Model Predictive Control and Reinforcement Learning
– Lecture 4: Constrained Nonlinear Optimization –

Joschka Boedecker and Moritz Diehl

University of Freiburg

Slides jointly developed with Armin Nurkanović, Florian Messerer, Katrin Baumgärtner, Andrea Ghezzi, Jonathan Frey.

Fall School on Model Predictive Control and Reinforcement Learning
Freiburg, 6-10 October 2025

universität freiburg

NTNU | Norwegian University of Science and Technology

Outline of the lecture

Nonlinear MPC solves Nonlinear Programs

Optimality Conditions for Constrained Optimization

Nonlinear Programming Algorithms

Sensitivity Computation

Outline of the lecture

Nonlinear MPC solves Nonlinear Programs

Optimality Conditions for Constrained Optimization

Nonlinear Programming Algorithms

Sensitivity Computation

Recap and Today's topic

Minimize (or maximize) an objective function $F(w)$ depending on decision variables w **subject to equality and/or inequality constrains**

Recap and Today's topic

Minimize (or maximize) an objective function $F(w)$ depending on decision variables w **subject to equality and/or inequality constrains**

An optimization problem

$$\min_{w} \; F(w) \tag{1a}$$
$$\text{s.t.} \quad G(w) = 0 \tag{1b}$$
$$H(w) \geq 0 \tag{1c}$$

Terminology

- w - decision variable
- F: objective/cost function
- G, H: equality and inequality constraint functions

- Optimization is a powerful tool used in all quantitative sciences
- Only in few special cases a closed form solution exist
- Use an iterative algorithm to find solution
- The optimization problem may be parametric, and all functions depend on a fixed parameter p

Direct optimal control methods formulate Nonlinear Programs (NLP)

Continuous-time OCP

$$\min_{x(\cdot),u(\cdot)} \int_0^T L_c(x(t),u(t))\,\mathrm{d}t + E(x(T))$$

$$\text{s.t.} \quad x(0) = \bar{x}_0$$

$$\dot{x}(t) = f_c(x(t),u(t))$$

$$0 \geq h(x(t),u(t)),\ t \in [0,T]$$

$$0 \geq r(x(T))$$

Direct methods like direct collocation, multiple shooting. *First discretize, then optimize.*

Direct optimal control methods formulate Nonlinear Programs (NLP)

Continuous-time OCP

$$\min_{x(\cdot),u(\cdot)} \int_0^T L_c(x(t),u(t))\,\mathrm{d}t + E(x(T))$$

$$\text{s.t.} \quad x(0) = \bar{x}_0$$

$$\dot{x}(t) = f_c(x(t),u(t))$$

$$0 \geq h(x(t),u(t)), \; t \in [0,T]$$

$$0 \geq r(x(T))$$

Discrete-time OCP (an NLP)

$$\min_{x,u} \sum_{k=0}^{N-1} \ell(x_k,u_k) + E(x_N)$$

$$\text{s.t.} \quad x_0 = \bar{x}_0$$

$$x_{k+1} = f(x_k,u_k)$$

$$0 \geq h(x_k,u_k), \; k = 0,\dots,N-1$$

$$0 \geq r(x_N)$$

Direct methods like direct collocation, multiple shooting. *First discretize, then optimize.*

Nonlinear MPC solves Nonlinear Programs (NLP)

Discrete time NMPC Problem (an NLP)

$$
\begin{aligned}
\min_{x,u} \ & \sum_{k=0}^{N-1} \ell(x_k, u_k) + E(x_N) \\
\text{s.t.} \quad & x_0 = \bar{x}_0 \\
& x_{k+1} = f(x_k, u_k) \\
& 0 \geq h(x_k, u_k), \ k = 0, \dots, N-1 \\
& 0 \geq r(x_N)
\end{aligned}
$$

Variables $x = (x_0, \dots, x_N)$ and $u = (u_0, \dots, u_{N-1})$ can be summarized in vector $w = (x, u) \in \mathbb{R}^n$.

Nonlinear MPC solves Nonlinear Programs (NLP)

Discrete time NMPC Problem (an NLP)

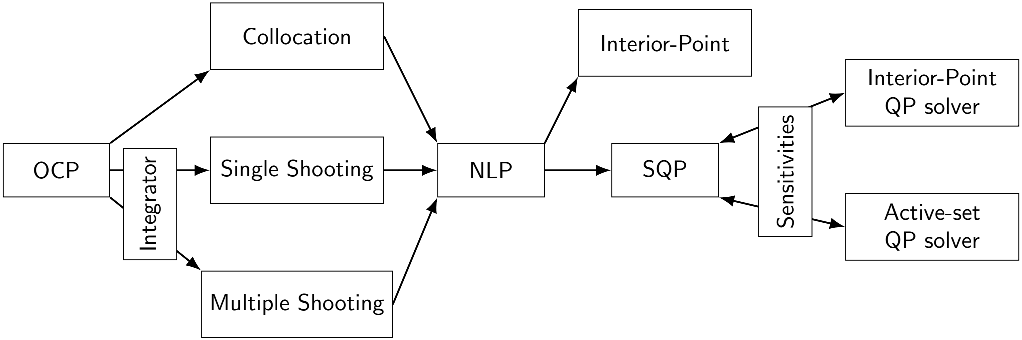
$$\begin{aligned} \min_{x,u} \; & \sum_{k=0}^{N-1} \ell(x_k, u_k) + E(x_N) \\ \text{s.t.} \; & x_0 = \bar{x}_0 \\ & x_{k+1} = f(x_k, u_k) \\ & 0 \geq h(x_k, u_k), \; k = 0, \dots, N-1 \\ & 0 \geq r(x_N) \end{aligned}$$

Variables $x = (x_0, \dots, x_N)$ and $u = (u_0, \dots, u_{N-1})$ can be summarized in vector $w = (x, u) \in \mathbb{R}^n$.

Nonlinear Program (NLP)

$$\begin{aligned} \min_{w \in \mathbb{R}^n} \; & F(w) \\ \text{s.t.} \; & G(w) = 0 \\ & H(w) \geq 0 \end{aligned}$$

The Map of Numerical Optimal Control



Outline of the lecture

Nonlinear MPC solves Nonlinear Programs

Optimality Conditions for Constrained Optimization

Nonlinear Programming Algorithms

Sensitivity Computation

FONC for equality constraints

Nonlinear Program (NLP)

$$\min_{w \in \mathbb{R}^n} \; F(w)$$
$$\text{s.t. } G(w) = 0$$

$\mathcal{L}(w, \lambda) = F(w) - \lambda^\top G(w)$ is the Lagrangian

FONC for equality constraints

Nonlinear Program (NLP)

$$\min_{w \in \mathbb{R}^n} \; F(w)$$
$$\text{s.t. } G(w) = 0$$

$\mathcal{L}(w, \lambda) = F(w) - \lambda^\top G(w)$ is the Lagrangian

Definition (LICQ)

A point w satisfies Linear Independence Constraint Qualification LICQ if and only if $\nabla G\,(w)$ is full column rank

FONC for equality constraints

Nonlinear Program (NLP)

$$\min_{w \in \mathbb{R}^n} \; F(w)$$
$$\text{s.t. } G(w) = 0$$

$\mathcal{L}(w, \lambda) = F(w) - \lambda^\top G(w)$ is the Lagrangian

Definition (LICQ)

A point w satisfies Linear Independence Constraint Qualification LICQ if and only if $\nabla G(w)$ is full column rank

First-order Necessary Conditions

Let F, G in $\mathcal{C}^1$. If w^* is a (local) minimizer, **and** w^* satisfies LICQ, then there is a unique vector λ such that:

$$\nabla_w \mathcal{L}(w^*, \lambda^*) = \nabla F(w^*) - \nabla G(w^*)\lambda = 0 \qquad \text{Dual feasibility}$$
$$\nabla_\lambda \mathcal{L}(w^*, \lambda^*) = G(w^*) = 0 \qquad \text{Primal feasibility}$$

The KKT conditions

Nonlinear Program (NLP)

$$\min_{w \in \mathbb{R}^n} \; F(w)$$
$$\text{s.t. } G(w) = 0$$
$$H(w) \geq 0$$

$$\mathcal{L}(w, \lambda) = F(w) - \lambda^\top G(w) - \mu^\top H(w)$$

The KKT conditions

Nonlinear Program (NLP)

$$\min_{w \in \mathbb{R}^n} \; F(w)$$
$$\text{s.t.} \; G(w) = 0$$
$$H(w) \geq 0$$

$$\mathcal{L}(w, \lambda) = F(w) - \lambda^\top G(w) - \mu^\top H(w)$$

Definition (LICQ)

A point w satisfies LICQ if and only if

$$[\nabla G(w), \quad \nabla H_{\mathcal{A}}(w)]$$

is full column rank

Active set $\mathcal{A} = \{i \mid H_i(w) = 0\}$

The KKT conditions

Nonlinear Program (NLP)

$$\min_{w \in \mathbb{R}^n} \; F(w)$$
$$\text{s.t. } G(w) = 0$$
$$H(w) \geq 0$$

$$\mathcal{L}(w, \lambda) = F(w) - \lambda^\top G(w) - \mu^\top H(w)$$

Definition (LICQ)

A point w satisfies LICQ if and only if

$$[\nabla G(w), \quad \nabla H_{\mathcal{A}}(w)]$$

is full column rank

Active set $\mathcal{A} = \{i \mid H_i(w) = 0\}$

Theorem (KKT conditions)

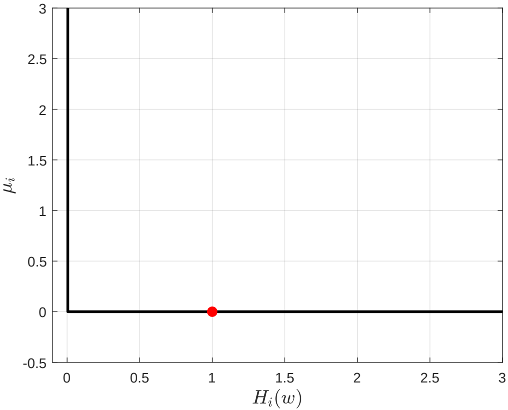
*Let F, G, H be $\mathcal{C}^1$. If w^* is a (local) minimizer **and** satisfies LICQ, then there are unique vectors λ^* and μ^* such that (w^*, λ^*, μ^*) satisfies:*

$$\nabla_w \mathcal{L}(w^*, \mu^*, \lambda^*) = 0, \quad \mu^* \geq 0, \qquad \textit{Dual feasibility}$$
$$G(w^*) = 0, \quad H(w^*) \geq 0 \qquad \textit{Primal feasibility}$$
$$\mu_i^* H_i(w^*) = 0, \quad \forall i \qquad \textit{Complementary slackness}$$

The complementary slackness condition

Active constraints:

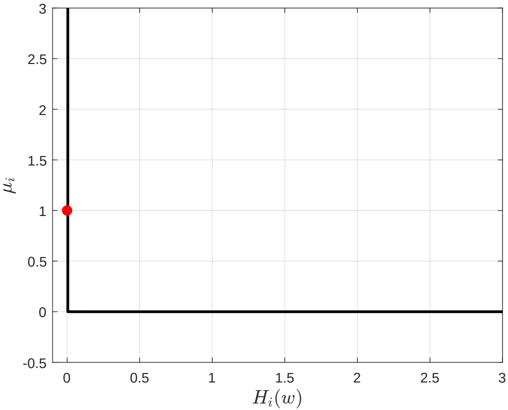
- $H_i(w^*) > 0$ then $\mu_i^* = 0$, and H_i is inactive



The complementary slackness condition

Active constraints:

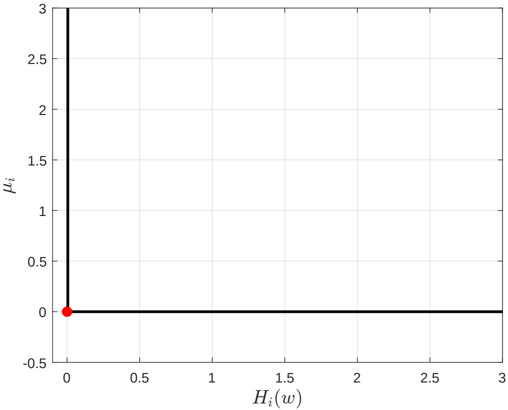
- $H_i(w^*) > 0$ then $\mu_i^* = 0$, and H_i is inactive
- $\mu_i^* > 0$ and $H_i(w) = 0$ then $H_i(w)$ is strictly active



The complementary slackness condition

Active constraints:

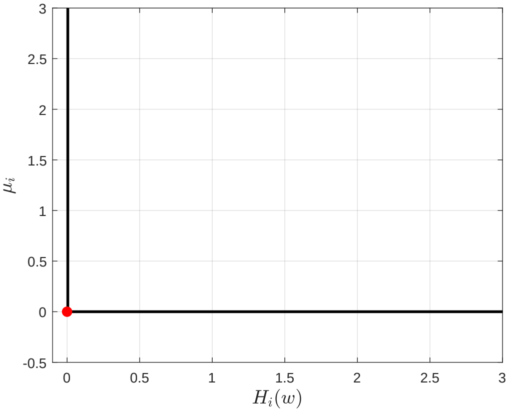
- $H_i(w^*) > 0$ then $\mu_i^* = 0$, and H_i is inactive
- $\mu_i^* > 0$ and $H_i(w) = 0$ then $H_i(w)$ is strictly active
- $\mu_i^* = 0$ and $H_i(w) = 0$ then then $H_i(w)$ is weakly active



The complementary slackness condition

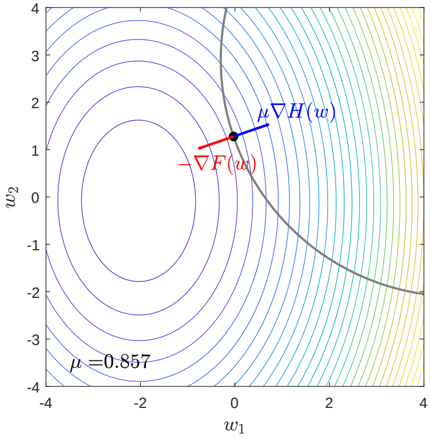
Active constraints:

- $H_i(w^*) > 0$ then $\mu_i^* = 0$, and H_i is inactive
- $\mu_i^* > 0$ and $H_i(w) = 0$ then $H_i(w)$ is strictly active
- $\mu_i^* = 0$ and $H_i(w) = 0$ then then $H_i(w)$ is weakly active
- We define the **active set** $\mathbb{A}^*$ as the set of indices i of the active constraints



Some intuitions on the KKT conditions

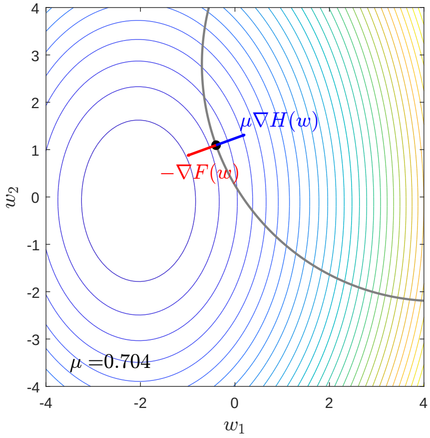
$$\min_{w \in \mathbb{R}^n} \; F(w)$$
$$\text{s.t. } H(w) \geq 0$$



Some intuitions on the KKT conditions

$$\min_{w \in \mathbb{R}^n} \; F(w)$$
$$\text{s.t. } H(w) \geq 0$$

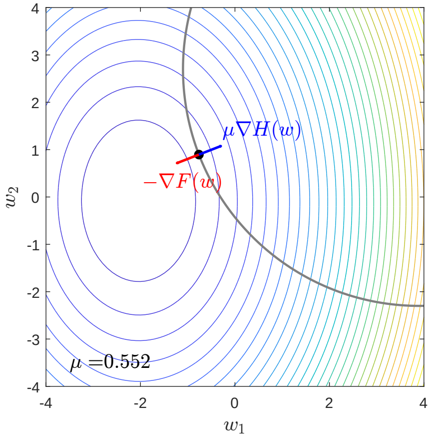
- $-\nabla F$ is the gravity



Some intuitions on the KKT conditions

$$\min_{w \in \mathbb{R}^n} \ F(w)$$
$$\text{s.t. } H(w) \geq 0$$

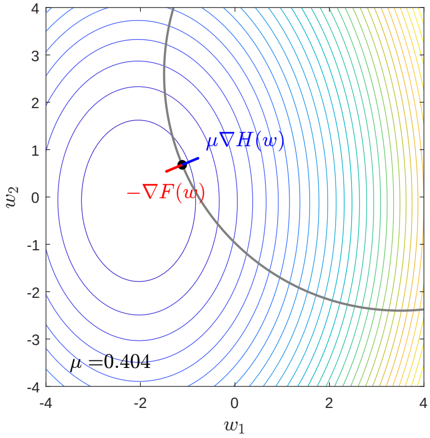
- $-\nabla F$ is the gravity



Some intuitions on the KKT conditions

$$\min_{w \in \mathbb{R}^n} \; F(w)$$
$$\text{s.t. } H(w) \geq 0$$

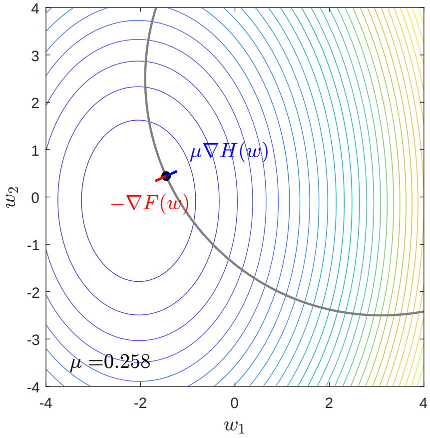
- $-\nabla F$ is the gravity
- $\mu \nabla H$ is the force of the fence. Sign $\mu \geq 0$ means the fence can only "push" the ball



Some intuitions on the KKT conditions

$$\min_{w \in \mathbb{R}^n} \; F(w)$$
$$\text{s.t. } H(w) \geq 0$$

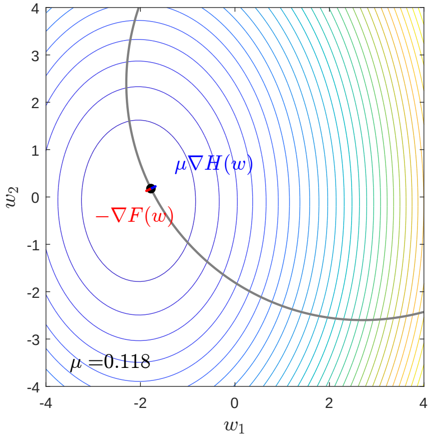
- $-\nabla F$ is the gravity
- $\mu \nabla H$ is the force of the fence. Sign $\mu \geq 0$ means the fence can only "push" the ball
- ∇H gives the direction of the force and μ adjusts the magnitude.



Some intuitions on the KKT conditions

$$\min_{w\in\mathbb{R}^n} \; F(w)$$
$$\text{s.t. } H(w) \geq 0$$

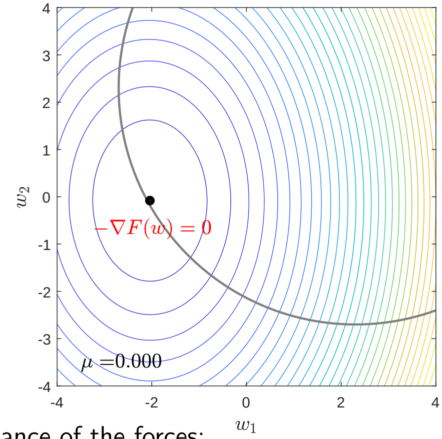
- $-\nabla F$ is the gravity
- $\mu\nabla H$ is the force of the fence. Sign $\mu \geq 0$ means the fence can only "push" the ball
- ∇H gives the direction of the force and μ adjusts the magnitude.



Some intuitions on the KKT conditions

$$\min_{w \in \mathbb{R}^n} \; F(w)$$
$$\text{s.t. } H(w) \geq 0$$

- $-\nabla F$ is the gravity
- $\mu \nabla H$ is the force of the fence. Sign $\mu \geq 0$ means the fence can only "push" the ball
- ∇H gives the direction of the force and μ adjusts the magnitude.
- Weakly active constraint: $H(w) = 0, \; \mu = 0$ the ball touches the fence but no force is needed



Balance of the forces:

$$\nabla \mathcal{L}(w, \mu) = \nabla F(w) - \mu \nabla H(w) = 0$$

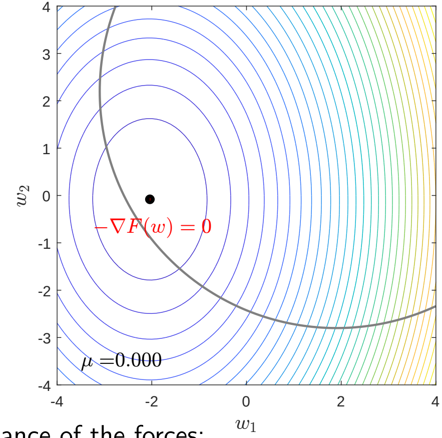
Some intuitions on the KKT conditions

$$\min_{w \in \mathbb{R}^n} \; F(w)$$
$$\text{s.t. } H(w) \geq 0$$

- $-\nabla F$ is the gravity
- $\mu \nabla H$ is the force of the fence. Sign $\mu \geq 0$ means the fence can only "push" the ball
- ∇H gives the direction of the force and μ adjusts the magnitude.
- Weakly active constraint: $H\left(w\right) = 0, \;\; \mu = 0$ the ball touches the fence but no force is needed
- Inactive constraint $H\left(w\right) > 0, \;\; \mu = 0$

$$H\left(w\right) > 0, \quad \mu = 0$$

- Complementary slackness $\mu H = 0$ describes a contact problem



Balance of the forces:

$$\nabla \mathcal{L}(w, \mu) = \nabla F(w) - \mu \nabla H(w) = 0$$

Summary of optimality conditions

Optimality conditions for NLP with equality and/or inequality constraints:

- **First-Order Necessary Conditions**: Under LICQ and differentiability, a local optimum of the NLP satisfies the KKT conditions.

Summary of optimality conditions

Optimality conditions for NLP with equality and/or inequality constraints:

- **First-Order Necessary Conditions**: Under LICQ and differentiability, a local optimum of the NLP satisfies the KKT conditions.
- **Second-Order (Necessary or) Sufficient Conditions** require positive-(semi)-definiteness of the Hessian in so called critical directions (feasible and non-ascent directions)

Summary of optimality conditions

Optimality conditions for NLP with equality and/or inequality constraints:

- **First-Order Necessary Conditions**: Under LICQ and differentiability, a local optimum of the NLP satisfies the KKT conditions.
- **Second-Order (Necessary or) Sufficient Conditions** require positive-(semi)-definiteness of the Hessian in so called critical directions (feasible and non-ascent directions)

Nonconvex problem $\Rightarrow$ minimum is not necessarily global.
But some nonconvex problems have a unique minimum

Outline of the lecture

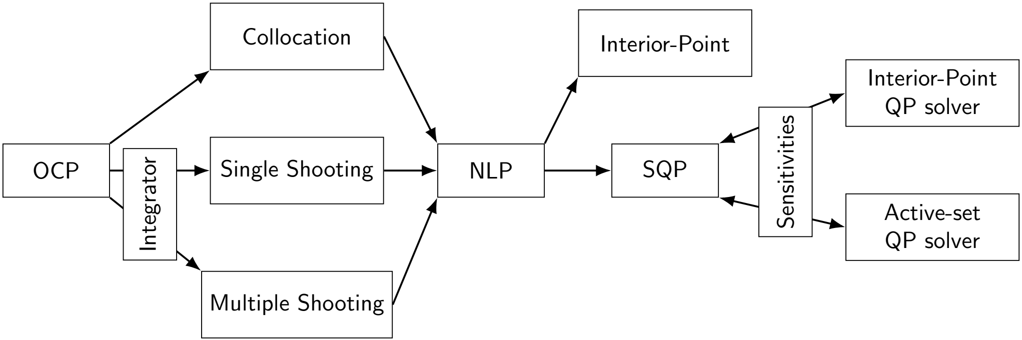
Nonlinear MPC solves Nonlinear Programs

Optimality Conditions for Constrained Optimization

Nonlinear Programming Algorithms

Sensitivity Computation

The Map of Numerical Optimal Control



General Nonlinear Program (NLP)

In direct methods, we have to solve the discretized optimal control problem, which is a Nonlinear Program (NLP)

General Nonlinear Program (NLP)

$$\min_{w} F(w) \quad \text{s.t.} \quad \begin{cases} G(w) & = & 0 \\ H(w) & \geq & 0 \end{cases}$$

We first treat the case without inequalities

NLP only with equality constraints

$$\min_{w} F(w) \quad \text{s.t.} \quad G(w) \; = \; 0$$

Lagrange function and optimality conditions

Lagrange function

$$\mathcal{L}(w, \lambda) = F(w) - \lambda^\top G(w)$$

Then for an optimal solution w^* exist multipliers λ^* such that

Nonlinear root-finding problem

$$\begin{aligned} \nabla_w \mathcal{L}(w^*, \lambda^*) &= 0 \\ G(w^*) &= 0 \end{aligned}$$

Newton's Method on optimality conditions

How to solve nonlinear equations

$$\begin{aligned} \nabla_w \mathcal{L}(w^*, \lambda^*) &= 0 \\ G(w^*) &= 0 \quad ? \end{aligned}$$

Linearize!

$$\begin{aligned} \nabla_w \mathcal{L}(w^k, \lambda^k) \quad + \nabla_w^2 \mathcal{L}(w^k, \lambda^k) \Delta w \quad - \nabla_w G(w^k) \Delta \lambda &= 0 \\ G(w^k) \quad + \nabla_w G(w^k)^\top \Delta w &= 0 \end{aligned}$$

This is equivalent, due to $\nabla \mathcal{L}(w^k, \lambda^k) = \nabla F(w^k) - \nabla G(w^k) \lambda^k$ with the shorthand $\lambda^+ = \lambda^k + \Delta\lambda$ to

$$\begin{aligned} \nabla_w F(w^k) \quad + \nabla_w^2 \mathcal{L}(w^k, \lambda^k) \Delta w \quad - \nabla_w G(w^k) \lambda^+ &= 0 \\ G(w^k) \quad + \nabla_w G(w^k)^\top \Delta w &= 0 \end{aligned}$$

Newton Step = Quadratic Program

Conditions

$$\begin{array}{rlrcl} \nabla_w F(w^k) & +\nabla_w^2 \mathcal{L}(w^k,\lambda^k)\Delta w & -\nabla_w G(w^k)\lambda^+ & = & 0 \\ G(w^k) & +\nabla_w G(w^k)^\top \Delta w & & = & 0 \end{array}$$

are optimality conditions of a quadratic program (QP), namely:

Quadratic program

$$\begin{array}{ll} \min\limits_{\Delta w} & \nabla F(w^k)^\top \Delta w + \frac{1}{2}\Delta w^\top A^k \Delta w \\ \text{s.t.} & G(w^k) + \nabla G(w^k)^\top \Delta w \quad = \quad 0, \end{array}$$

with

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

The full step Newton's Method iterates by solving in each iteration the Quadratic Progam

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

with $A^k = \nabla_w^2 \mathcal{L}(w^k, \lambda^k)$. This obtains as solution the step Δw^k and the new multiplier $\lambda_{\mathrm{QP}}^+ = \lambda^k + \Delta\lambda^k$

New iterate

$$\begin{aligned} w^{k+1} \quad &= \quad w^k + \Delta w^k \\ \lambda^{k+1} \quad &= \quad \lambda^k + \Delta\lambda^k = \lambda_{\mathrm{QP}}^+ \end{aligned}$$

This Newton's method is also called "Sequential Quadratic Programming (SQP) for equality constrained optimization" (with "exact Hessian" and "full steps")

NLP with Inequalities

Regard again NLP with both, equalities and inequalities:

NLP with equality and inequality constraints

$$\min_{w} F(w) \quad \text{s.t.} \quad \begin{cases} G(w) & = & 0 \\ H(w) & \geq & 0 \end{cases}$$

Lagrangian function for NLP with equality and inequality constraints

$$\mathcal{L}(w, \lambda, \mu) = F(w) - \lambda^\top G(w) - \mu^\top H(w)$$

Optimality conditions with inequalities

Theorem (Karush-Kuhn-Tucker (KKT) conditions)

Let F, G, H be $\mathcal{C}^2$. If w^* is a (local) minimizer and satisfies LICQ, then there are unique vectors λ^* and μ^* such that (w^*, λ^*, μ^*) satisfies:

$$\nabla_w \mathcal{L}(w^*, \mu^*, \lambda^*) = 0$$
$$G(w^*) = 0$$
$$H(w^*) \geq 0$$
$$\mu^* \geq 0$$
$$H(w^*)^\top \mu^* = 0$$

- These contain nonsmooth conditions (the last three) which are called *complementarity conditions*
- This system cannot be solved by Newton's Method. But still with SQP...

Sequential Quadratic Programming (SQP)

By Linearizing all functions within the KKT Conditions, and setting $\lambda^+ = \lambda^k + \Delta\lambda$ and $\mu^+ = \mu^k + \Delta\mu$, we obtain the KKT conditions of a Quadratic Program (QP) (we omit these conditions).

QP with inequality constraints

$$\min_{\Delta w} \quad \nabla F(w^k)^\top \Delta w + \frac{1}{2}\Delta w^\top A^k \Delta w$$

$$\text{s.t.} \quad \begin{cases} G(w^k) + \nabla G(w^k)^\top \Delta w & = & 0 \\ H(w^k) + \nabla H(w^k)^\top \Delta w & \geq & 0 \end{cases}$$

with

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

and its solution delivers

$$\Delta w^k, \quad \lambda_{\mathrm{QP}}^+, \quad \mu_{\mathrm{QP}}^+$$

Constrained Gauss-Newton Method

In special case of least squares objectives

Least squares objective function

$$F(w) = \frac{1}{2}\|R(w)\|_2^2$$

can approximate Hessian $\nabla_w^2 \mathcal{L}(w^k, \lambda^k, \mu^k)$ by much cheaper

$$A^k = \nabla R(w) \nabla R(w)^\top.$$

Need no multipliers to compute A^k! QP= linear least squares:

Gauss-Newton QP

$$\begin{aligned} \min_{\Delta w} \quad & \frac{1}{2}\|R(w^k) + \nabla R(w^k)^\top \Delta w\|_2^2 \\ \text{s.t.} \quad & G(w^k) + \nabla G(w^k)^\top \Delta w = 0 \\ & H(w^k) + \nabla H(w^k)^\top \Delta w \geq 0 \end{aligned}$$

Convergence: linear (better if $\|R(w^*)\|$ small)

Interior point methods

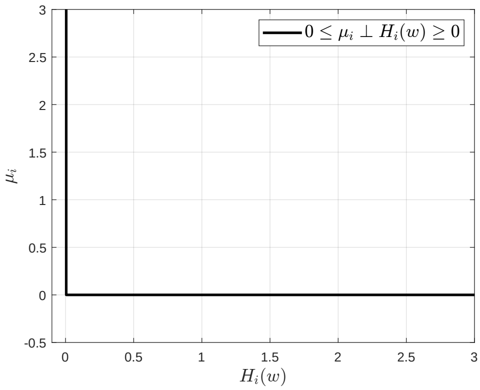
NLP with inequalites

$$\min_{w} \; F(w)$$
$$\text{s.t.} \quad H(w) \geq 0$$

KKT conditions

$$\nabla F(w) - \nabla H(w)^\top \mu = 0$$
$$0 \leq \mu \perp H(w) \geq 0$$

Main difficulty: inequality conditions introduce nonsmoothness in the KKT conditions



The barrier problem

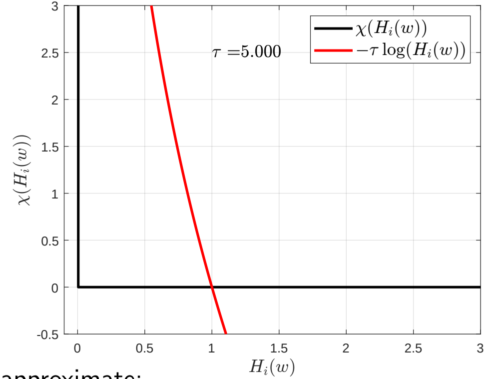
NLP with inequalites

$$\min_{w} \; F(w)$$
$$\text{s.t.} \quad H(w) \geq 0$$

Barrier problem

$$\min_{w} \; F(w) - \tau \sum_{i=1}^{m} \log(H_i(w)) =: F_\tau(w)$$

Main idea: put inequality constraint into objective



approximate:

$$\chi(H_i(w)) = \begin{cases} 0 & \text{if } H_i(w) \geq 0 \\ \infty & \text{if } H_i(w) < 0 \end{cases}$$

The barrier problem

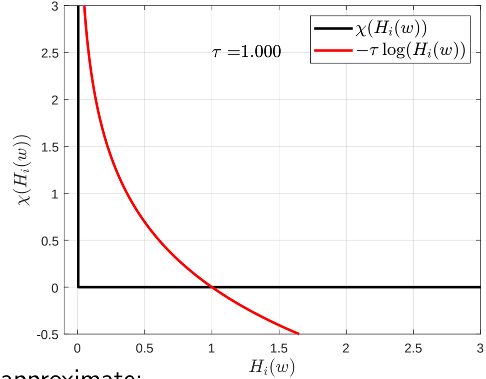
NLP with inequalites

$$\min_{w} \; F(w)$$
$$\text{s.t.} \quad H(w) \geq 0$$

Barrier problem

$$\min_{w} \; F(w) - \tau \sum_{i=1}^{m} \log(H_i(w)) =: F_\tau(w)$$

Main idea: put inequality constraint into objective



approximate:

$$\chi(H_i(w)) = \begin{cases} 0 & \text{if } H_i(w) \geq 0 \\ \infty & \text{if } H_i(w) < 0 \end{cases}$$

The barrier problem

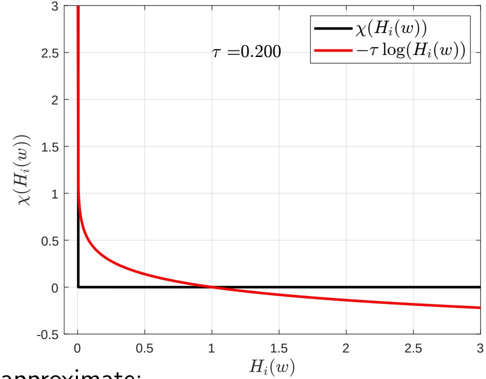
NLP with inequalites

$$\min_{w} \; F(w)$$
$$\text{s.t.} \quad H(w) \geq 0$$

Barrier problem

$$\min_{w} \; F(w) - \tau \sum_{i=1}^{m} \log(H_i(w)) =: F_\tau(w)$$

Main idea: put inequality constraint into objective



approximate:

$$\chi(H_i(w)) = \begin{cases} 0 & \text{if } H_i(w) \geq 0 \\ \infty & \text{if } H_i(w) < 0 \end{cases}$$

The barrier problem

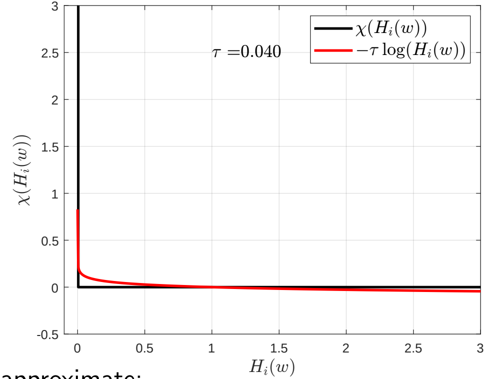
NLP with inequalites

$$\min_{w} \; F(w)$$
$$\text{s.t.} \quad H(w) \geq 0$$

Barrier problem

$$\min_{w} \; F(w) - \tau \sum_{i=1}^{m} \log(H_i(w)) =: F_\tau(w)$$

Main idea: put inequality constraint into objective



approximate:

$$\chi(H_i(w)) = \begin{cases} 0 & \text{if } H_i(w) \geq 0 \\ \infty & \text{if } H_i(w) < 0 \end{cases}$$

The barrier problem

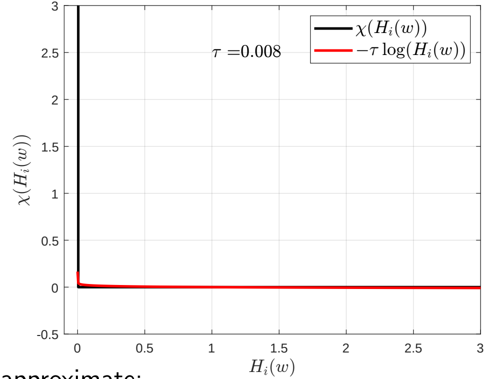
NLP with inequalites

$$\min_{w} \; F(w)$$
$$\text{s.t.} \quad H(w) \geq 0$$

Barrier problem

$$\min_{w} \; F(w) - \tau \sum_{i=1}^{m} \log(H_i(w)) =: F_\tau(w)$$

Main idea: put inequality constraint into objective



approximate:

$$\chi(H_i(w)) = \begin{cases} 0 & \text{if } H_i(w) \geq 0 \\ \infty & \text{if } H_i(w) < 0 \end{cases}$$

The barrier problem

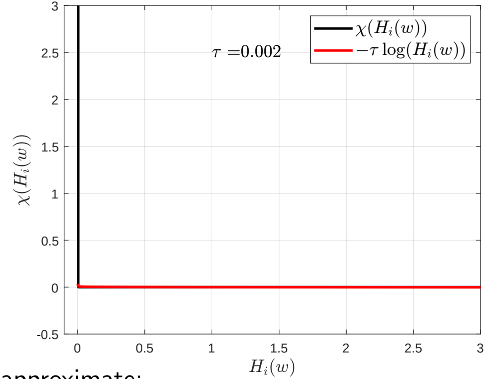
NLP with inequalites

$$\min_{w} \; F(w)$$
$$\text{s.t.} \quad H(w) \geq 0$$

Barrier problem

$$\min_{w} \; F(w) - \tau \sum_{i=1}^{m} \log(H_i(w)) =: F_\tau(w)$$

Main idea: put inequality constraint into objective



approximate:

$$\chi(H_i(w)) = \begin{cases} 0 & \text{if } H_i(w) \geq 0 \\ \infty & \text{if } H_i(w) < 0 \end{cases}$$

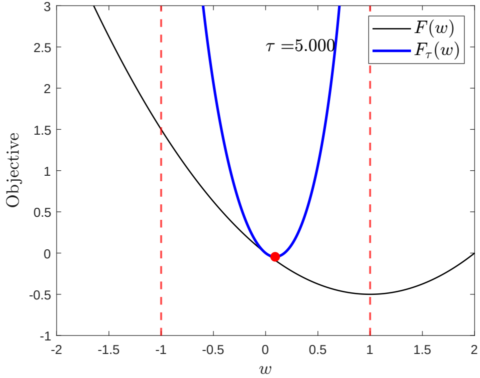
An example of the barrier problem

Example NLP

$$\min_{w} \; 0.5w^2 - 2w$$
$$\text{s.t.} \quad -1 \leq w \leq 1$$

Barrier problem

$$\min_{w} \; 0.5w^2 - 2 - \tau \log(w+1) - \tau \log(1-w)$$



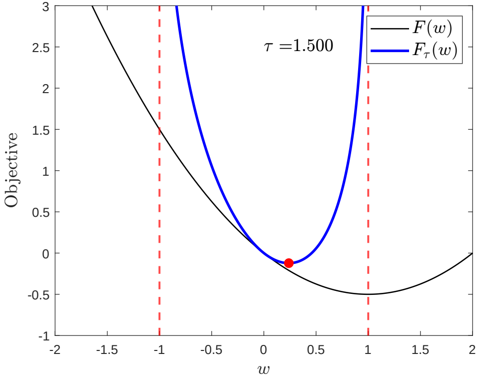
An example of the barrier problem

Example NLP

$$\min_{w} \; 0.5w^2 - 2w$$
$$\text{s.t.} \quad -1 \leq w \leq 1$$

Barrier problem

$$\min_{w} \; 0.5w^2 - 2 - \tau \log(w+1) - \tau \log(1-w)$$



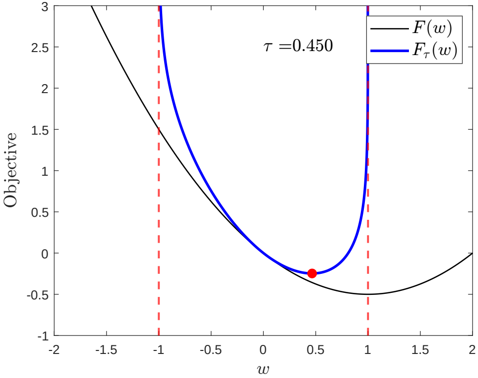
An example of the barrier problem

Example NLP

$$\min_{w} \; 0.5w^2 - 2w$$
$$\text{s.t.} \quad -1 \leq w \leq 1$$

Barrier problem

$$\min_{w} \; 0.5w^2 - 2 - \tau \log(w+1) - \tau \log(1-w)$$



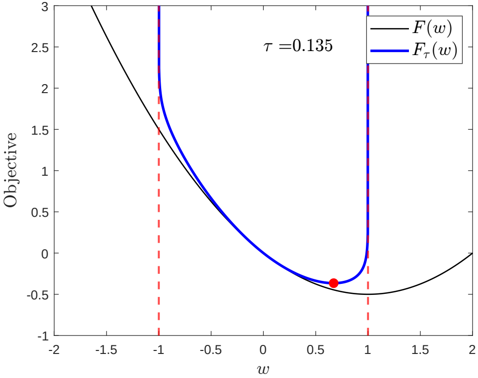
An example of the barrier problem

Example NLP

$$\min_{w} \; 0.5w^2 - 2w$$
$$\text{s.t.} \quad -1 \leq w \leq 1$$

Barrier problem

$$\min_{w} \; 0.5w^2 - 2 - \tau \log(w+1) - \tau \log(1-w)$$



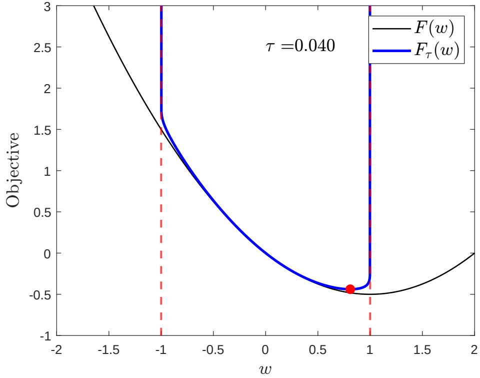
An example of the barrier problem

Example NLP

$$\min_{w} \; 0.5w^2 - 2w$$
$$\text{s.t.} \quad -1 \leq w \leq 1$$

Barrier problem

$$\min_{w} \; 0.5w^2 - 2 - \tau \log(w+1) - \tau \log(1-w)$$



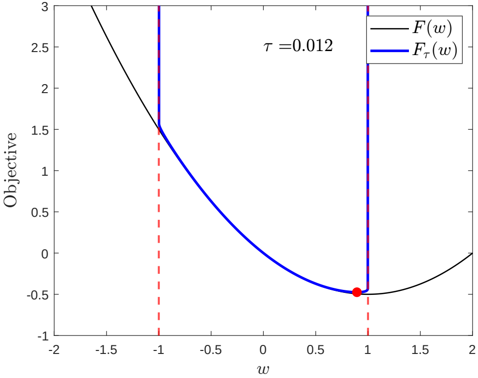
An example of the barrier problem

Example NLP

$$\min_{w} \ 0.5w^2 - 2w$$
$$\text{s.t.} \quad -1 \leq w \leq 1$$

Barrier problem

$$\min_{w} \ 0.5w^2 - 2 - \tau \log(w+1) - \tau \log(1-w)$$



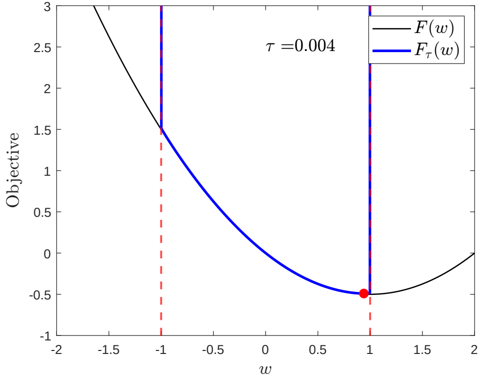
An example of the barrier problem

Example NLP

$$\min_{w} \; 0.5w^2 - 2w$$
$$\text{s.t.} \quad -1 \leq w \leq 1$$

Barrier problem

$$\min_{w} \; 0.5w^2 - 2 - \tau \log(w+1) - \tau \log(1-w)$$



Primal-dual interior point methods

Barrier problem

$$\min_{w} \; F(w) - \tau \sum_{i=1}^{m} \log(H_i(w)) =: F_\tau(w)$$

KKT conditions

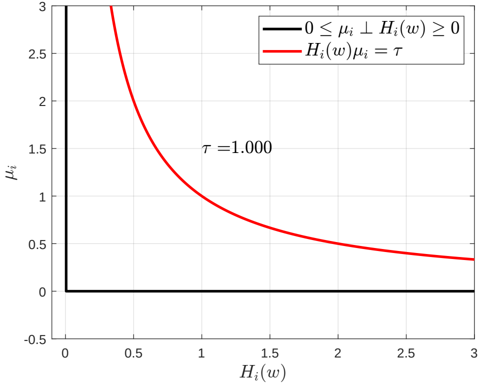
$$\nabla F(w) - \tau \sum_{i-1}^{m} \frac{1}{H_i(w)} \nabla H_i(w) = 0$$

Introduce variable $\mu_i = \frac{\tau}{H_i(w)}$

Smoothed KKT conditions

$$\nabla F(w) - \nabla H(w)^\top \mu = 0$$
$$H_i(w)\mu_i = \tau$$
$$(H_i(w) > 0, \mu_i > 0)$$

Solve nonsmooth system with Newtons' method



Primal-dual interior point methods

Barrier problem

$$\min_{w} \; F(w) - \tau \sum_{i=1}^{m} \log(H_i(w)) =: F_\tau(w)$$

KKT conditions

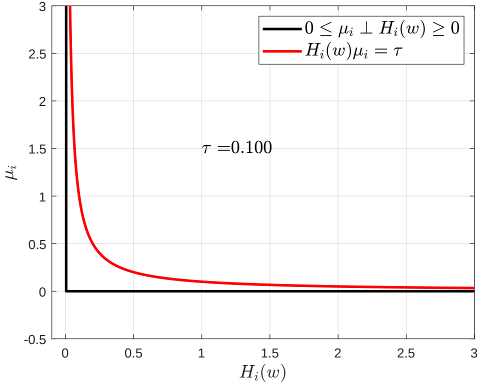
$$\nabla F(w) - \tau \sum_{i-1}^{m} \frac{1}{H_i(w)} \nabla H_i(w) = 0$$

Introduce variable $\mu_i = \frac{\tau}{H_i(w)}$

Smoothed KKT conditions

$$\nabla F(w) - \nabla H(w)^\top \mu = 0$$
$$H_i(w)\mu_i = \tau$$
$$(H_i(w) > 0, \mu_i > 0)$$

Solve nonsmooth system with Newtons' method



Primal-dual interior point methods

Barrier problem

$$\min_{w} \; F(w) - \tau \sum_{i=1}^{m} \log(H_i(w)) =: F_\tau(w)$$

KKT conditions

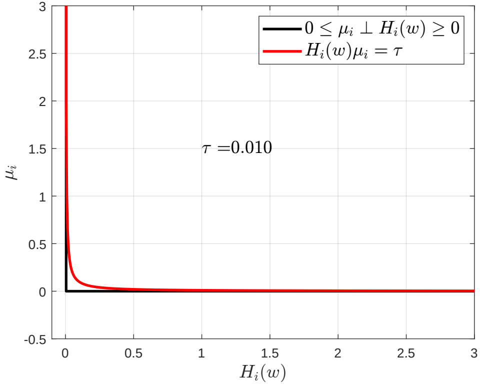
$$\nabla F(w) - \tau \sum_{i-1}^{m} \frac{1}{H_i(w)} \nabla H_i(w) = 0$$

Introduce variable $\mu_i = \frac{\tau}{H_i(w)}$

Smoothed KKT conditions

$$\nabla F(w) - \nabla H(w)^\top \mu = 0$$
$$H_i(w)\mu_i = \tau$$
$$(H_i(w) > 0, \mu_i > 0)$$

Solve nonsmooth system with Newtons' method



Primal-dual interior point methods

Barrier problem

$$\min_{w}\ F(w) - \tau \sum_{i=1}^{m} \log(H_i(w)) =: F_\tau(w)$$

KKT conditions

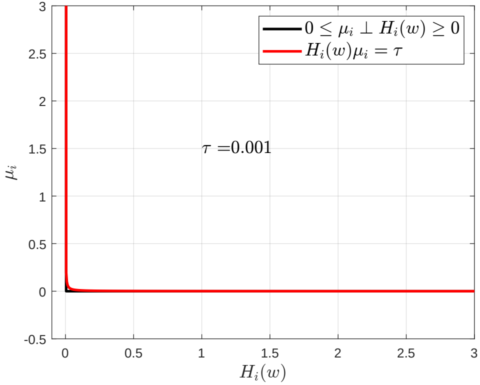
$$\nabla F(w) - \tau \sum_{i-1}^{m} \frac{1}{H_i(w)} \nabla H_i(w) = 0$$

Introduce variable $\mu_i = \frac{\tau}{H_i(w)}$

Smoothed KKT conditions

$$\nabla F(w) - \nabla H(w)^\top \mu = 0$$
$$H_i(w)\mu_i = \tau$$
$$(H_i(w) > 0, \mu_i > 0)$$

Solve nonsmooth system with Newtons' method



Primal-dual interior point method

Nonlinear programming problem

$$\min_{w} \; F(w)$$
$$\text{s.t.} \quad G(w) = 0$$
$$H(w) \geq 0$$

Smoothed KKT conditions

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

$(s, \mu > 0)$

$e = (1, \dots, 1)$

Solve approximately with Newton's method for fixed τ

$$R_\tau(w, s, \lambda, \mu) + \nabla R_\tau(w, s, \lambda, \mu)^\top \Delta z = 0$$

with $z = (w, s, \lambda, \mu)$

Line-search

Find $\alpha \in (0, 1)$

$$w^{k+1} = w^k + \alpha \Delta w$$
$$s^{k+1} = s^k + \alpha \Delta s$$
$$\lambda^{k+1} = \lambda^k + \alpha \Delta \lambda$$
$$\mu^{k+1} = \mu^k + \alpha \Delta \mu$$

such that $s^{k+1} > 0, \mu^{k+1} > 0$

and reduce τ...

Summary Newton-type optimization

- Newton type optimization solves the necessary optimality conditions
- Newton's method linearizes the nonlinear system in each iteration
- for constraints, need Lagrangian function, and KKT conditions
- for equalities KKT conditions are smooth, can apply Newton's method
- for inequalities KKT conditions are non-smooth, can apply Sequential Quadratic Programming (SQP)
- QPs with inequalities can be solved with interior point methods
- Also NLPs with inequalities can be solved with interior point methods (e.g. by the IPOPT solver)

Outline of the lecture

Nonlinear MPC solves Nonlinear Programs

Optimality Conditions for Constrained Optimization

Nonlinear Programming Algorithms

Sensitivity Computation

Solution sensitivities – Intro

Motivation

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

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

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

Solution sensitivities – Intro

Motivation

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

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

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

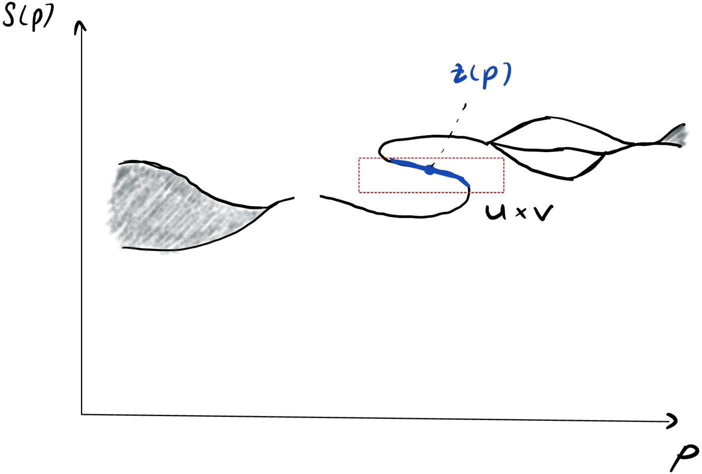
Need to introduce implicit functions

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

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

  with $R : \mathbb{R}^{n_z} \times \mathbb{R}^{n_p} \to \mathbb{R}^{n_z}$. 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:

$$z(p) \approx z(p^*) + \frac{\mathrm{d}z(p^*)}{\mathrm{d}p}(p - p^*).$$

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

Implicit function theorem

Theorem (Implicit function theorem)

Let $R : \mathbb{R}^{n_z} \times \mathbb{R}^{n_p} \to \mathbb{R}^{n_z}$ *be a* $\mathcal{C}^1$ *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* $\mathcal{C}^1$ *differentiable on* U, *with*

$$\frac{\mathrm{d}z(p)}{\mathrm{d}p} = -\left(\frac{\partial R}{\partial z}(z,p)\right)^{-1} \frac{\partial R}{\partial p}(z,p) \text{ for every } p \in U \, . \qquad (2)$$

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* $L = \max_{z \in C} \|\nabla f(z)\|$.

KKT conditions for parametric NLPs (p-NLP)

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

KKT conditions for parametric NLPs (p-NLP)

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

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

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

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

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

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.

Active set definitions

- The active set

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

- Strongly (strictly) 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 $\mathcal{A}^0(w,\mu,p)$.

The critical cone

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:

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

Second-order sufficent conditions for parametric NLPs

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

Second-order sufficent conditions for parametric NLPs

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

Definition (Second-order sufficient condition (SOSC))

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

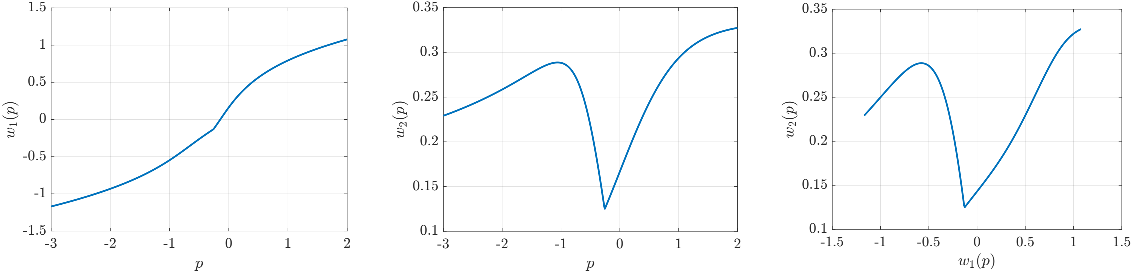
$$d^\top \nabla^2_{ww} \mathcal{L}(w, \lambda, \mu, p) d > 0 \quad \forall d \in \mathcal{C}(w, \mu, p) \setminus \{0\}. \tag{3}$$

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

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

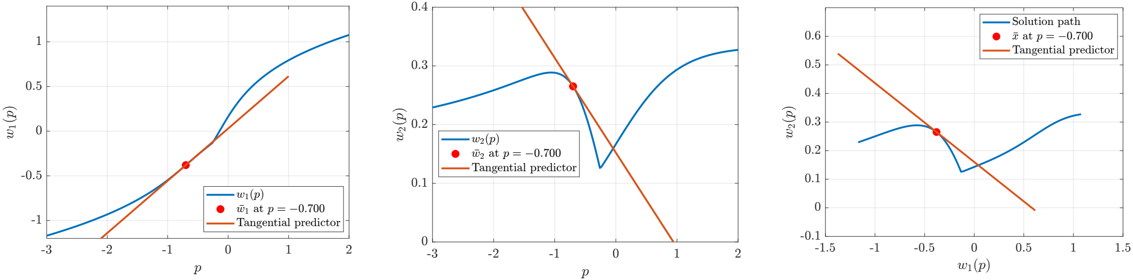
$$d^\top \nabla^2_{ww} \mathcal{L}(w, \lambda, \mu, p) d > 0 \quad \forall d \in \mathcal{D}(w, \mu, p) \setminus \{0\}. \tag{4}$$

Parametric solution $z(p)$ - solution manifold



Parametric solution as function of p - linear approximation

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



KKT conditions at the fixed active set and SCS

- 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:

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

- From IFT: $\frac{\mathrm{d}z(p)}{\mathrm{d}p} = -\left(\frac{\partial R}{\partial z}(z,p)\right)^{-1} \frac{\partial R}{\partial p}(z,p)$ for every $p \in U$, can we directly apply 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

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

- 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.

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

$$= \begin{bmatrix} \nabla^2_{ww}\mathcal{L} & -\nabla_w \bar{G}_\mathcal{A}(w,p) \\ -\nabla_w \bar{G}_\mathcal{A}(w,p)^\top & 0 \end{bmatrix},$$

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

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*
$$\frac{\mathrm{d}z(p^*)}{\mathrm{d}p} = -M^{-1}r \; . \tag{5}$$
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.

QP interpretation

In detail, Equation (5) gives the sensitivity

$$\begin{bmatrix} \frac{\mathrm{d}w(p)}{\mathrm{d}p} \\ -\frac{\mathrm{d}\lambda(p)}{\mathrm{d}p} \\ -\frac{\mathrm{d}\mu_{\mathcal{A}}(p)}{\mathrm{d}p} \end{bmatrix} = -\begin{bmatrix} \nabla^2_{ww}\mathcal{L}(w,\lambda,\mu,p) & \nabla_w G(w,p) & \nabla_w H_{\mathcal{A}}(w,p) \\ \nabla_w G(w,p)^\top & 0 & 0 \\ \nabla_w H_{\mathcal{A}}(w,p)^\top & 0 & 0 \end{bmatrix}^{-1} \begin{bmatrix} \nabla^2_{wp}\mathcal{L}(w,\lambda,\mu,p) \\ \nabla_p G(w,p)^\top \\ \nabla_p H_{\mathcal{A}}(w,p)^\top \end{bmatrix}. \tag{6}$$

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

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

Computing a linear approximation directly

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

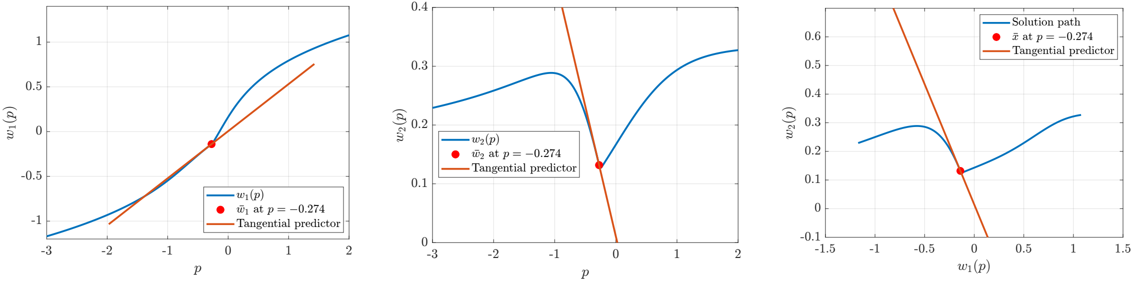
$$\hat{w}(p) = w(\bar{p}) + \frac{\mathrm{d}w}{\mathrm{d}p}(p - \bar{p}) = w(\bar{p}) + \underbrace{\frac{\mathrm{d}w}{\mathrm{d}p}\Delta p}_{:=\Delta w}$$

can be computed directly:

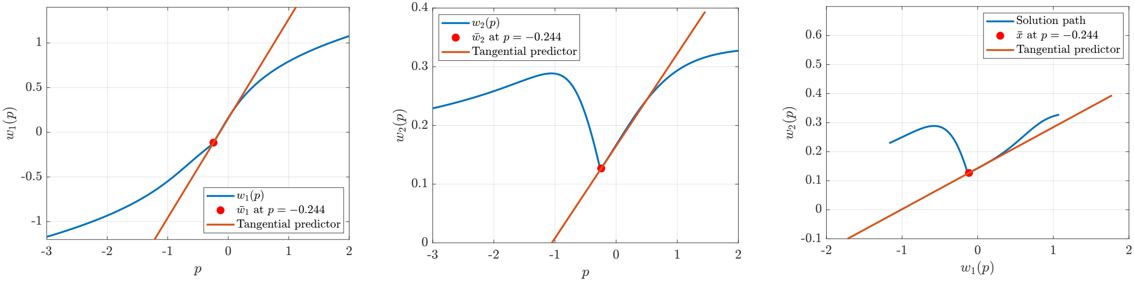
$$\begin{aligned} \min_{\Delta w}\ & \frac{1}{2}\Delta w^\top(\nabla^2_{ww}\mathcal{L})\Delta w + (\nabla^2_{wp}\mathcal{L}\Delta p)\Delta w \\ \text{s.t.}\quad & \nabla_w G(w,p)^\top \Delta w + \nabla_p G(w,p)^\top \Delta p = 0 \\ & \nabla_w H_{\mathcal{A}}(w,p)^\top \Delta w + \nabla_p H_{\mathcal{A}}(w,p)^\top \Delta p = 0. \end{aligned} \tag{8}$$

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: NLP(t) instead of 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* $(w(0), \lambda(0), \mu(0))$*:*

$$\begin{aligned} \min_{\delta w} \quad & \frac{1}{2}\delta w^\top (\nabla^2_{ww}\mathcal{L})\delta w + (\nabla_{wt}\mathcal{L})^\top \delta w \\ s.t. \quad & \nabla_t G(w,t)^\top + \nabla_w G(w,t)^\top \delta w = 0, \\ & \nabla_t H_{\mathcal{A}^+}(w,t)^\top + \nabla_w H_{\mathcal{A}^+}(w,t)^\top \delta w = 0, \\ & \nabla_t H_{\mathcal{A}^0}(w,t)^\top + \nabla_w H_{\mathcal{A}^0}(w,t)^\top \delta w \geq 0, \end{aligned} \tag{9}$$

Sensitivity at active set changes (continued)

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

] ... which satisfies the strict complementarity for the multiplier vector $\delta\mu_{\mathcal{A}^0}$ of the QP inequality constraints.
Then there exist an $\varepsilon > 0$ and a differentiable function $z : [0, \varepsilon) \to \mathbb{R}^{n_w} \times \mathbb{R}^{n_G} \times \mathbb{R}^{n_H}$, which are KKT points of NLP(ε) that satisfy LICQ and SSOSC for $t \in [0, \varepsilon)$. At $t = 0$, the one-sided derivative of $z(t)$ is given by:

$$\lim_{t \to 0, t > 0} \frac{1}{t} \begin{pmatrix} w(t) - w(0) \\ \lambda(t) - \lambda(0) \\ \mu(t) - \mu(0) \end{pmatrix} = \begin{pmatrix} \delta w \\ \delta \lambda \\ \begin{pmatrix} \\ \delta\mu \\ \\ \end{pmatrix} \end{pmatrix} = \begin{pmatrix} \delta w \\ \delta\lambda \\ \delta\mu_{\mathcal{A}^+} \\ \delta\mu_{\mathcal{A}^0} \\ 0 \end{pmatrix} \tag{10}$$

Some remarks on the theorem

- Note that the QP (9) 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 $p = p^* + t\Delta p, t \in [0, 1]$.

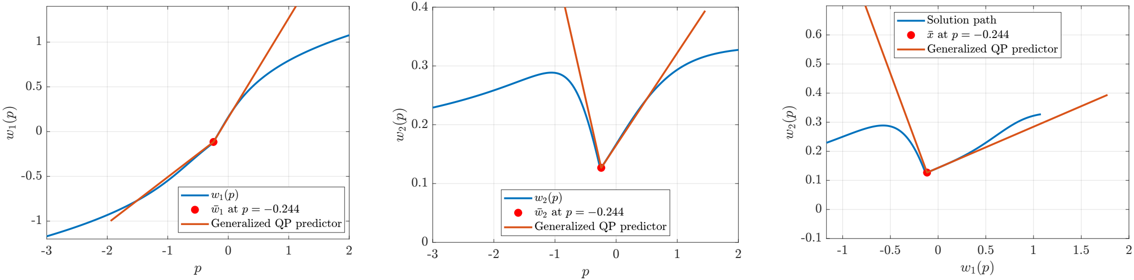
Under SCS, solve an equality-constrained QP (IFT)

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

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

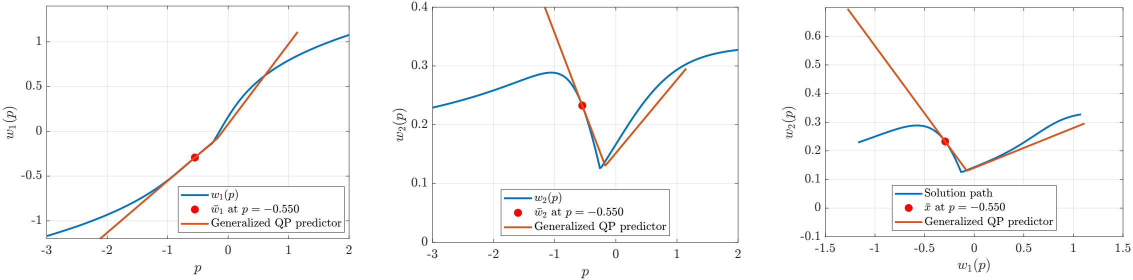
$$
\begin{aligned}
\min_{\delta w} \quad & \frac{1}{2}\delta w^\top (\nabla^2_{ww}\mathcal{L})\delta w + (\nabla_{wt}\mathcal{L})^\top \delta w \\
\text{s.t.} \quad & \nabla_t G(w,t) + \nabla_w G(w,t)^\top \delta w = 0, \\
& \nabla_t H_{\mathcal{A}^+}(w,t) + \nabla_w H_{\mathcal{A}^+}(w,t)^\top \delta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

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 $z(p)$.

Recall: QP Subproblem in Sequential Quadratic Programming (SQP)

SQP subproblem with inequality constraints

$$\min_{\Delta w} \quad \nabla F(w^k)^T \Delta w + \frac{1}{2} \Delta w^T A^k \Delta w$$

$$\text{s.t.} \quad \begin{cases} G(w^k) + \nabla G(w^k)^T \Delta w & = & 0 \\ H(w^k) + \nabla H(w^k)^T \Delta w & \geq & 0 \end{cases}$$

with

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

The solution to the QP subproblem delivers

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

Generalized Tangential Predictor

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^*)*:*

$$\begin{aligned} \min_{\Delta w} \quad & \frac{1}{2}\Delta w^\top (\nabla^2_{ww}\mathcal{L})\Delta w + \Delta w^\top((\nabla_{wt}\mathcal{L})\Delta p + \nabla_w F) \\ \text{s.t.} \quad & \nabla_p G^\top \Delta p + \nabla_w G^\top \Delta w = 0, \\ & \nabla_p H^\top \Delta p + \nabla_w H^\top \Delta w \geq 0, \end{aligned} \tag{11}$$

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

Generalized Tangential Predictor - SQP Interpretation

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^*)*:*

$$\begin{aligned} \min_{\Delta y = (\Delta w, \Delta p)} \quad & \frac{1}{2} \Delta y^\top (\nabla^2_{yy} \mathcal{L}) \Delta y + \Delta y^\top \nabla_y F \\ \text{s.t.} \quad & p^* - \bar{p} + \Delta p = 0, \\ & G + \nabla_y G^\top \Delta y = 0, \\ & H + \nabla_y H^\top \Delta y \geq 0, \end{aligned} \tag{12}$$

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

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 $p - \bar{p} = 0$.

Alternative ways to deal with inequality constraints

- 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 $\frac{\mathrm{d}z(p)}{\mathrm{d}p}\Delta p$ 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 way
- heuristic and approximate

Log-barrier approximation - usually paired with IP methods

Log-barrier parametric NLP:

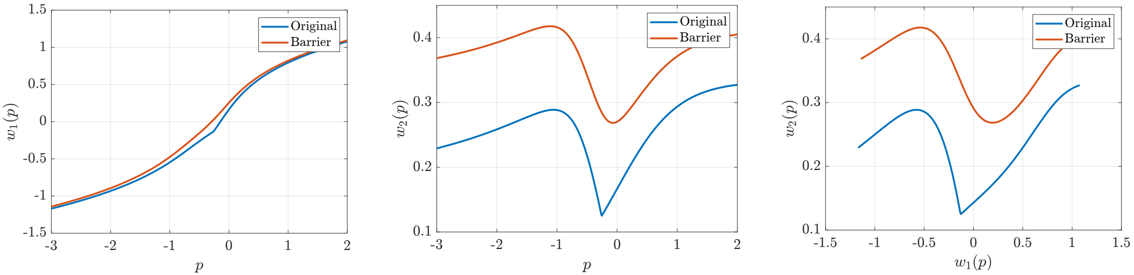
$$\begin{aligned} \min_{w \in \mathbb{R}^{n_w}} \quad & F(w,p) - \tau \sum_{i=1}^{n_H} \log(H_i(w,p)) \\ \text{s.t.} \quad & G(w,p) = 0. \end{aligned}$$

- 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:

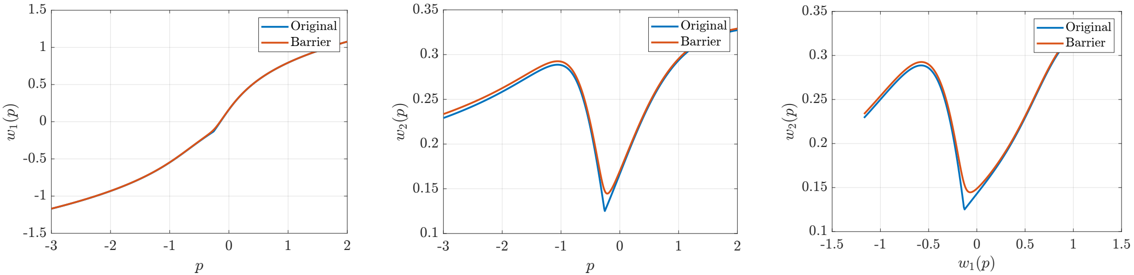
$$R_\tau(w,s,\lambda,\mu,p) = \begin{bmatrix} \nabla_w \mathcal{L}(w,\lambda,\mu,p) \\ G(w,p) \\ H(w,p) - s \\ \mathrm{diag}(s)\mu - \tau e \end{bmatrix} = 0, \qquad (s,\mu > 0)$$

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

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



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



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 $\Delta w_{\mathrm{QP}}^{\mathrm{sol}}(p, z^\star)$. For exact Hessian QP, the solution maps $w^{\mathrm{sol}}(\theta)$ and $w^\star + \Delta w_{\mathrm{QP}}^{\mathrm{sol}}(p, z^\star)$, and their sensitivities, $\frac{\partial w^{\mathrm{sol}}}{\partial p}(p)$ and $\frac{\partial \Delta w_{\mathrm{QP}}^{\mathrm{sol}}}{\partial p}(p, z^\star)$ 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 $\Delta w_{\mathrm{QP}}^{\mathrm{sol}}(p, z^\star)$. For exact Hessian QP, the solution maps $w^{\mathrm{sol}}(\theta)$ and $w^\star + \Delta w_{\mathrm{QP}}^{\mathrm{sol}}(p, z^\star)$, and their sensitivities, $\frac{\partial w^{\mathrm{sol}}}{\partial p}(p)$ and $\frac{\partial \Delta w_{\mathrm{QP}}^{\mathrm{sol}}}{\partial p}(p, z^\star)$ coincide.

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

- Not an SQP method for $\tau_{\min} > 0$
- Convergence to IP-smoothed KKT solution

Hessian approximations & Two-solver approach

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

$\Longrightarrow$ **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

References

- Moritz Diehl, Sébastien Gros. "Numerical optimal control (Draft)," Lecture notes, 2019.
- Jorge Nocedal, Stephen J. Wright, Numerical optimization. New York, NY: Springer New York, 2006.