable* covariance Σ_k

Dual control

▶ The choice of trajectory affects the information gain (*dual control effect*)

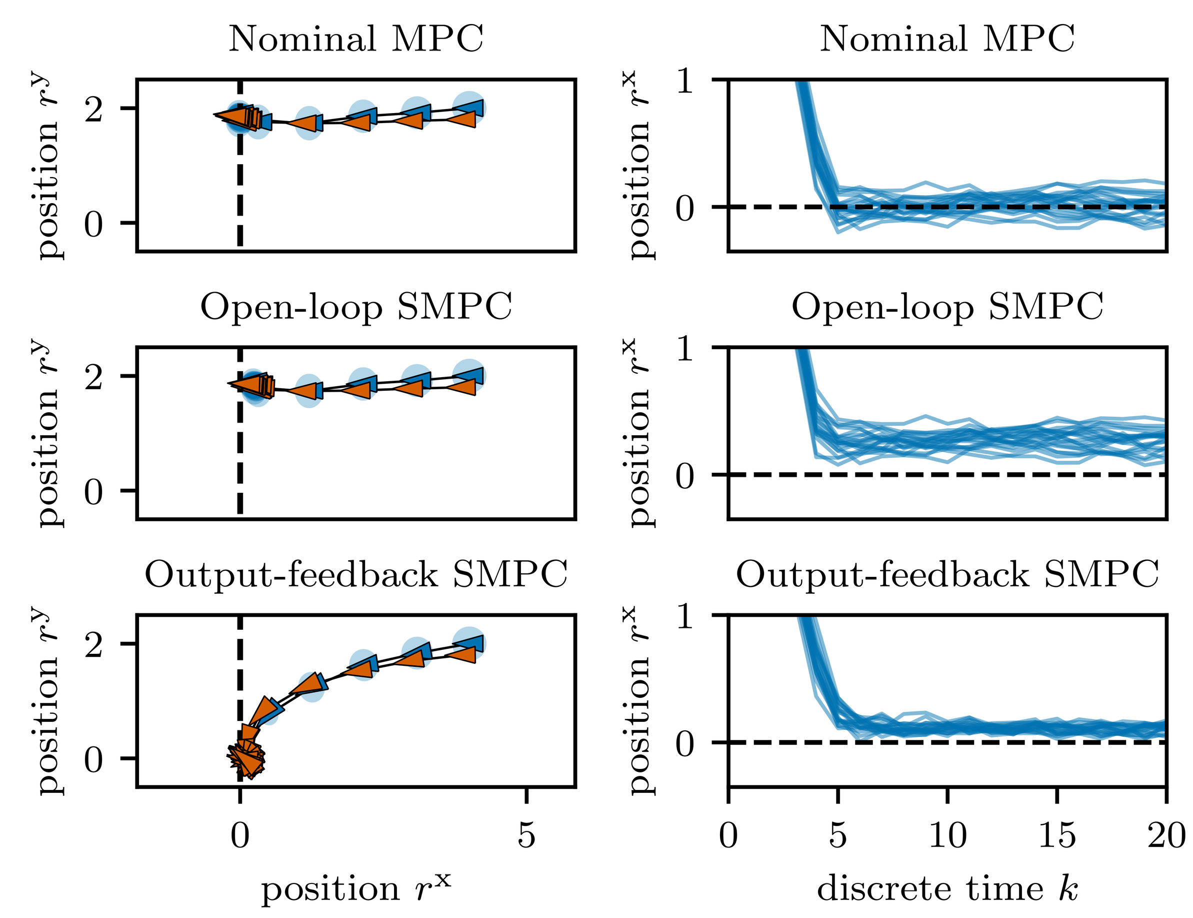
▶ A performance metric implictly defines the value of information

▶ The induced explore-exploit trade-off can be captured by: (a) modelling how information is obtained, (b) optimizing over information feedback

Illustrative example

Objective: Go as far left as possible without constraint violation.

Output model: Less measurement noise when close to bottom of frame.



Tractable OCP

$$
\begin{aligned}
\min_{\bar{x}, \bar{u}, \beta, \Sigma, K} \quad & \tilde{J}(\bar{x}, \bar{u}, \Sigma, K) + \tilde{\Phi}(h(\bar{x}, \bar{u}), \beta) + \varepsilon_\beta \|\beta\|_2^2 + r(\Sigma, K)\\
\text{s.t.} \quad & \bar{x}_0 = \hat{x}_0,\ \Sigma_0 = \hat{\Sigma}_0(\hat{P}_0),\\
& \bar{x}_{k+1} = f_k(\bar{x}_k, \bar{u}_k, 0) && k = 0, \dots, N-1,\\
& \Sigma_{k+1} = \psi_k(\bar{x}_k, \bar{u}_k, \Sigma_k, K_k), && k = 0, \dots, N-1,\\
& 0 \geq h_k^u(\bar{u}_k), && k = 0, \dots, N-1,\\
& \beta \geq 0,\\
& \beta_k^i \geq H_k^i(\bar{x}_k, \bar{u}_k, \Sigma_k, K_k), && k = 0, \dots, N-1,\\
& && i = 1, \dots, n_{h_k},\\
& \beta_N^i \geq H_N^i(\bar{x}_N, \Sigma_N), && i = 1, \dots, n_{h_N},
\end{aligned}
$$

where H_k^i corresponds to the linearized variance in constraint direction, and $r(\cdot, \cdot)$ is a regularization term.