Dual control effect preserving stochastic MPC

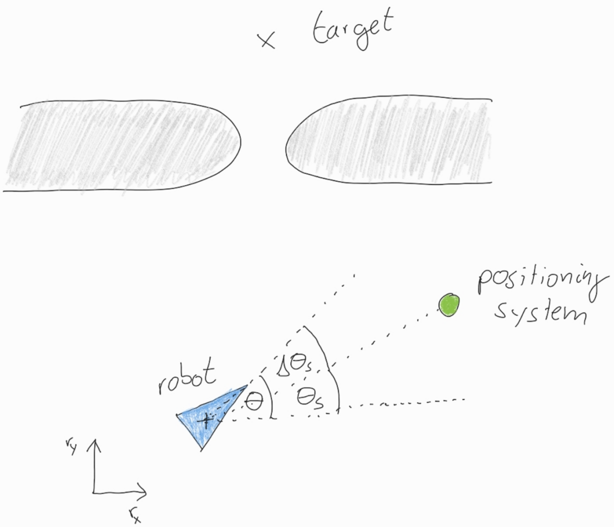
Florian Messerer Katrin Baumgärtner Moritz Diehl

Systems Control and Optimization Laboratory (syscop), University of Freiburg

Workshop on Robust Model Predictive Control

Feb 27, 2023

Scenario



target
positioning system
robot
$\Delta\theta_s$
θ_s
θ
r_y
r_x

Dual control

- In general, the choice of trajectory affects the information gain (*dual control effect*)
- For uncertain systems there are two types of *conflicting* control goals
  - Exploring: Gathering information to reduce uncertainty
  - Exploiting: Bringing the system into an actually desired state (greedily)
- How do we find the optimal trade-off?
- How do we **define** the optimal trade-off?

Key ideas

- We have a performance-based objective, which is what we actually care about.
- Information is only relevant in as much as it advances this objective.
- We can model how we will obtain information.
- We can model how we will use this information.

$\rightarrow$ Optimal Control Problem (OCP)

Overview

Part I: Defining the ideal optimal control problem

- Output-feedback stochastic OCP.
- This is the problem we would like to solve.
- It will turn out that this problem is intractable.

Part II: Deriving a tractable approximation

- Using linearization and normal distributions
- Preserving the dual control effect (implicit dual control)

Partially observed stochastic nonlinear system

$$\begin{aligned} x_0 &= p_0(\xi_0), & \\ x_{k+1} &= f_k(x_k, u_k, w_k), & k = 0, \ldots, N-1, \\ y_{k+1} &= g_{k+1}(x_{k+1}, v_{k+1}), & k = 0, \ldots, N, \end{aligned}$$

- initial state uncertainty ξ_0, process noise $w = (w_0, \ldots, w_{N-1})$ and output noise $v = (v_1, \ldots, v_{N-1})$, distributed as $\xi = (\xi_0, w, v) \sim \mathcal{N}(0, \mathbb{I})$.
- Pick control u_k at time k based on information vector

$$I_1 = (u_0, y_1), \quad I_k = (I_{k-1}, u_{k-1}, y_k), \quad k = 2, \ldots, N$$

- Aim: find policy $\pi = (\bar{u}_0, \pi_1(\cdot), \ldots \pi_{N-1}(\cdot))$
  - Defining how we react to future information

$$u_0 = \bar{u}_0, \quad u_k = \pi_k(I_k), \; k = 1, \ldots, N-1$$

OCP components

- Denote by $x_k^\pi(\xi), u_k^\pi(\xi)$ the simulation of the system under $\pi(\cdot)$.

- Incurred cost $J^\pi(\xi) = \sum_{k=0}^{N-1} l_k(x_k^\pi(\xi), u_k^\pi(\xi)) + l_N(x_N^\pi(\xi))$
  - Stochastic $\rightarrow$ not a well-defined objective function
  - Choose deterministic criterion, e.g., expected value

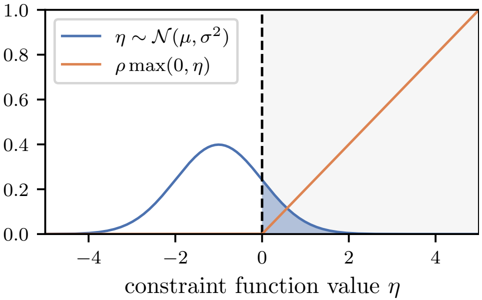
- Constraints $h_k(x_k^\pi(\xi), u_k^\pi(\xi)) \leq 0, \quad k = 0, \ldots, N-1, \quad h_N(x_N^\pi(\xi)) \leq 0,$
  - In general impossible to enforce for all values of ξ (unbounded support)
  - Option A: Chance constraints

$$\mathcal{P}\left\{\begin{array}{ll} h_k(x_k^\pi(\xi), u_k^\pi(\xi)) \leq 0, & k = 0, \ldots, N-1, \\ \& \; h_N(x_N^\pi(\xi)) \leq 0, & \end{array}\right\} \geq 1 - \varepsilon$$

  - **Option B: Penalize constraint violation and add to objective**

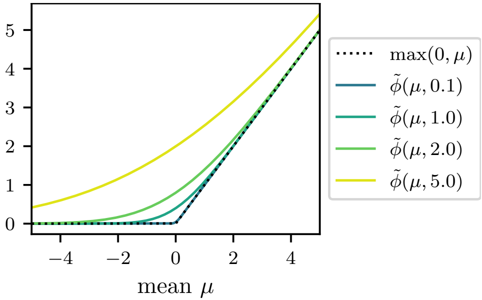
Constraint treatment

- Constraint $h(x) \leq 0$
- Consider $\eta = h(x) \sim \mathcal{N}(\mu, \sigma^2)$
- *Exact* penalty $\phi(\eta)$ on $\eta > 0$



- Expected penalty

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



The ideal dual OCP

Output-feedback stochastic nonlinear OCP

$$\min_{\pi(\cdot)} \quad \mathbb{E}_{\xi\sim\mathcal{N}(0,\mathbb{I})}\left\{ J^\pi(\xi) + \sum_{i=0}^{n_h} \phi_i^h(h_i^\pi(\xi)) \right\}$$

with $\phi_i^h(\eta) := \rho_i \max(0,\eta)$ and $h^\pi(\xi) = (h_0(x_0^\pi(\xi), u_0^\pi(\xi)), \dots, h_N(x_N^\pi(\xi)))$

- Models how future information is obtained, and how to react to it

  $\Rightarrow$ Perfectly encodes explore-exploit trade-off as induced by objective
- Expectation over nonlinear transformation of ξ
- Optimization over general policies in infinite-dimensional function space

Overview

Part I: Defining the ideal optimal control problem

- Output-feedback stochastic OCP.
- This is the problem we would like to solve.
- It will turn out that this problem is intractable.

Part II: Deriving a tractable approximation

- Using linearization and normal distributions
- Preserving the dual control effect (implicit dual control)

Implicit vs explicit dual control

The ideal OCP is intractable and needs to be approximated.

Two types of approximation

Implicit dual control

- Dual control effect is (qualitatively) preserved
- Explore-exploit tradeoff remains implicitly encoded

Explicit dual control

- Dual control effect is lost
- Heuristics to encourage uncertainty reduction

Linearize with respect to uncertainty

- Nominal trajectory with open-loop controls $\bar{u} = (\bar{u}_0, \dots, \bar{u}_{N-1})$.

$$
\begin{aligned}
\bar{x}_0 &= p_0(0), \\
\bar{x}_{k+1} &= f_k(\bar{x}_k, \bar{u}_k, 0), && k = 0, \dots, N-1, \\
\bar{y}_k &= g_k(\bar{x}_k, 0), && k = 1, \dots, N-1,
\end{aligned}
$$

- Linearize in stochastic variables, at nominal trajectory

$$
\begin{aligned}
x_0 - \bar{x}_0 &\approx P_0^{\mathrm{r}} \xi_0, \\
x_{k+1} - \bar{x}_{k+1} &\approx A_k(\bar{x}_k, \bar{u}_k)(x_k - \bar{x}_k) + B_k(\bar{x}_k, \bar{u}_k)(u_k - \bar{u}_k) + \Gamma_{k+1}(\bar{x}_k, \bar{u}_k) w_k, \\
y_k - \bar{y}_k &\approx C_k(\bar{x}_k)(x_k - \bar{x}_k) + D_k(\bar{x}_k) v_k,
\end{aligned}
$$

Kalman filter and linear feedback law

- We now have a *linear* uncertain system around a *nonlinear* nominal trajectory
  - Linear system matrices A_k, B_k, Γ_k, C_k, D_k depend on nominal trajectory $(\bar{x}, \bar{u})$
  - Explicit dependency dropped for notational simplicity
- With respect to the *linearized* system...
  - ...the information vectors I_k can be (w.r.t. to the linearized system) perfectly summarized by Kalman estimate $\hat{x}_k$, $\hat{P}_k$ with Kalman gain $\hat{K}_k$:

$$\hat{x}_0 - x_0 = -P_0^{\mathrm{r}}\xi_0,$$
$$\hat{x}_{k+1} - x_{k+1} = (\mathbb{I} - \hat{K}_{k+1}C_{k+1})A_k(\hat{x}_k - x_k) + (\hat{K}_{k+1}C_{k+1} - \mathbb{I})\Gamma_k w_k + \hat{K}_{k+1}D_{k+1}v_{k+1}.$$

  - ... the future KF means $\hat{x}_k$ are uncertain, but the covariances $\hat{P}_k$ are perfectly predictable
  - ... the covariances are independent of the linear system trajectory: no dual control effect in a linear system.
  - (but we retain dual control effect since $(\bar{x}_k, \bar{u}_k)$ affects the linear system matrices)
- Restrict control policies to linear feedback $\kappa_k(\cdot)$ based on the current state estimate:

$$u_k = \kappa_k(\hat{x}_k) = \bar{u}_k + K_k(\hat{x}_k - \bar{x}_k).$$

Linear Uncertainty dynamics

Augmented linear system

$$\tilde{x}_0^\kappa := \begin{bmatrix} \hat{P}_0^{\mathrm{r}} \\ -\hat{P}_0^{\mathrm{r}} \end{bmatrix} \xi_0, \quad \tilde{x}_k^\kappa := \begin{bmatrix} x_k^\kappa - \bar{x}_k \\ \hat{x}_k^\kappa - x_k^\kappa \end{bmatrix}, \quad \tilde{w}_k := \begin{bmatrix} w_k \\ v_{k+1} \end{bmatrix},$$

$$\underbrace{\tilde{A}_k^\kappa := \begin{bmatrix} \bar{A}_k + \bar{B}_k K_k & \bar{B}_k K_k \\ 0 & (\mathbb{I} - \hat{K}_{k+1} \bar{C}_{k+1}) \bar{A}_k \end{bmatrix}, \quad \tilde{\Gamma}_k := \begin{bmatrix} \bar{\Gamma}_k & 0 \\ (\hat{K}_{k+1} \bar{C}_{k+1} - \mathbb{I}) \bar{\Gamma}_k & \hat{K}_{k+1} \bar{D}_{k+1} \end{bmatrix}}$$

$$\tilde{x}_{k+1}^\kappa(\xi) = \tilde{A}_k^\kappa \tilde{x}_k^\kappa(\xi) + \tilde{\Gamma}_k \tilde{w}_k$$

Mean $\mathbb{E}_\xi\{\tilde{x}_k(\xi)\} = 0$ and covariance $\mathrm{Cov}_\xi\{\tilde{x}_k(\xi)\} =: \Sigma_k,\ k = 0, \dots, N$

$$\Sigma_0 = \underbrace{\begin{bmatrix} \hat{P}_0 & -\hat{P}_0 \\ -\hat{P}_0 & \hat{P}_0 \end{bmatrix}}_{=:\hat{\Sigma}_0(\hat{P}_0)}, \quad \Sigma_k = \begin{bmatrix} P_k & \breve{P}_k^\top \\ \breve{P}_k & \hat{P}_k \end{bmatrix}, \quad \Sigma_{k+1} = \underbrace{\tilde{A}_k^\kappa \Sigma_k \tilde{A}_k^{\kappa\top} + \tilde{\Gamma}_k \tilde{\Gamma}_k^\top}_{=:\psi_k(\bar{x}_k, \bar{u}_k, \Sigma_k, K_k)},$$

Resulting dynamics approximation

- We have approximated the evolution of the original system by
  1. Propagating a nonlinear nominal trajectory $(\bar{x}, \bar{u})$
  2. Linearizing w.r.t. to uncertainty, using linear feedback based on a Kalman filter estimate.

Approximated uncertainty dynamics

$$\bar{x}_0 = \hat{x}_0, \qquad \bar{x}_{k+1} = f_k(\bar{x}_k, \bar{u}_k, 0),$$
$$\Sigma_0 = \hat{\Sigma}_0(\hat{P}_0), \quad \Sigma_{k+1} = \psi_k(\bar{x}_k, \bar{u}_k, \Sigma_k, K_k), \quad k = 0, \ldots, N-1.$$

- Approximation retains a mechanism through which
  1. the choice of nominal controls $\bar{u}$ influences the estimation uncertainty
  2. the estimation uncertainty is fed-back into the "real" system...
  3. ...and thus impacts the predictive uncertainty of the true state
  4. (which will hurt the objective)

  $\Rightarrow$ Dual control effect is preserved

Expectation of cost

- State and control distribution approximated as

$$\underbrace{\begin{bmatrix} x_k \\ u_k \end{bmatrix}}_{=:z_k} \sim \mathcal{N}\left(\underbrace{\begin{bmatrix} \bar{x}_k \\ \bar{u}_k \end{bmatrix}}_{=:\bar{z}_k}, \underbrace{\begin{bmatrix} \mathbb{I} & 0 \\ K_k & K_k \end{bmatrix} \Sigma_k \begin{bmatrix} \mathbb{I} & 0 \\ K_k & K_k \end{bmatrix}^\top}_{=:\tilde{\Sigma}_k(\Sigma_k, K_k)} \right), \; k = 0, \dots, N-1, \quad x_N \sim \mathcal{N}(\bar{x}_N, P_N)$$

- We consider quadratic cost functions l_k with Hessian $B_k := \nabla^2 l_k(\cdot)$

$$\mathbb{E}_{z_k \sim \mathcal{N}(\bar{z}_k, \tilde{\Sigma}_k)} \{l_k(z_k)\} = l_k(\bar{z}_k) + \tfrac{1}{2} \operatorname{Tr}(B_k \tilde{\Sigma}_k),$$

- Resulting cost approximation (error due to linearized propagation)

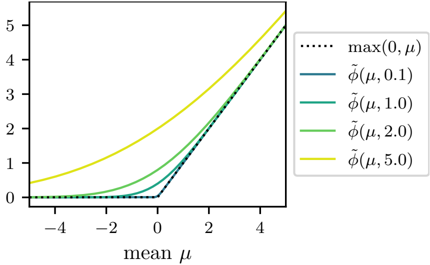
$$\mathbb{E}_{\xi \sim \mathcal{N}(0, \mathbb{I})} \{J^\pi(\xi)\} \approx \underbrace{\sum_{k=0}^{N-1} l_k(\bar{x}_k, \bar{u}_k) + \tfrac{1}{2} \operatorname{Tr}\left(B_k \tilde{\Sigma}_k(\Sigma_k, K_k)\right) + l_N(\bar{x}_N) + \tfrac{1}{2} \operatorname{Tr}\left(B_N \tilde{\Sigma}_N(\Sigma_N)\right)}_{=:\tilde{J}(\bar{x}, \bar{u}, \Sigma, K)}.$$

Constraints

- Additional linearization of the constraint functions

$$\mathbb{E}_{z_k \sim \mathcal{N}(\bar{z}_k, \tilde{\Sigma}_k)}\{\phi_i(h_k^i(z_k))\}$$
$$\approx \mathbb{E}_{z_k \sim \mathcal{N}(\bar{z}_k, \tilde{\Sigma}_k)}\{\phi_i(h_k^i(\bar{z}_k) + \nabla h_k^i(\bar{z}_k)(z - z_k))\}$$
$$= \mathbb{E}_{\eta \sim \mathcal{N}(\bar{h}_k^i, \beta_k^i)}\{\phi_i(\eta)\},$$

  - nominal constraint value $\bar{h}_k^i := h_k^i(\bar{z}_k)$
  - variance orthogonal to constraint boundary $\beta_k^i := \nabla h_k^i(\bar{z}_k)^\top \tilde{\Sigma}_k \nabla h_k^i(\bar{z}_k)$



- Compute expected constraint penalty analytically

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

with $p_{\mathcal{N}}$ resp. $P_{\mathcal{N}}$ the PDF resp. CDF of standard normal distribution

Resulting approximated OCP

Implicit dual OCP

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

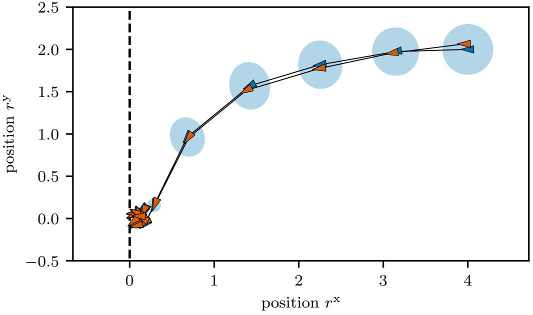
with trajectory cost $\tilde{J}(\cdot)$, constraint penalty $\tilde{\Phi}(\cdot)$, regularization $r(\cdot)$, variances in constraint direction β resp. $H(\cdot)$

Example problem

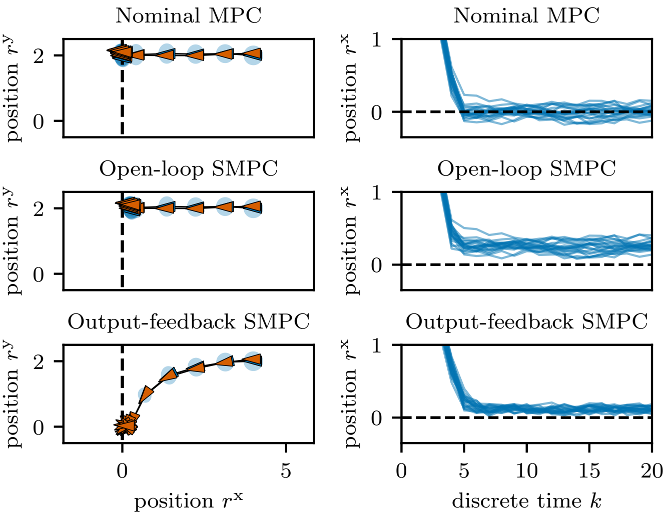
$$x = \begin{bmatrix} r^{\mathrm{x}} \\ r^{\mathrm{y}} \\ \theta \end{bmatrix}, \quad \dot{x} = \begin{bmatrix} v\cos\theta \\ v\sin\theta \\ \omega \end{bmatrix} + w, \qquad u = \begin{bmatrix} v \\ \omega \end{bmatrix}.$$

$$y_k = x_k + \sigma_{\mathrm{y}}(x) v_k, \quad \sigma_{\mathrm{y}}(x) := 1 + 10\left(\sqrt{r_k^{\mathrm{y}} + \varepsilon^2} - \varepsilon\right)$$

$$\begin{aligned}
\min_{x,\,u} \quad & \sum_{k=0}^{N-1} r_k^x + \varepsilon_u \|u\|_2^2 + r_N^x \\
\text{s.t.} \quad & x_0 = \bar{x}_0, \\
& x_{k+1} = f(x_k, u_k, 0), \quad && k = 0, \dots, N-1, \\
& u_{\min} \le u_k \le u_{\max}, \quad && k = 0, \dots, N-1, \\
& 0 \le r_k^x, \quad && k = 1, \dots, N.
\end{aligned}$$



Comparison to non-dual methods



Summary / Conclusion

- The dual effect can be captured by
  - Including output-model in OCP
  - Optimizing over output-feedback policies
- The ideal dual OCP implicitly encodes the explore-exploit trade-off ...
- ... but is intractable
- We used linearization to derive a tractable OCP ...
- ... while preserving the dual control effect.

F. Messerer, K. Baumgärtner, M. Diehl. *A dual-control effect preserving formulation for nonlinear output-feedback stochastic model predictive control with constraints.* IEEE Control Systems Letters, 2023