

Model Predictive Control and Reinforcement Learning

– Lecture 08: Differentiable Programming –

Moritz Diehl

University of Freiburg

Slides jointly developed with Armin Nurkanović, Jonathan Frey, Leonard Fichtner

School on Model Predictive Control and Reinforcement Learning
Freiburg, 7-16 October 2026

universität freiburg



NTNU

Norwegian University of
Science and Technology

Sensitivity Computation

Generalized Tangential Predictors via SQP

Generalized Tangential Predictors and Derivatives via SBQP

Motivation

- ▶ Embedding optimization solvers in neural networks requires solution sensitivities
- ▶ Learning-enhanced MPC schemes, MPC-RL

$$\begin{aligned} w^{\text{sol}}(p) &:= \arg \min_{w \in \mathbb{R}^{n_w}} && F(w; p) \\ &\text{subject to} && G(w; p) = 0, \\ &&& H(w; p) \leq 0 \end{aligned}$$

Wanted: $\frac{\partial w^{\text{sol}}}{\partial p}(p)$

Motivation

- ▶ Embedding optimization solvers in neural networks requires solution sensitivities
- ▶ Learning-enhanced MPC schemes, MPC-RL

$$\begin{aligned} w^{\text{sol}}(p) := \arg \min_{w \in \mathbb{R}^{n_w}} \quad & F(w; p) \\ \text{subject to} \quad & G(w; p) = 0, \\ & H(w; p) \leq 0 \end{aligned}$$

Wanted: $\frac{\partial w^{\text{sol}}}{\partial p}(p)$

Need to introduce implicit functions

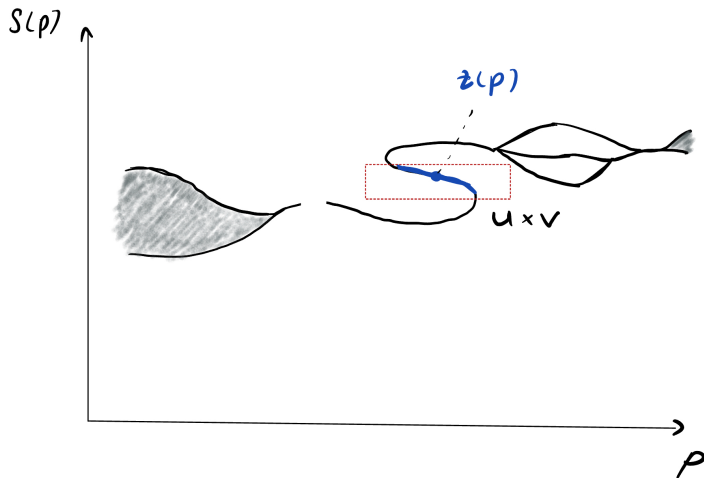
- ▶ Given a parameter p , we are interested in finding solution $z(p)$:

$$R(z, p) = 0,$$

with $R : \mathbb{R}^{n_z} \times \mathbb{R}^{n_p} \rightarrow \mathbb{R}^{n_z}$. The resulting function $z(p)$ is defined *implicitly*.

- ▶ Central questions are:
 - ▶ When does a solution exist?
 - ▶ Is the solution (locally) unique? Is it the function $z(p)$ single or multi-valued?
 - ▶ Is $z(p)$ differentiable and how do we compute its derivatives?
 - ▶ How to efficiently compute approximations $\hat{z}(p) \approx z(p)$, and how can the numerical error be quantified?

Illustration of a solution map



Example solution map $S(p)$ which can be single-valued (*unique minimizer*), set-valued (*nonunique local minimizers*), set-valued with isolated arcs (*multiple strict local minima*) empty.

First-order approximations of solution maps

- ▶ Often, we may solve an equation $R(z, p^*) = 0$ and obtain $z(p^*)$. Then our problem data may slightly change to a new p , but it may be computationally expensive to evaluate the new $z(p)$.
- ▶ Instead, we may compute a first-order Taylor approximation:

$$z(p) \approx z(p^*) + \frac{dz(p^*)}{dp}(p - p^*).$$

- ▶ The evaluation $z(p^*)$ is available from the last solve, the derivative $\nabla_p z(p^*) = \frac{dz(p^*)}{dp}^\top$ from the implicit function theorem.
- ▶ Recall that the main idea of Newton's method was to use a sequence of linearizations.
 1. use Newton's method to compute evaluations of $z(p^*)$,
 2. show how to directly compute a linear approximation of $z(p)$.

Theorem (Implicit function theorem)

Let $R : \mathbb{R}^{n_z} \times \mathbb{R}^{n_p} \rightarrow \mathbb{R}^{n_z}$ be a C^1 function and $(z^*, p^*) \in \mathbb{R}^{n_z} \times \mathbb{R}^{n_p}$ such that $R(z^*, p^*) = 0$.

There exists a neighborhood $U \subset \mathbb{R}^{n_p}$ of p^* and a differentiable function $z(p)$ such that $R(z(p), p) = 0$ for all $p \in U$, and $z(p)$ is the unique solution in V , a neighborhood of z^* , **if and only if** the partial Jacobian $\frac{\partial R}{\partial z}(z^*, p^*)$ is invertible.

In addition, $z(\cdot)$ is C^1 differentiable on U , with

$$\frac{dz(p)}{dp} = - \left(\frac{\partial R}{\partial z}(z, p) \right)^{-1} \frac{\partial R}{\partial p}(z, p) \text{ for every } p \in U. \quad (1)$$

Theorem (Implicit function theorem)

Let $R : \mathbb{R}^{n_z} \times \mathbb{R}^{n_p} \rightarrow \mathbb{R}^{n_z}$ be a C^1 function and $(z^*, p^*) \in \mathbb{R}^{n_z} \times \mathbb{R}^{n_p}$ such that $R(z^*, p^*) = 0$.

There exists a neighborhood $U \subset \mathbb{R}^{n_p}$ of p^* and a differentiable function $z(p)$ such that $R(z(p), p) = 0$ for all $p \in U$, and $z(p)$ is the unique solution in V , a neighborhood of z^* , **if and only if** the partial Jacobian $\frac{\partial R}{\partial z}(z^*, p^*)$ is invertible.

In addition, $z(\cdot)$ is C^1 differentiable on U , with

$$\frac{dz(p)}{dp} = -\left(\frac{\partial R}{\partial z}(z, p)\right)^{-1} \frac{\partial R}{\partial p}(z, p) \text{ for every } p \in U. \quad (1)$$

Proposition (Lipschitz from differentiability)

If $f(\cdot)$ is continuously differentiable on an open set O and C is a compact convex subset of O , then $f(\cdot)$ is Lipschitz continuous relative to C with constant $L = \max_{z \in C} \|\nabla f(z)\|$.



Disclaimer: we will emphasize the dependence on p , otherwise everything identical to before!

KKT conditions for parametric NLPs (p-NLP)

Disclaimer: we will emphasize the dependence on p , otherwise everything identical to before!

The KKT conditions of (p-NLP) read as:

$$\begin{aligned}\nabla_w F(w, p) - \sum_{i=1}^{n_G} \lambda_i \nabla_w G_i(w, p) - \sum_{i=1}^{n_H} \mu_i \nabla_w H_i(w, p) &= 0, \\ G(w, p) &= 0, \\ H(w, p) &\geq 0, \quad \mu \geq 0, \\ H_i(w, p) \mu_i &= 0, i = 1, \dots, n_H.\end{aligned}$$

Under a CQ they are necessary for optimality. Under convexity they are also sufficient. A primal-dual KKT point is the vector

$$z(p) = (w(p), \lambda(p), \mu(p)) \in \mathbb{R}^{n_w} \times \mathbb{R}^{n_G} \times \mathbb{R}^{n_H}.$$

Given a point w and parameter p , the set of multipliers (λ, μ) that satisfy the KKT conditions are denoted by $\mathcal{M}(w, p)$. Recall that, if LICQ holds, at a stationary point the set $\mathcal{M}(w, p)$ is a singleton.

- The active set

$$\mathcal{A}(w, p) = \{i \in \{1, \dots, n_H\} \mid H_i(w, p) = 0\}$$

- Strongly active set

$$\mathcal{A}^+(w, \mu, p) = \{i \in \mathcal{A}(w, p) \mid \mu_i > 0\}$$

- Weakly active set

$$\mathcal{A}^0(w, \mu, p) = \{i \in \mathcal{A}(w, p) \mid \mu_i = 0\}$$

- Relations:

$$\mathcal{A}^0(w, \mu, p) \cup \mathcal{A}^+(w, \mu, p) = \mathcal{A}(w, p),$$

$$\mathcal{A}^0(w, \mu, p) \cap \mathcal{A}^+(w, \mu, p) = \emptyset.$$

- We say that **strict complementarity slackness (SCS)** holds if $\mathcal{A}^0(w, \mu, p) = \emptyset$. Most difficulties arise because usually small changes of p result in changes of $\mathcal{A}^0(w, \mu, p)$.

Definition (Critical cone)

Let (w, λ, μ) be a KKT point.

The critical cone $\mathcal{C}(w, \mu, p)$ is the following set:

$$\mathcal{C}(w, \mu, p) := \left\{ d \in \mathbb{R}^{n_w} \mid \nabla_w G(w, p)^\top d = 0, \nabla_w H_i(w, p)^\top d \begin{cases} = 0, & i \in \mathcal{A}^+(w, \mu, p), \\ \geq 0, & i \in \mathcal{A}^0(w, \mu, p) \end{cases} \right\}.$$

For more strict conditions, often a larger set than the critical cone is used:

$$\mathcal{D}(w, \mu, p) := \{ d \in \mathbb{R}^n \mid \nabla G(w)^\top d = 0, \nabla H_i(w)^\top d = 0, i \in \mathcal{A}^+(w, \mu) \}$$



Disclaimer: we will emphasize the dependence on p , otherwise everything identical to before!

Disclaimer: we will emphasize the dependence on p , otherwise everything identical to before!

Definition (Second-order sufficient condition (SOSC))

SOSC holds at w if there exists $(\lambda, \mu) \in \mathcal{M}(w, p)$ such that

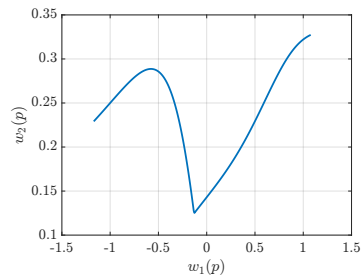
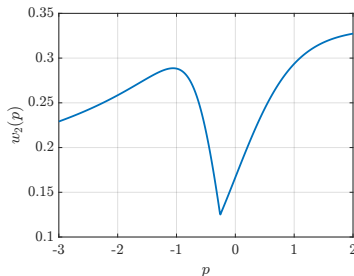
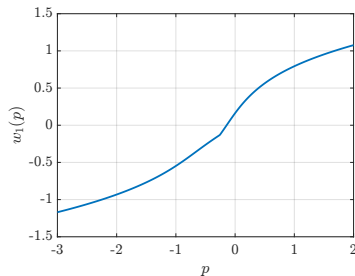
$$d^\top \nabla_{ww}^2 \mathcal{L}(w, \lambda, \mu, p) d > 0 \quad \forall d \in \mathcal{C}(w, \mu, p) \setminus \{0\}. \quad (2)$$

Definition (Strong second-order sufficient condition (SSOSC))

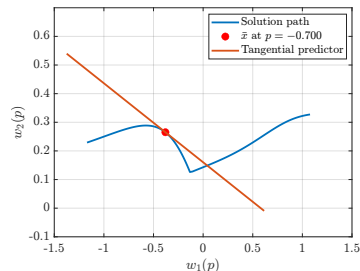
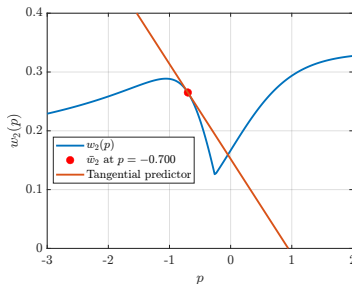
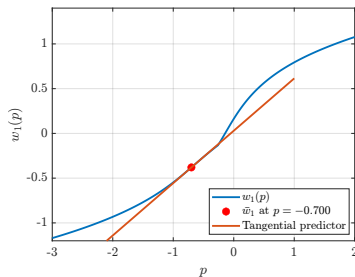
SSOSC holds at w if there exists $(\lambda, \mu) \in \mathcal{M}(w, p)$ such that

$$d^\top \nabla_{ww}^2 \mathcal{L}(w, \lambda, \mu, p) d > 0 \quad \forall d \in \mathcal{D}(w, \mu, p) \setminus \{0\}. \quad (3)$$

Parametric solution $z(p)$ - solution manifold



The active set is fixed and SCS holds (we are far from the kinks)



- ▶ Note that under SCS $\mu_{\mathcal{A}} > 0$, and for inactive constraints it holds $\mu_{\bar{\mathcal{A}}} = 0$, where $\bar{\mathcal{A}}(w, p) := \{1, \dots, n_H\} \setminus \mathcal{A}(w, p)$.
- ▶ With a slight abuse of notation we redefine z and write $z = (w, \lambda, \mu_{\mathcal{A}})$, and $\mu_{\bar{\mathcal{A}}} = 0$, $H_{\bar{\mathcal{A}}}(w, p) > 0$.
- ▶ If the active set is known, then the KKT conditions are a smooth nonlinear root-finding problem:

$$R(z, p) = \begin{pmatrix} \nabla_w \mathcal{L}(w, \lambda, \mu_{\mathcal{A}}, p) \\ -G(w, p) \\ -H_{\mathcal{A}}(w, p) \end{pmatrix} = 0.$$

- ▶ Recall the Implicit Function Theorem (IFT):
 $\frac{dz(p)}{dp} = -\left(\frac{\partial R}{\partial z}(z, p)\right)^{-1} \frac{\partial R}{\partial p}(z, p)$ for every $p \in U$ where U is a neighborhood.
- ▶ Can we directly apply the IFT to the KKT system?

IFT for a KKT system with fixed active set and SCS

1) $H_{\mathcal{A}}(w, p)$ should not change. The active set is fixed for all $p \in U \implies \mu_i H_i(w, p) = 0$, by differentiating we obtain

$$\frac{\partial \mu_i}{\partial p} H_i(w, p) + \mu_i \left(\frac{\partial H_i(w, p)}{\partial w} \frac{\partial w}{\partial p} + \frac{\partial H_i}{\partial p} \right) = 0, \quad i = 1, \dots, n_H.$$

- ▶ If $i \in \mathcal{A}(w, p)$, then $\mu_i > 0$, thus $\frac{\partial H_i(w, p)}{\partial w} \frac{\partial w}{\partial p} + \frac{\partial H_i}{\partial p} = 0$, i.e., the constraints stay active.
- ▶ If $i \in \bar{\mathcal{A}}(w, p)$, then $H_i(w, p) > 0$, $\mu_i = 0$, thus $\frac{\partial \mu_i}{\partial p} = 0$, i.e., the multiplier stays zero.

2) The Jacobian must be invertible.

$$\begin{aligned} M := \frac{\partial R(z, p)}{\partial z} &= \begin{bmatrix} \nabla_{ww}^2 \mathcal{L} & -\nabla_w G(w, p) & -\nabla_w H_{\mathcal{A}}(w, p) \\ -\nabla_w G(w, p)^\top & 0 & 0 \\ -\nabla_w H_{\mathcal{A}}(w, p)^\top & 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} \nabla_{ww}^2 \mathcal{L} & -\nabla_w \bar{G}_{\mathcal{A}}(w, p) \\ -\nabla_w \bar{G}_{\mathcal{A}}(w, p)^\top & 0 \end{bmatrix}, \\ r := \frac{\partial R(z, p)}{\partial p} &= \begin{bmatrix} \nabla_{wp}^2 \mathcal{L} \\ -\nabla_p G(w, p)^\top \\ -\nabla_p H_{\mathcal{A}}(w, p)^\top \end{bmatrix} = \begin{bmatrix} \nabla_{wp}^2 \mathcal{L} \\ -\nabla_p \bar{G}_{\mathcal{A}}(w, p)^\top \end{bmatrix} \end{aligned}$$



Theorem (Fiacco, 1976)

Suppose w^* satisfies the KKT conditions for $NLP(p^*)$, SOSC, LICQ and SCS hold. Then,

1. the primal-dual solution $z^* = (w^*, \lambda^*, \mu_{\mathcal{A}}^*)$ is unique;
2. there exists a unique continuously differentiable function $z(\cdot)$ defined in a neighborhood U of p^* such that $z(p) = (w(p), \lambda(p), \mu_{\mathcal{A}}(p))$ is a KKT point and $w(p)$ a local minimizer of $NLP(p)$, with

$$\frac{dz(p^*)}{dp} = -M^{-1}r. \quad (4)$$

3. SOSC, LICQ and SCS hold for $z(p)$, $p \in U$.

Sketch of proof: SOSC and LICQ imply the KKT matrix M being invertible. The active set does not change for a ball U with a sufficiently small radius. The IFT can be applied to the KKT system.

Sensitivity Computation

Generalized Tangential Predictors via SQP

Generalized Tangential Predictors and Derivatives via SBQP

In detail, Equation (4) gives the sensitivity

$$\begin{bmatrix} \frac{dw(p)}{dp} \\ -\frac{d\lambda(p)}{dp} \\ -\frac{d\mu_{\mathcal{A}}(p)}{dp} \end{bmatrix} = - \begin{bmatrix} \nabla_{ww}^2 \mathcal{L}(w, \lambda, \mu, p) & \nabla_w G(w, p) & \nabla_w H_{\mathcal{A}}(w, p) \\ \nabla_w G(w, p)^\top & 0 & 0 \\ \nabla_w H_{\mathcal{A}}(w, p)^\top & 0 & 0 \end{bmatrix}^{-1} \begin{bmatrix} \nabla_{wp}^2 \mathcal{L}(w, \lambda, \mu, p) \\ \nabla_p G(w, p)^\top \\ \nabla_p H_{\mathcal{A}}(w, p)^\top \end{bmatrix}. \quad (5)$$

Solving this linear system is equivalent to solving the QP, with $\delta w = \frac{dw(p)}{dp}$:

$$\begin{aligned} \min_{\delta w} \quad & \frac{1}{2} \delta w^\top (\nabla_{ww}^2 \mathcal{L}) \delta w + (\nabla_{wp}^2 \mathcal{L})^\top \delta w \\ \text{s.t.} \quad & \nabla_w G(w, p)^\top \delta w + \nabla_p G(w, p)^\top = 0 \\ & \nabla_w H_{\mathcal{A}}(w, p)^\top \delta w + \nabla_p H_{\mathcal{A}}(w, p)^\top = 0. \end{aligned} \quad (6)$$

The directional derivative in $\Delta p = p - \bar{p}$:

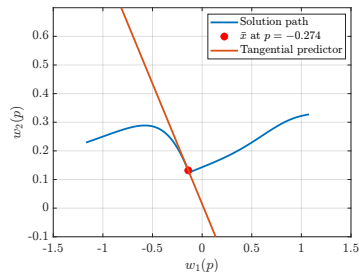
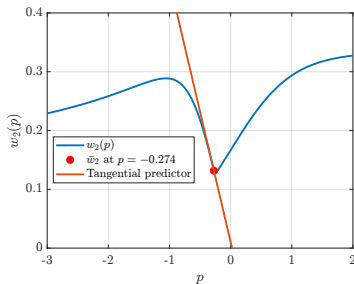
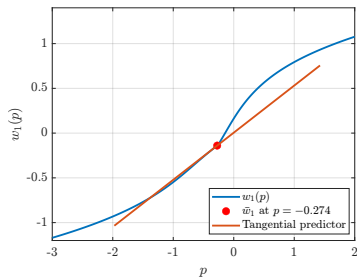
$$\hat{w}(p) = w(\bar{p}) + \frac{dw}{dp}(p - \bar{p}) = w(\bar{p}) + \underbrace{\frac{dw}{dp} \Delta p}_{:= \Delta w}$$

can be computed directly:

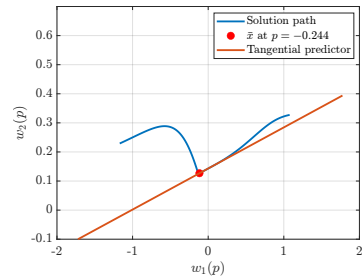
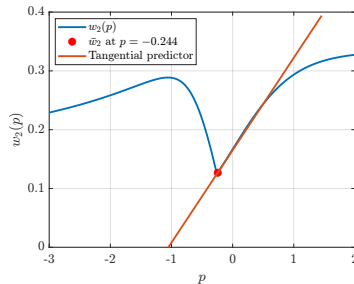
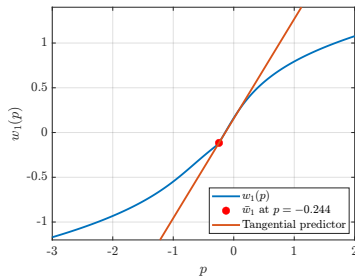
$$\begin{aligned} \min_{\Delta w} \quad & \frac{1}{2} \Delta w^\top (\nabla_{ww}^2 \mathcal{L}) \Delta w + (\nabla_{wp}^2 \mathcal{L} \Delta p) \Delta w \\ \text{s.t.} \quad & \nabla_w G(w, p)^\top \Delta w + \nabla_p G(w, p)^\top \Delta p = 0 \\ & \nabla_w H_{\mathcal{A}}(w, p)^\top \Delta w + \nabla_p H_{\mathcal{A}}(w, p)^\top \Delta p = 0. \end{aligned} \tag{7}$$

This QP is even more similar to an SQP subproblem.

Tangential predictor - just before the kink



Tangential predictor - just after the kink



The IFT tangential predictors are valid at smooth pieces in neighborhoods where SCS holds, at a kink and past it they are not defined.

We need a better way to compute directional derivatives at kinks, and approximations across them.

Sensitivity at active set changes

For a direction $\Delta p = p - p^*$, we consider a scalar parameter $t \in \mathbb{R}$, with $p = p^* + t\Delta p$, $t \in [0, 1]$. Thus we can regard the parametric NLP: $\text{NLP}(t)$ instead of $\text{NLP}(p)$.

Theorem (One-sided differentiability, Jittorntrum 1981, Diehl 2001)

Consider the parametric NLP(t), with $p = p^* + t(p - p^*)$, $t \in [0, 1]$. Let $z(0) = (w(0), \lambda(0), \mu(0))$ be a KKT point at p^* that satisfies LICQ and SSOSC, with a partition of strongly and weakly active constraints into $H_{\mathcal{A}^+}(w, t)$ and $H_{\mathcal{A}^0}(w, t)$, resp. Assume $(\delta w, \delta \lambda, \delta \mu_{\mathcal{A}^+}, \delta \mu_{\mathcal{A}^0})$ is the solution of the following QP, with functions evaluated at the solution $(w(0), \lambda(0), \mu(0))$:

$$\begin{aligned}
 \min_{\delta w} \quad & \frac{1}{2} \delta w^\top (\nabla_{ww}^2 \mathcal{L}) \delta w + (\nabla_{wt} \mathcal{L})^\top \delta w \\
 \text{s.t.} \quad & \nabla_t G(w, t)^\top + \nabla_w G(w, t)^\top \delta w = 0, \\
 & \nabla_t H_{\mathcal{A}^+}(w, t)^\top + \nabla_w H_{\mathcal{A}^+}(w, t)^\top \delta w = 0, \\
 & \nabla_t H_{\mathcal{A}^0}(w, t)^\top + \nabla_w H_{\mathcal{A}^0}(w, t)^\top \delta w \geq 0,
 \end{aligned} \tag{8}$$

Theorem (One-sided differentiability (continued), Jittorntrum 1981, Diehl 2001)

] ... which satisfies the strict complementarity for the multiplier vector $\delta\mu_{A^0}$ of the QP inequality constraints.

Then there exist an $\varepsilon > 0$ and a differentiable function $z : [0, \varepsilon) \rightarrow \mathbb{R}^{n_w} \times \mathbb{R}^{n_G} \times \mathbb{R}^{n_H}$, which are KKT points of $NLP(\varepsilon)$ that satisfy LICQ and SSOSC for $t \in [0, \varepsilon)$. At $t = 0$, the one-sided derivative of $z(t)$ is given by:

$$\lim_{t \rightarrow 0, t > 0} \frac{1}{t} \begin{pmatrix} w(t) - w(0) \\ \lambda(t) - \lambda(0) \\ \mu(t) - \mu(0) \end{pmatrix} = \begin{pmatrix} \delta w \\ \delta \lambda \\ \begin{pmatrix} \delta \mu \end{pmatrix} \end{pmatrix} = \begin{pmatrix} \delta w \\ \delta \lambda \\ \delta \mu_{A^+} \\ \delta \mu_{A^0} \\ 0 \end{pmatrix} \quad (9)$$

- ▶ Note that the QP (8) has always a unique solution. The QP is feasible, since $\delta w = 0$ is a feasible solution. By the hypothesis of the theorem LICQ and SSOSC hold, hence the QP is solvable and has a unique solution.
- ▶ We have the same assumptions as Fiacco's theorem, but show on top of that, starting at $z(0)$, any $z(\varepsilon)$, moving along $\varepsilon \Delta p$, which as **no new active constraints**, is differentiable. Only **weakly active** constraints are allowed to **become inactive**.
- ▶ The only additional assumption was the strict complementarity in the QP, which ensures no active set changes along our direction of interest.
- ▶ We looked only at the r.h.s. derivative, but regarding a $t \in (-\varepsilon, 0]$, we can make similar conclusions for the l.h.s. derivative.
- ▶ If SCS holds, there is no ineq. constraint in the QP, and the curve $z(t)$ is cont. differentiable at $t = 0$ (Fiacco's theorem under SCS).

Summary of QPs for computing linear predictions

We need to compute directional derivatives for $\Delta p = p - p^*$. This can be represented via a scalar parameter t , with $p = p^* + t\Delta p, t \in [0, 1]$.

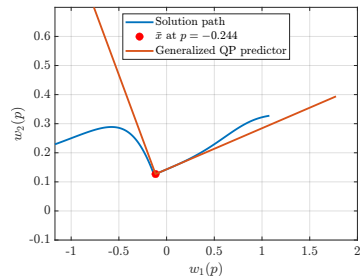
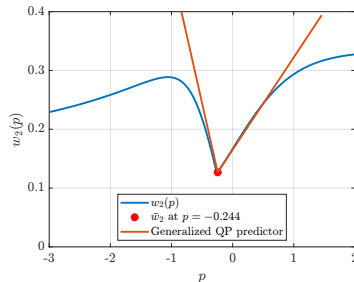
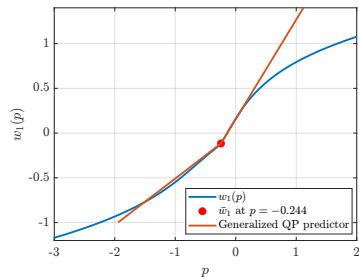
Under SCS, solve an equality-constrained QP (IFT)

$$\begin{aligned} \min_{\delta w} \quad & \frac{1}{2} \delta w^\top (\nabla_{ww}^2 \mathcal{L}) \delta w + (\nabla_{wt} \mathcal{L})^\top \delta w \\ \text{s.t.} \quad & \nabla_t G(w, t) + \nabla_w G(w, t)^\top \delta w = 0, \\ & \nabla_t H_{\mathcal{A}}(w, t) + \nabla_w H_{\mathcal{A}}(w, t)^\top \delta w = 0. \end{aligned}$$

Without SCS, solve an inequality-constrained QP (complete Fiacco's theorem)

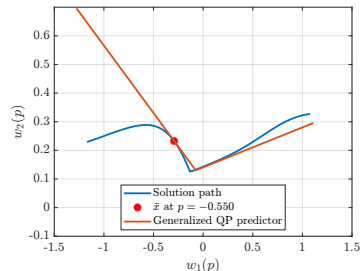
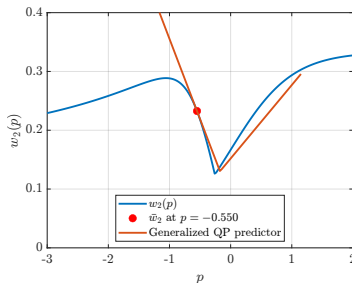
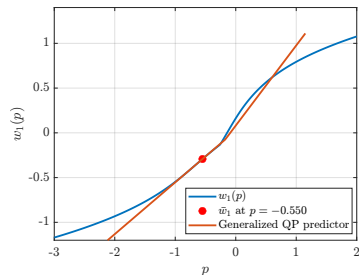
$$\begin{aligned} \min_{\delta w} \quad & \frac{1}{2} \delta w^\top (\nabla_{ww}^2 \mathcal{L}) \delta w + (\nabla_{wt} \mathcal{L})^\top \delta w \\ \text{s.t.} \quad & \nabla_t G(w, t) + \nabla_w G(w, t)^\top \delta w = 0, \\ & \nabla_t H_{\mathcal{A}^+}(w, t) + \nabla_w H_{\mathcal{A}^+}(w, t)^\top \delta w = 0, \\ & \nabla_t H_{\mathcal{A}^0}(w, t) + \nabla_w H_{\mathcal{A}^0}(w, t)^\top \delta w \geq 0. \end{aligned}$$

Parametric solution as function of p - generalized approximation



Solving a QP at a kink gives a reasonable way to compute directional derivatives, as they always exist. This is formalized next.

What about active set changes?



- ▶ As long as the active set match, and the derivative is Lipschitz (at least along the direction of interest) we have quadratic accuracy $O(\|\Delta p\|^2)$.
- ▶ Past a kink, the (directional) derivative is invalid. However, by solving an even more general QP (see next three slides) we can still obtain a good approximation. We can interpret this QP as a **piecewise linear approximation** of $z(p)$.

SQP subproblem with inequality constraints

$$\begin{aligned} \min_{\Delta w} \quad & \nabla F(w^k)^T \Delta w + \frac{1}{2} \Delta w^T A^k \Delta w \\ \text{s.t.} \quad & \begin{cases} G(w^k) + \nabla G(w^k)^T \Delta w = 0 \\ H(w^k) + \nabla H(w^k)^T \Delta w \geq 0 \end{cases} \end{aligned}$$

with

$$A^k = \nabla_w^2 \mathcal{L}(w^k, \lambda^k, \mu^k)$$

The solution to the QP subproblem delivers

$$w^k + \Delta w_{\text{QP}} \approx w^*, \quad \lambda_{\text{QP}} \approx \lambda^*, \quad \mu_{\text{QP}} \approx \mu^*$$

Theorem (Generalized Tangential Predictor)

Consider the parametric NLP(p) and let $z^* = (w^*, \lambda^*, \mu^*)$ be a KKT point at p^* that satisfies LICQ and SSOSC. For a parameter step $\Delta p := \bar{p} - p^*$ from p^* to a neighboring parameter $\bar{p}$ regard the solution of the following QP, with functions evaluated at (z^*, p^*) :

$$\begin{aligned} \min_{\Delta w} \quad & \frac{1}{2} \Delta w^\top (\nabla_{ww}^2 \mathcal{L}) \Delta w + \Delta w^\top ((\nabla_{wt} \mathcal{L}) \Delta p + \nabla_w F) \\ \text{s.t.} \quad & \nabla_p G^\top \Delta p + \nabla_w G^\top \Delta w = 0, \\ & \nabla_p H^\top \Delta p + \nabla_w H^\top \Delta w \geq 0, \end{aligned} \tag{10}$$

The solution delivers a generalized tangential predictor to the possibly nonsmooth parametric solution $w^{\text{sol}}(p)$, i.e., $\|w^* + \Delta w - w^{\text{sol}}(p)\| = O(\|\Delta p\|^2)$.

Theorem (Generalized Tangential Predictor via SQP, Diehl 2001)

Consider the parametric NLP(p) and let $z^* = (w^*, \lambda^*, \mu^*)$ be a KKT point at p^* that satisfies LICQ and SSOSC. For a parameter step $\Delta p := \bar{p} - p^*$ from p^* to a neighboring parameter $\bar{p}$ regard the solution of the following QP, with functions evaluated at (z^*, p^*) :

$$\begin{aligned} \min_{\Delta y = (\Delta w, \Delta p)} \quad & \frac{1}{2} \Delta y^\top (\nabla_{yy}^2 \mathcal{L}) \Delta y + \Delta y^\top \nabla_y F \\ \text{s.t.} \quad & p^* - \bar{p} + \Delta p = 0, \\ & G + \nabla_y G^\top \Delta y = 0, \\ & H + \nabla_y H^\top \Delta y \geq 0, \end{aligned} \tag{11}$$

The solution delivers a generalized tangential predictor to the possibly nonsmooth parametric solution $w^{\text{sol}}(p)$, i.e., $\|w^* + \Delta w - w^{\text{sol}}(p)\| = O(\|\Delta p\|^2)$.

Note: The above QP is identical to the SQP subproblem for a "lifted" NLP where the parameters p are extra variables constrained by the extra constraint $p - \bar{p} = 0$.

- ▶ Under SCS, SOSC and LICQ $\implies$ IFT $\implies z(p)$ is a $\mathcal{C}^1$ function (and thus Lipschitz).
- ▶ Without SCS $\implies$ no IFT since $z(p)$ is not differentiable, only directional derivatives $\frac{dz(p)}{dp} \Delta p$ exist.
 - ▶ even in case of (some) changes in the active set, we can compute the directional derivatives, and obtain $z(p)$ as a piecewise linear function

All complications arise because of inequality constraints... can we get rid of them?

- ▶ use the barrier reformulation of the NLP
- ▶ simple and practical, but approximate

Sensitivity Computation

Generalized Tangential Predictors via SQP

Generalized Tangential Predictors and Derivatives via SBQP

Log-barrier parametric NLP:

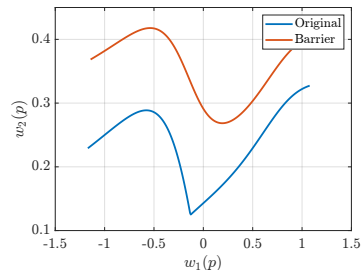
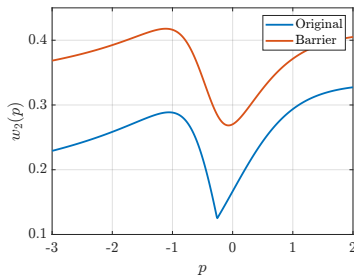
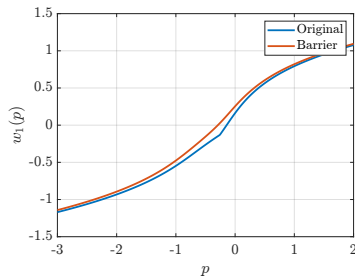
$$\begin{aligned} \min_{w \in \mathbb{R}^{n_w}} \quad & F(w, p) - \tau \sum_{i=1}^{n_H} \log(H_i(w, p)) \\ \text{s.t.} \quad & G(w, p) = 0. \end{aligned}$$

- ▶ only equality constraints (SSOSC = SSOC), and under SOSC and LICQ we can apply the IFT to compute the sensitivity
- ▶ IFT is applied to the smoothed KKT system:

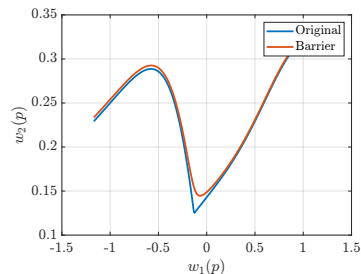
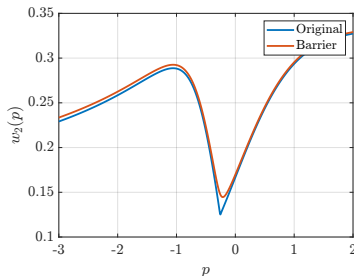
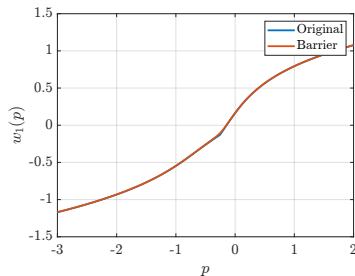
$$R_\tau(w, s, \lambda, \mu, p) = \begin{bmatrix} \nabla_w \mathcal{L}(w, \lambda, \mu, p) \\ G(w, p) \\ H(w, p) - s \\ \text{diag}(s)\mu - \tau e \end{bmatrix} = 0, \quad (s, \mu > 0)$$

- ▶ Note that both τ and p are parameters, but we only vary p and assume τ to be fixed.

Parametric solution $z(p)$ – smoothed solution manifold $\tau = 0.1$



Parametric solution $z(p)$ – smoothed solution manifold $\tau = 0.002$



Setting: solve NLP with acados SQP

- SQP solves QP in Δw space of primal variables

Theorem: Denote QP solution map at NLP solution $\Delta w_{\text{QP}}^{\text{sol}}(p, z^*)$. For **exact** Hessian QP, the solution maps $w^{\text{sol}}(\theta)$ and $w^* + \Delta w_{\text{QP}}^{\text{sol}}(p, z^*)$, and their sensitivities, $\frac{\partial w^{\text{sol}}}{\partial p}(p)$ and $\frac{\partial \Delta w_{\text{QP}}^{\text{sol}}}{\partial p}(p, z^*)$ coincide.

An efficient implementation of sensitivity computation in acados

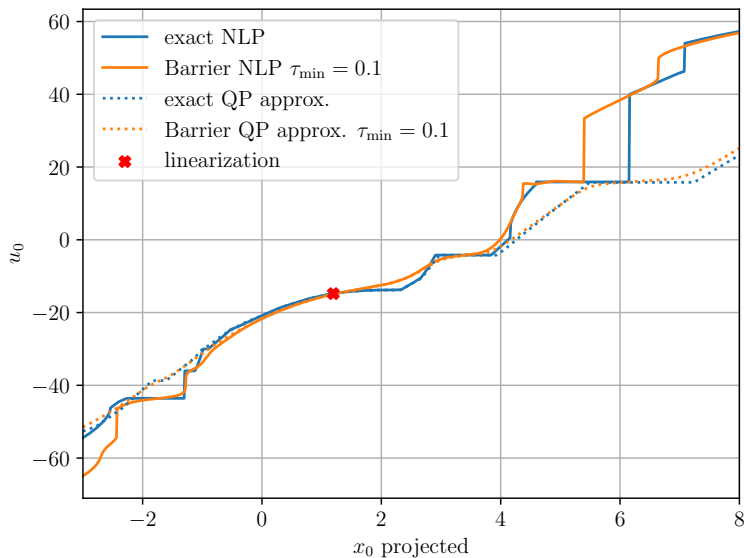
Setting: solve NLP with acados SQP

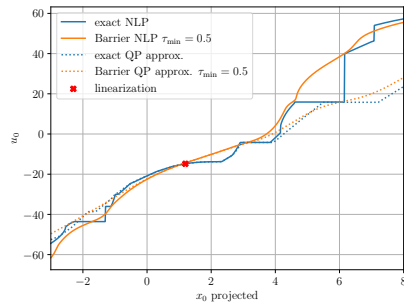
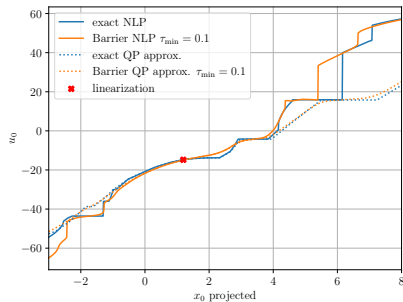
- ▶ SQP solves QP in Δw space of primal variables

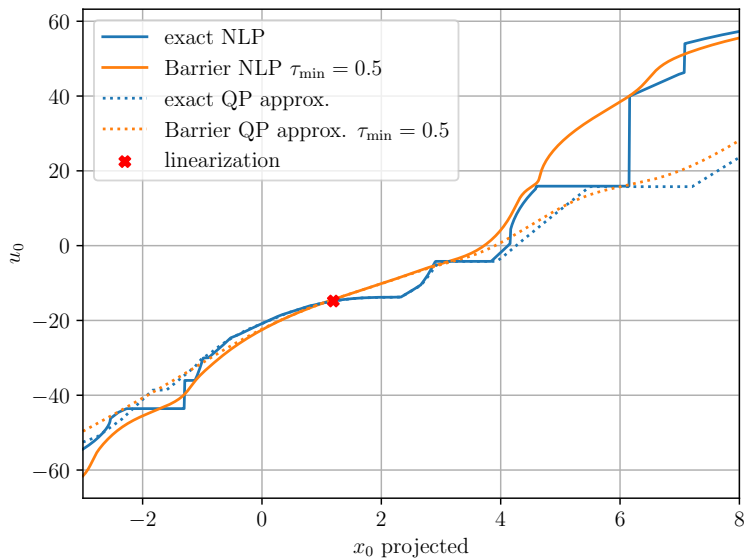
Theorem: Denote QP solution map at NLP solution $\Delta w_{\text{QP}}^{\text{sol}}(p, z^*)$. For **exact** Hessian QP, the solution maps $w^{\text{sol}}(\theta)$ and $w^* + \Delta w_{\text{QP}}^{\text{sol}}(p, z^*)$, and their sensitivities, $\frac{\partial w^{\text{sol}}}{\partial p}(p)$ and $\frac{\partial \Delta w_{\text{QP}}^{\text{sol}}}{\partial p}(p, z^*)$ coincide.

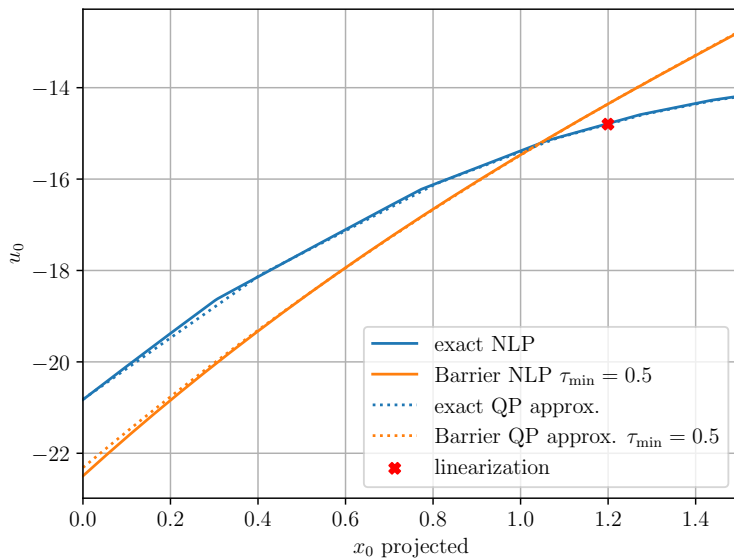
Blending SQP with IP QP solver (HPIPM): Shrink τ in QP solver to $\tau_{\min} > 0$ instead of 0.

- ▶ Not SQP but Sequential Barrier Quadratic Programming (SBQP)
- ▶ Convergence to IP-smoothed KKT solution









- ▶ Hessian approximations often beneficial in SQP
 - ▶ convergence
 - ▶ computational complexity
 - ▶ regularity
- ▶ Regularization needed when dealing with indefinite Hessians
- ▶ IFT requires **exact** Hessian

⇒ **Two-solver approach**

1. Nominal solver: can use approximate Hessian, regularization etc.
2. Sensitivity solver
 - ▶ load solution
 - ▶ evaluate exact Hessian
 - ▶ evaluate partial derivatives w.r.t. θ
 - ▶ solve linear system **efficiently** with HPIPM Riccati

Thank you very much for your attention

- ▶ NLP solutions depend locally differentiably on parameters, but have kinks at active set changes (and sometimes even jumps, e.g. when the global optimizer changes)
- ▶ Generalized tangential predictors can be obtained via a QP at kinky points where LICQ and SSOCS hold
- ▶ this QP can automatically be obtained via SQP, when embedding the parameter as an optimization variable into the NLP (this happens automatically in the real-time iteration)
- ▶ Alternatively, the same trick can be used with Sequential Barrier Quadratic Programming (SBQP), which can deliver generalized (and non-generalized) tangential predictors for the corresponding barrier NLP
- ▶ acados delivers the exact SBQP sensitivity using a two-solver approach (that needs to use the exact Hessian only for the sensitivity)