Nonlinear Optimization

Moritz Diehl

Systems Control and Optimization Laboratory, University of Freiburg, Germany

Robust Nonlinear MPC Course, University of California Santa Barbara
March 25-28, 2024

(slides jointly developed with **Armin Nurkanović**, Titus Quah, Katrin Baumgärtner, Jim Rawlings)

universität freiburg

Outline of the lecture

What is an optimization problem?

Optimization is used in all quantitative sciences and engineering. Its aim is to minimize (or maximize) an objective function $F(w)$ depending on decision variables $w = (w_1, \dots, w_n)$ subject to constraints.

What is an optimization problem?

Optimization is used in all quantitative sciences and engineering. Its aim is to minimize (or maximize) an objective function $F(w)$ depending on decision variables $w = (w_1, \ldots, w_n)$ subject to constraints.

Optimization Problem

$$\min_{w \in \mathbb{R}^n} \quad F(w) \tag{1a}$$

$$\text{s.t.} \quad G(w) = 0 \tag{1b}$$

$$H(w) \geq 0 \tag{1c}$$

What is an optimization problem?

Optimization is used in all quantitative sciences and engineering. Its aim is to minimize (or maximize) an objective function $F(w)$ depending on decision variables $w = (w_1, \ldots, w_n)$ subject to constraints.

Optimization Problem

$$\min_{w \in \mathbb{R}^n} \; F(w) \tag{1a}$$

$$\text{s.t.} \quad G(w) = 0 \tag{1b}$$

$$H(w) \geq 0 \tag{1c}$$

Terminology

- $w \in \mathbb{R}^n$ - vector of decision variables
- $F : \mathbb{R}^n \rightarrow \mathbb{R}$ - objective function
- $G : \mathbb{R}^n \rightarrow \mathbb{R}^{n_G}$ - equality constraints
- $H : \mathbb{R}^n \rightarrow \mathbb{R}^{n_H}$ - inequality constraints

What is an optimization problem?

Optimization is used in all quantitative sciences and engineering. Its aim is to minimize (or maximize) an objective function $F(w)$ depending on decision variables $w = (w_1, \dots, w_n)$ subject to constraints.

Optimization Problem

$$\min_{w \in \mathbb{R}^n} \; F(w) \tag{1a}$$

$$\text{s.t.} \quad G(w) = 0 \tag{1b}$$

$$H(w) \geq 0 \tag{1c}$$

Terminology

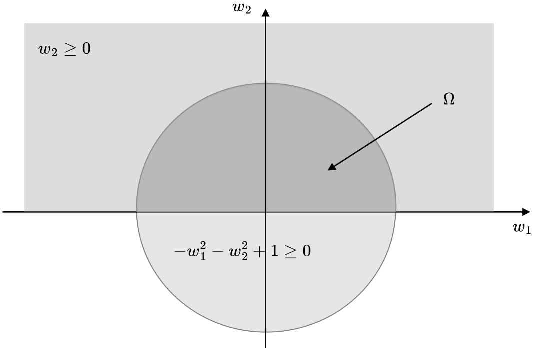
- $w \in \mathbb{R}^n$ - vector of decision variables
- $F : \mathbb{R}^n \to \mathbb{R}$ - objective function
- $G : \mathbb{R}^n \to \mathbb{R}^{n_G}$ - equality constraints
- $H : \mathbb{R}^n \to \mathbb{R}^{n_H}$ - inequality constraints

- only in a few special cases a closed form solution exists
- if F, G, H are nonlinear and smooth, we speak of a *nonlinear programming problem (NLP)*
- usually we need iterative algorithms to find an approximate solution
- in NMPC, the problem depends on parameters that change every sampling time

Basic definitions: the feasible set

Definition

The feasible set of the optimization problem (1) is defined as
$\Omega = \{w \in \mathbb{R}^n \mid G(w) = 0, H(w) \geq 0\}$. A point $w \in \Omega$ is is called a feasible point.



In the example, the feasible set is the intersection of the two grey areas (halfspace and circle)

Basic definitions: global and local minimizer

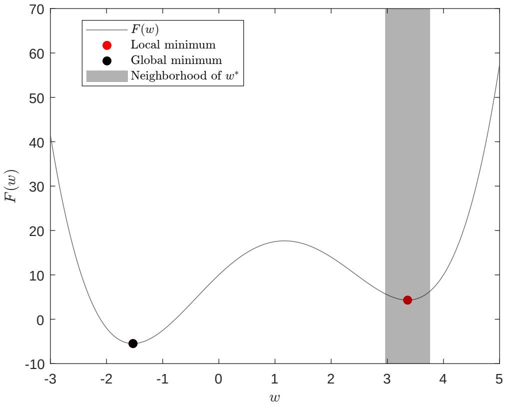
Definition (Global Minimizer)

Point $w^* \in \Omega$ is a **global minimizer** of the NLP (1) if for all $w \in \Omega$ it holds that $F(w) \geq F(w^*)$.

Definition (Local Minimizer)

Point $w^* \in \Omega$ is a **local minimizer** of the NLP (1) if there exists a ball $\mathcal{B}_\epsilon(w^*) = \{w \,|\, \|w - w^*\| \leq \epsilon\}$ with $\epsilon > 0$, such that for all $w \in \mathcal{B}_\epsilon(w^*) \cap \Omega$ it holds that $F(w) \geq F(w^*)$

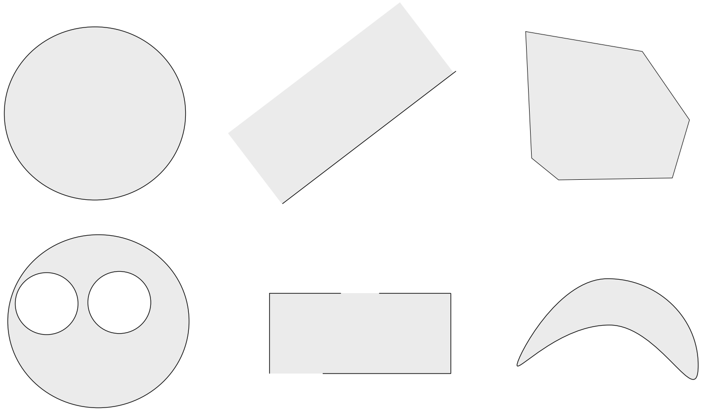
The value $F(w^*)$ at a local/global minimizer w^* is called local/global *minimum*, or *minimum value*.



$$F(w) = \frac{1}{2}w^4 - 2w^3 - 3w^2 + 12w + 10$$

Convex sets

a key concept in optimization



A set Ω is said to be convex if for any w_1, w_2 and any $\theta \in [0, 1]$ it holds $\theta w_1 + (1 - \theta) w_2 \in \Omega$

Figure inspired by Figure 2.2 in S. Boyd and L. Vandenberghe. Convex optimization. Cambridge university press, 2004.

Convex functions

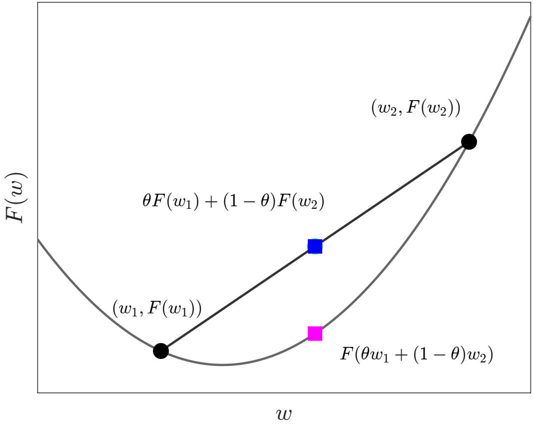
- A function $F : \Omega \to \mathbb{R}$ is convex if for every $w_1, w_2 \in \Omega \subset \mathbb{R}^n$ and $\theta \in [0,1]$ it holds that

  $$F(\theta w_1 + (1-\theta) w_2) \leq \theta F(w_1) + (1-\theta) F(w_2)$$

- F is concave if and only if $-F$ is convex
- F is convex if and only if the epigraph

  $$\text{epi}F = \{(w,t) \in \mathbb{R}^{n_w+1} \mid w \in \Omega, F(w) \leq t\}$$

  is a convex set



Convex optimization problems

A convex optimization problem

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

An optimization problem is convex if the objective function F is convex and the feasible set Ω is convex.

- For convex problems, every locally optimal solution is globally optimal
- First order conditions are necessary and sufficient
- *"...in fact, the great watershed in optimization isn't between linearity and nonlinearity, but convexity and nonconvexity."* R. T. Rockafellar, SIAM Review, 1993

Outline of the lecture

Some classification of optimization problems

Optimization problems can be:

- unconstrained ($\Omega = \mathbb{R}^n$) or constrained ($\Omega \subset \mathbb{R}^n$)
- convex or nonconvex
- linear or nonlinear
- differentiable or nonsmooth
- continuous or (mixed-)integer
- finite or infinite dimensional

Some classification of optimization problems

Optimization problems can be:

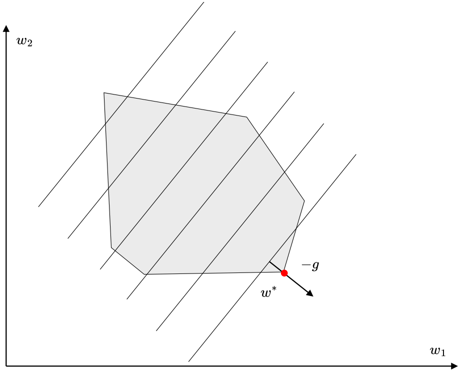
- unconstrained ($\Omega = \mathbb{R}^n$) or constrained ($\Omega \subset \mathbb{R}^n$)
- convex or nonconvex
- linear or nonlinear
- differentiable or nonsmooth
- continuous or (mixed-)integer
- finite or infinite dimensional

"... the main fact, which should be known to any person dealing with optimization models, is that in general, optimization problems are unsolvable."
Yurii Nesterov, Lectures on Convex Optimization, 2018.
("solvable" refers to finding a global minimizer)

Class 1: Linear Programming (LP)

Linear program

$$\min_{w \in \mathbb{R}^n} \; g^\top w$$
$$\text{s.t.} \quad Aw - b = 0$$
$$Cw - d \geq 0$$

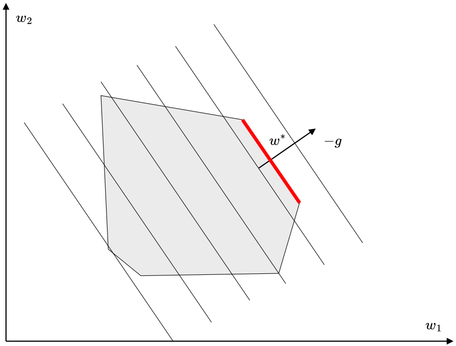


- convex optimization problem
- 1947: simplex method by G. Dantzig
- a solution is always at a vertex of the feasible set (possibly a whole facet if nonunique)
- very mature and reliable

Class 1: Linear Programming (LP)

Linear program

$$\min_{w \in \mathbb{R}^n} \; g^\top w$$
$$\text{s.t.} \quad Aw - b = 0$$
$$Cw - d \geq 0$$

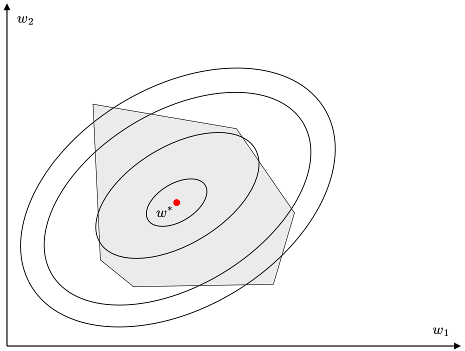


- convex optimization problem
- 1947: simplex method by G. Dantzig
- a solution is always at a vertex of the feasible set (possibly a whole facet if nonunique)
- very mature and reliable

Class 2: Quadratic Programming (QP)

Quadratic Program (QP)

$$\min_{w \in \mathbb{R}^n} \quad \frac{1}{2} w^\top Q w + g^\top w$$
$$\text{s.t.} \quad Aw - b = 0$$
$$Cw - d \geq 0$$

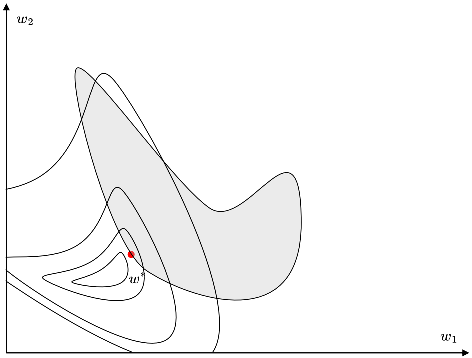


- depending on Q, can be convex and nonconvex
- solved online in linear model predictive control
- many good solvers: `Gurobi`, `OSQP`, `HPIPM`, `qpOASES`, `OOQP`, `DAQP`...
- subsproblems in nonlinear optimization

Class 3: Nonlinear Programming (NLP)

Nonlinear Rrogram (NLP)

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

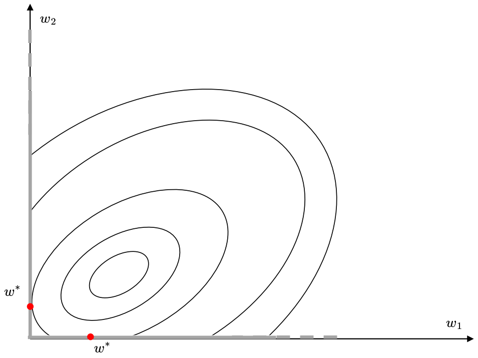


- can be convex and nonconvex
- solved with iterative Newton-type algorithms
- solved in nonlinear model predictive control

MPCC

$$
\begin{aligned}
\min_{w_0, w_1, w_2} \quad & F(w) \\
\text{s.t.} \quad & G(w) = 0 \\
& H(w) \geq 0 \\
& 0 \leq w_1 \perp w_2 \geq 0
\end{aligned}
$$

$$w = [w_0^\top, w_1^\top, w_2^\top]^\top, \; w_1 \perp w_1 \Leftrightarrow w_1^\top w_2 = 0$$



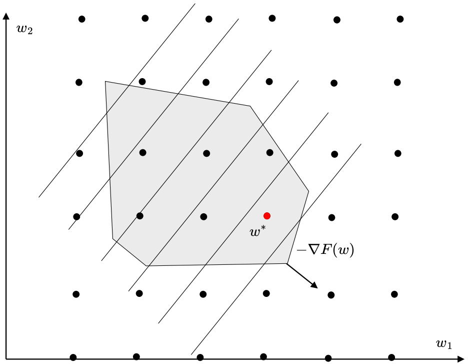
- more difficult than standard nonlinear programming
- feasible set is inherently nonsmooth and nonconvex
- powerful modeling concept
- requires specialized theory and algorithms

Class 5: Mixed-Integer Nonlinear Programming (MINLP)

Mixed-Integer Nonlinear Program (MINLP)

$$\min_{w_0 \in \mathbb{R}^p, w_1 \in \mathbb{Z}^q} \quad F(w)$$
$$\text{s.t.} \quad G(w) = 0$$
$$H(w) \geq 0$$

$w = [w_0^\top, w_1^\top]^\top, n = p + q$



- inherently nonconvex feasible set
- due to combinatorial nature, NP-hard even for linear F, G, H
- branch and bound, branch and cut algorithms based on iterative solution of relaxed continuous problems

Class 6: Continuous-Time Optimal Control

Optimal Control Problem (OCP)

$$\begin{aligned} \min_{x(\cdot),u(\cdot)} \quad & \int_0^T L_c(x(t),u(t))\,\mathrm{d}t + E(x(T)) \\ \text{s.t.} \quad & x(0) = \bar{x}_0 \\ & \dot{x}(t) = f_c(x(t),u(t)) \\ & 0 \geq h(x(t),u(t)),\ t \in [0,T] \\ & 0 \geq r(x(T)) \end{aligned}$$

- decision variables $x(\cdot)$, $u(\cdot)$ in infinite dimensional function space
- infinitely many constraints ($t \in [0,T]$)
- smooth ordinary differential equation (ODE) $\dot{x}(t) = f_c(x(t),u(t))$
- more generally, dynamic model can be based on
  - differential algebraic equations (DAE)
  - partial differential equations (PDE)
  - nonsmooth ODE
  - stochastic ODE
- OCP can be convex or nonconvex
- all or some components of $u(t)$ may take integer values (mixed-integer OCP)

Direct optimal control methods solve Nonlinear Programs (NLP)

Continuous-time OCP

$$
\begin{aligned}
\min_{x(\cdot),u(\cdot)} \quad & \int_0^T L_c(x(t),u(t))\,\mathrm{d}t + E(x(T)) \\
\text{s.t.} \quad x(0) &= \bar{x}_0 \\
\dot{x}(t) &= f_c(x(t),u(t)) \\
0 &\geq h(x(t),u(t)), \ t \in [0,T] \\
0 &\geq r(x(T))
\end{aligned}
$$

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

Direct optimal control methods solve Nonlinear Programs (NLP)

Continuous-time OCP

$$\begin{aligned}\min_{x(\cdot),u(\cdot)} \quad & \int_0^T L_c(x(t),u(t))\,\mathrm{d}t + E(x(T)) \\ \text{s.t.} \quad & x(0) = \bar{x}_0 \\ & \dot{x}(t) = f_c(x(t),u(t)) \\ & 0 \geq h(x(t),u(t)),\ t \in [0,T] \\ & 0 \geq r(x(T))\end{aligned}$$

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

Discrete-time OCP (an NLP)

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

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

Direct optimal control methods solve Nonlinear Programs (NLP)

Discrete-time OCP (an NLP)

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

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

Nonlinear MPC solves Nonlinear Programs (NLP)

Discrete time NMPC Problem (an NLP)

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

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)

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

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)

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

Outline of the lecture

Algebraic characterization of unconstrained local minimizers

Consider the unconstrained problem: $\min_{w \in \mathbb{R}^n} \quad F(w)$

First-Order Necessary Condition of Optimality (FONC) (in convex case also sufficient)

w^* local optimizer $\quad \Rightarrow \quad \nabla F(w^*) = 0, \; w^*$ stationary point

Second-Order Necessary Condition of Optimality (SONC)

w^* local minimizer $\quad \Rightarrow \quad \nabla^2 F(w^*) \succeq 0$

Algebraic characterization of unconstrained local minimizers

Consider the unconstrained problem: $\min_{w \in \mathbb{R}^n} \quad F(w)$

First-Order Necessary Condition of Optimality (FONC) (in convex case also sufficient)

w^* local optimizer $\Rightarrow$ $\nabla F(w^*) = 0$, w^* stationary point

Second-Order Necessary Condition of Optimality (SONC)

w^* local minimizer $\Rightarrow$ $\nabla^2 F(w^*) \succeq 0$

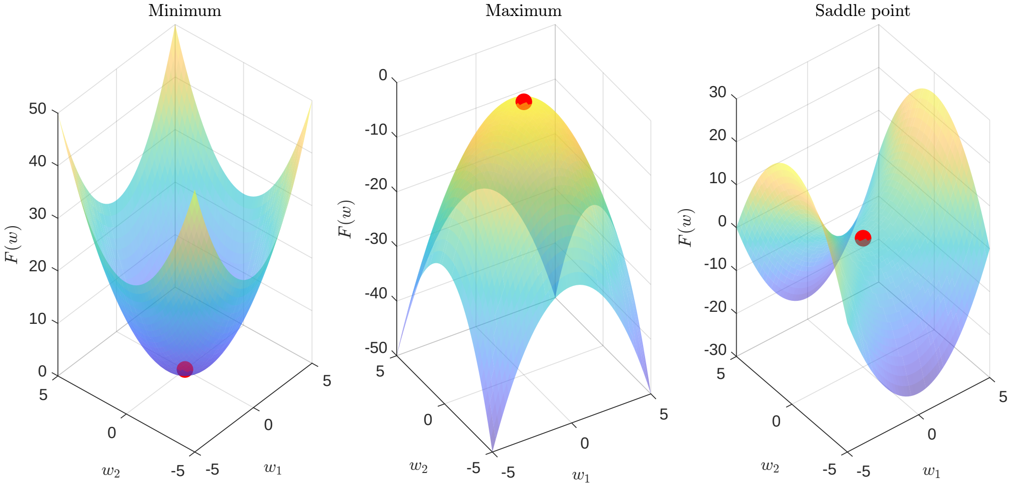
Second-Order Sufficient Conditions of Optimality (SOSC)

$\nabla F(w^*) = 0$ and $\nabla^2 F(w^*) \succ 0$ $\Rightarrow$ x^* strict local minimizer

$\nabla F(w^*) = 0$ and $\nabla^2 F(w^*) \prec 0$ $\Rightarrow$ x^* strict local maximizer

no conclusion can be drawn in the case $\nabla^2 F(w^*)$ is indefinite

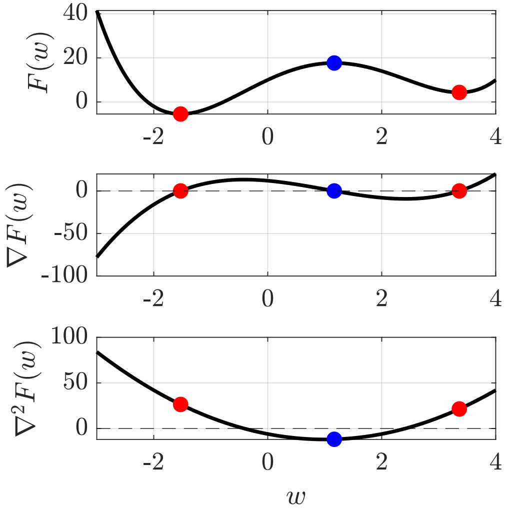
Types of stationary points



a stationary point w with $\nabla F(w) = 0$ can be a minimizer, a maximizer, or a saddle point

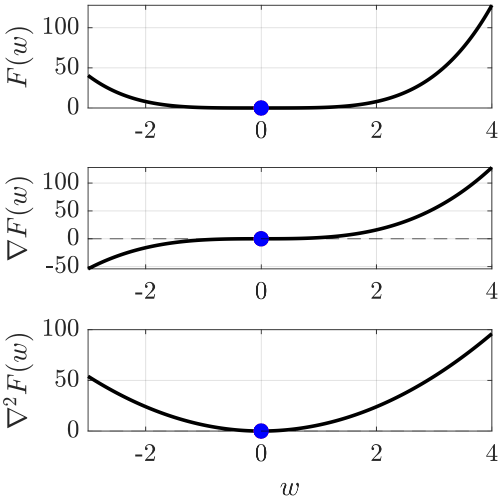
Optimality conditions - unconstrained

- necessary conditions: find a candidate point (or to exclude points)
- sufficient conditions: verify optimality of a candidate point



Optimality conditions - unconstrained

- necessary conditions: find a candidate point (or to exclude points)
- sufficient conditions: verify optimality of a candidate point
- a minimizer must satisfy SONC, but does not have to satisfy SOSC



First order necessary conditions for equality constrained optimization

Nonlinear Program (NLP)

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

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

First order necessary conditions for equality constrained optimization

Nonlinear Program (NLP)

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

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

Definition (LICQ)

A point w satisfies *Linear Independence Constraint Qualification (LICQ)* if and only if $\nabla G(w) := \frac{\partial G}{\partial w}(w)^\top$ is full column rank

First order necessary conditions for equality constrained optimization

Nonlinear Program (NLP)

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

Definition (LICQ)

A point w satisfies *Linear Independence Constraint Qualification (LICQ)* if and only if $\nabla G(w) := \frac{\partial G}{\partial w}(w)^\top$ is full column rank

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

First-Order Necessary Conditions (in convex case also sufficient)

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:

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

Duality in a nutshell

for equality constrained optimization

Primal Problem

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

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

Lagrange dual function $\mathcal{Q}(\lambda) := \inf_{w \in \mathbb{R}^n} \mathcal{L}(w, \lambda)$

- $\mathcal{Q}(\lambda)$ - concave in λ by construction
- $\mathcal{Q}(\lambda) \leq p^*$ for all $\lambda \in \mathbb{R}^{n_G}$

Duality in a nutshell

for equality constrained optimization

Primal Problem

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

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

Lagrange dual function $\mathcal{Q}(\lambda) := \inf_{w \in \mathbb{R}^n} \mathcal{L}(w, \lambda)$

- $\mathcal{Q}(\lambda)$ - concave in λ by construction
- $\mathcal{Q}(\lambda) \leq p^*$ for all $\lambda \in \mathbb{R}^{n_G}$

Dual Problem

$$d^* = \max_{\lambda \in \mathbb{R}^{n_G}} \mathcal{Q}(\lambda)$$

- weak duality: $d^* \leq p^*$, always holds
- strong duality: $d^* = p^*$, only holds for some problems (e.g. convex ones)

Duality in a nutshell

for equality constrained optimization

Primal Problem

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

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

Lagrange dual function $\mathcal{Q}(\lambda) := \inf_{w \in \mathbb{R}^n} \mathcal{L}(w, \lambda)$

- $\mathcal{Q}(\lambda)$ - concave in λ by construction
- $\mathcal{Q}(\lambda) \leq p^*$ for all $\lambda \in \mathbb{R}^{n_G}$

Dual Problem

$$d^* = \max_{\lambda \in \mathbb{R}^{n_G}} \mathcal{Q}(\lambda)$$

- weak duality: $d^* \leq p^*$, always holds
- strong duality: $d^* = p^*$, only holds for some problems (e.g. convex ones)

Wolfe Dual (in convex case)

$$d^* = \max_{w \in \mathbb{R}^n, \lambda \in \mathbb{R}^{n_G}} \mathcal{L}(w, \lambda)$$
$$\text{s.t. } \nabla_w \mathcal{L}(w, \lambda) = 0$$

(w constrained by lower level optimality)

The Karush-Kuhn-Tucker (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 Karush-Kuhn-Tucker (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

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

is full column rank

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

The Karush-Kuhn-Tucker (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

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

is full column rank

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

Theorem (KKT conditions - FONC for constrained optimization)

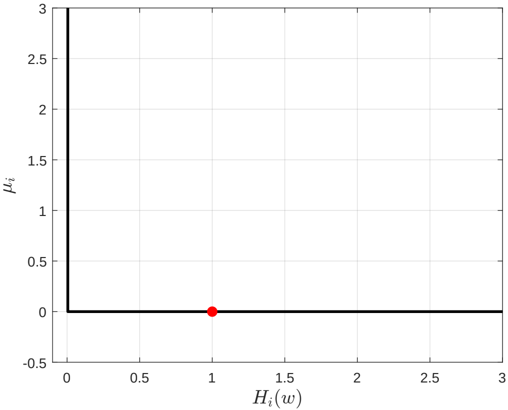
Let F, G, H be $\mathcal{C}^1$. If w^* is a (local) minimizer and satisfies LICQ, then there are unique vectors λ^* and μ^* such that (w^*, λ^*, μ^*) satisfies:

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

Complementarity Conditions

Complementarity conditions $0 \geq \mu \perp H(w) \geq 0$ form an L-shaped, nonsmooth manifold.

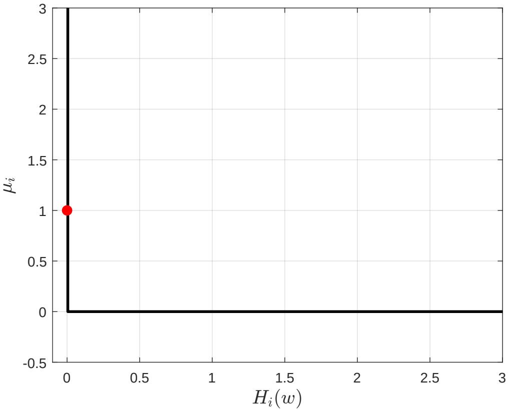
- $H_i(w^*) > 0$ then $\mu_i^* = 0$, and H_i is inactive



Complementarity Conditions

Complementarity conditions $0 \geq \mu \perp H(w) \geq 0$ form an L-shaped, nonsmooth manifold.

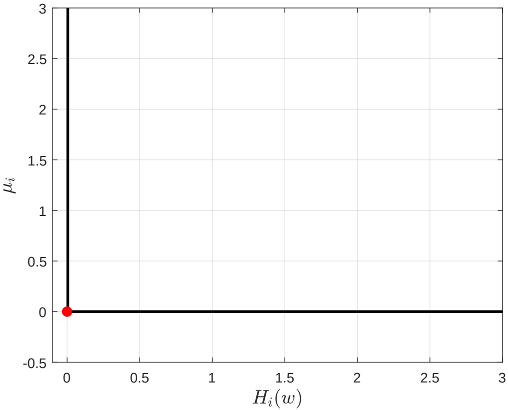
- $H_i(w^*) > 0$ then $\mu_i^* = 0$, and H_i is inactive
- $\mu_i^* > 0$ and $H_i(w) = 0$ then $H_i(w)$ is strictly active



Complementarity Conditions

Complementarity conditions
$0 \geq \mu \perp H(w) \geq 0$ form an L-shaped, nonsmooth manifold.

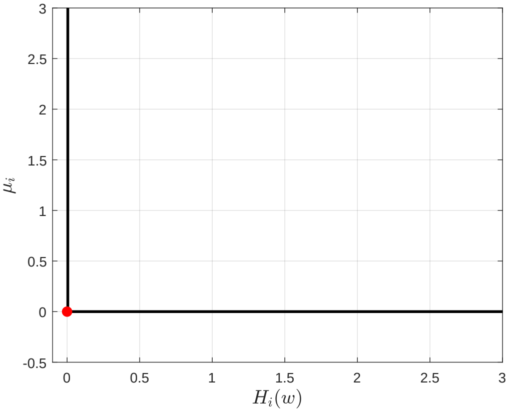
- $H_i(w^*) > 0$ then $\mu_i^* = 0$, and H_i is inactive
- $\mu_i^* > 0$ and $H_i(w) = 0$ then $H_i(w)$ is strictly active
- $\mu_i^* = 0$ and $H_i(w) = 0$ then then $H_i(w)$ is weakly active



Complementarity Conditions

Complementarity conditions $0 \geq \mu \perp H(w) \geq 0$ form an L-shaped, nonsmooth manifold.

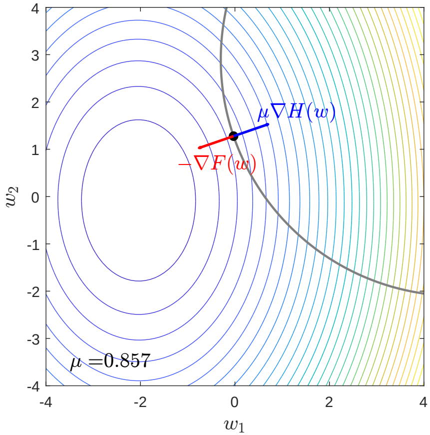
- $H_i(w^*) > 0$ then $\mu_i^* = 0$, and H_i is inactive
- $\mu_i^* > 0$ and $H_i(w) = 0$ then $H_i(w)$ is strictly active
- $\mu_i^* = 0$ and $H_i(w) = 0$ then then $H_i(w)$ is weakly active
- We define the **active set** $\mathbb{A}$ as the set of indices i of the active constraints



Some intuition on the KKT conditions

Ball rolling down a valley blocked by a fence - test problem with two variables and one inequality constraint

$$\min_{w \in \mathbb{R}^2} \; F(w)$$
$$\text{s.t.} \; H(w) \geq 0$$



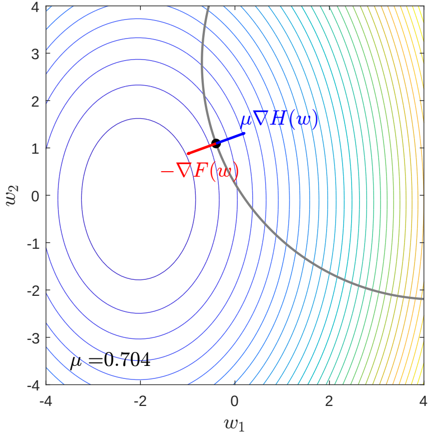
Animation inspired by Lecture 2 of the Winter School on Numerical Optimal Control with Differential Algebraic Equations by S. Gros and M. Diehl, Freiburg, 2016.

Some intuition on the KKT conditions

Ball rolling down a valley blocked by a fence - test problem with two variables and one inequality constraint

$$\min_{w \in \mathbb{R}^2} \ F(w)$$
$$\text{s.t. } H(w) \geq 0$$

- $-\nabla F$ is the gravity



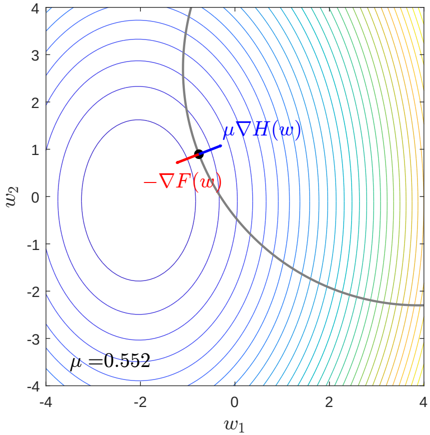
Animation inspired by Lecture 2 of the Winter School on Numerical Optimal Control with Differential Algebraic Equations by S. Gros and M. Diehl, Freiburg, 2016.

Some intuition on the KKT conditions

Ball rolling down a valley blocked by a fence - test problem with two variables and one inequality constraint

$$\min_{w \in \mathbb{R}^2} \; F(w)$$
$$\text{s.t. } H(w) \geq 0$$

- $-\nabla F$ is the gravity



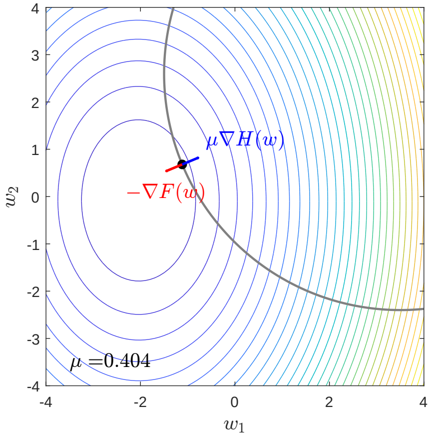
Animation inspired by Lecture 2 of the Winter School on Numerical Optimal Control with Differential Algebraic Equations by S. Gros and M. Diehl, Freiburg, 2016.

Some intuition on the KKT conditions

Ball rolling down a valley blocked by a fence - test problem with two variables and one inequality constraint

$$\min_{w \in \mathbb{R}^2} \; F(w)$$
$$\text{s.t. } H(w) \geq 0$$

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



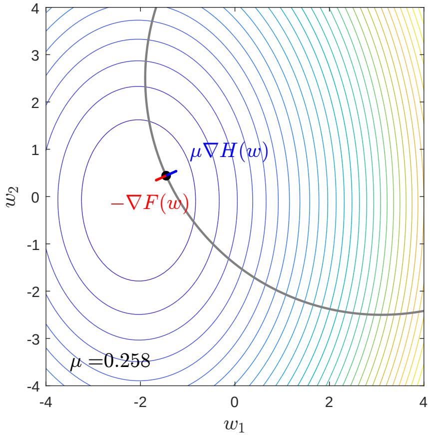
Animation inspired by Lecture 2 of the Winter School on Numerical Optimal Control with Differential Algebraic Equations by S. Gros and M. Diehl, Freiburg, 2016.

Some intuition on the KKT conditions

Ball rolling down a valley blocked by a fence - test problem with two variables and one inequality constraint

$$\min_{w \in \mathbb{R}^2} \; F(w)$$
$$\text{s.t. } H(w) \geq 0$$

- $-\nabla F$ 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



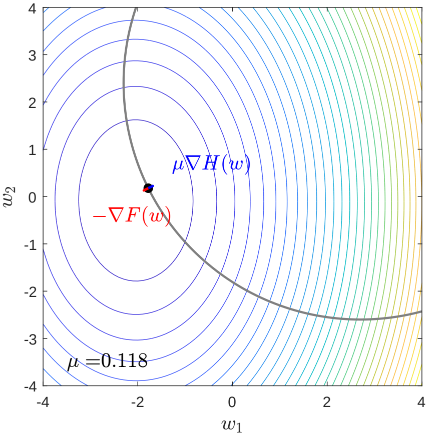
Animation inspired by Lecture 2 of the Winter School on Numerical Optimal Control with Differential Algebraic Equations by S. Gros and M. Diehl, Freiburg, 2016.

Some intuition on the KKT conditions

Ball rolling down a valley blocked by a fence - test problem with two variables and one inequality constraint

$$\min_{w \in \mathbb{R}^2} \; F(w)$$
$$\text{s.t. } H(w) \geq 0$$

- $-\nabla F$ 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



Animation inspired by Lecture 2 of the Winter School on Numerical Optimal Control with Differential Algebraic Equations by S. Gros and M. Diehl, Freiburg, 2016.

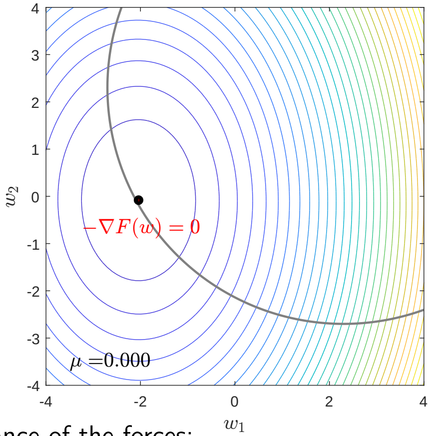
Some intuition on the KKT conditions

Ball rolling down a valley blocked by a fence - test problem with two variables and one inequality constraint

$$\min_{w \in \mathbb{R}^2} \; F(w)$$

$$\text{s.t. } H(w) \geq 0$$

- $-\nabla F$ 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:

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

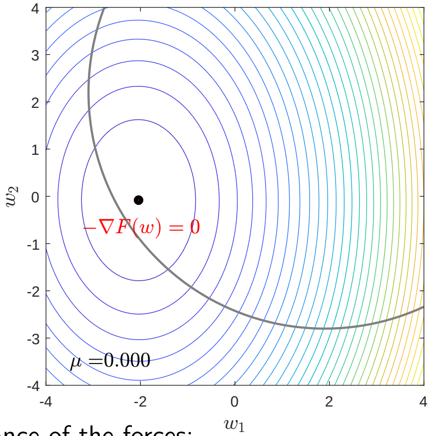
Animation inspired by Lecture 2 of the Winter School on Numerical Optimal Control with Differential Algebraic Equations by S. Gros and M. Diehl, Freiburg, 2016.

Ball rolling down a valley blocked by a fence - test problem with two variables and one inequality constraint

$$\min_{w \in \mathbb{R}^2} \; F(w)$$
$$\text{s.t. } H(w) \geq 0$$

- $-\nabla F$ 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
- inactive constraint $H(w) > 0, \;\; \mu = 0$

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



Balance of the forces:

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

Animation inspired by Lecture 2 of the Winter School on Numerical Optimal Control with Differential Algebraic Equations by S. Gros and M. Diehl, Freiburg, 2016.

Outline of the lecture

Newton's method

To solve a nonlinear system, solve a sequence of linear systems

Linearization of F at linearization point $\bar{w}$

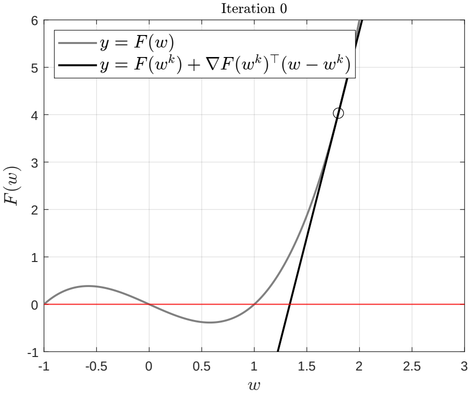
equals

First order Taylor series at $\bar{w}$

equals

$$F_{\mathrm{L}}(w;\bar{w}) := F(\bar{w}) + \frac{\partial F}{\partial w}(\bar{w}) \quad (w - \bar{w})$$

(for continuously differentiable $F : \mathbb{R}^n \to \mathbb{R}^n$)



Linearization of F at linearization point $\bar{w}$

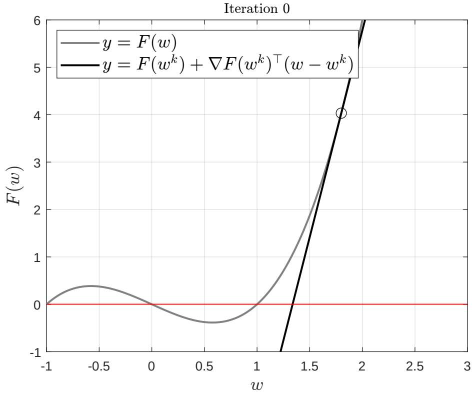
equals

First order Taylor series at $\bar{w}$

equals

$$F_{\mathrm{L}}(w;\bar{w}) := F(\bar{w}) + \nabla_w F(\bar{w})^\top (w - \bar{w})$$

(for continuously differentiable $F : \mathbb{R}^n \to \mathbb{R}^n$)



Newton's method

To solve a nonlinear system, solve a sequence of linear systems

Linearization of F at linearization point $\bar{w}$

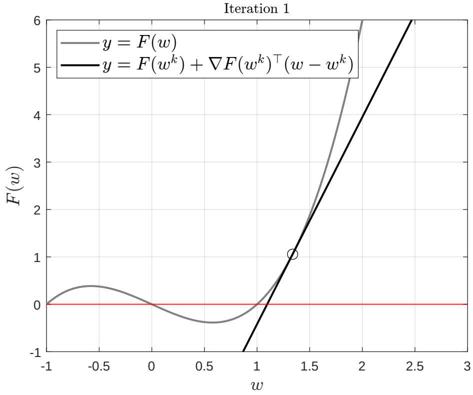
equals

First order Taylor series at $\bar{w}$

equals

$$F_{\mathrm{L}}(w;\bar{w}) := F(\bar{w}) + \nabla_w F(\bar{w})^\top (w - \bar{w})$$

(for continuously differentiable $F : \mathbb{R}^n \to \mathbb{R}^n$)



Newton's method

To solve a nonlinear system, solve a sequence of linear systems

Linearization of F at linearization point $\bar{w}$

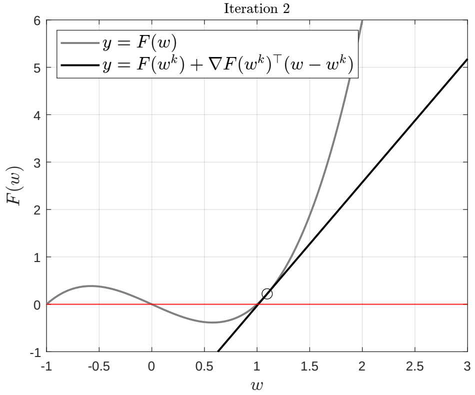
equals

First order Taylor series at $\bar{w}$

equals

$$F_{\mathrm{L}}(w;\bar{w}) := F(\bar{w}) + \nabla_w F(\bar{w})^\top (w - \bar{w})$$

(for continuously differentiable $F : \mathbb{R}^n \to \mathbb{R}^n$)



Newton's method

To solve a nonlinear system, solve a sequence of linear systems

Linearization of F at linearization point $\bar{w}$

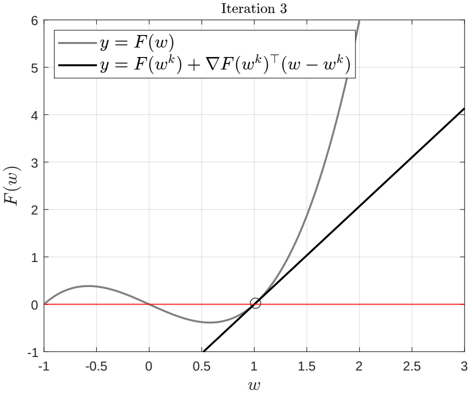
equals

First order Taylor series at $\bar{w}$

equals

$$F_{\mathrm{L}}(w;\bar{w}) := F(\bar{w}) + \nabla_w F(\bar{w})^\top (w - \bar{w})$$

(for continuously differentiable $F : \mathbb{R}^n \to \mathbb{R}^n$)



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)

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

We first treat the case without inequalities

NLP only with equality constraints

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

Lagrange function and optimality conditions

Lagrange function

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

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

Newton's method to solve

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

results, at iterate (w^k, λ^k), in the following linear system:

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

Due to $\nabla \mathcal{L}(w^k, \lambda^k) = \nabla F(w^k) - \nabla G(w^k)\lambda^k$ this is equivalent to

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

with the shorthand $\lambda^+ = \lambda^k + \Delta\lambda$

Newton Step = Quadratic Program

Conditions

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

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

Quadratic program

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

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

Newton's method

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

Quadratic Program in Sequential Quadratic Programming (SQP)

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

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

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

This is the "full step, exact Hessian SQP method for equality constrained optimization".

NLP with Inequalities

Regard again NLP with both, equalities and inequalities:

NLP with equality and inequality constraints

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

Lagrangian function for NLP with equality and inequality constraints

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

Recall necessary 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:

$$\begin{aligned}
&\nabla_w \mathcal{L}\left(w^*, \mu^*, \lambda^*\right) = 0 \\
&G\left(w^*\right) = 0 \\
&H(w^*) \geq 0 \\
&\mu^* \geq 0 \\
&H(w^*)^\top \mu^* = 0
\end{aligned}$$

- Last three "complementarity conditions" are nonsmooth
- Thus, this system cannot be solved by Newton's Method. But still with SQP...

Sequential Quadratic Programming (SQP) with Inequalities

By linearizing all functions and setting $\lambda^+ = \lambda^k + \Delta\lambda$, $\mu^+ = \mu^k + \Delta\mu$, we obtain the KKT conditions of the following Quadratic Program (QP)

Inequality Constrained Quadratic Program within SQP method

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

with

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

Its solution delivers the next SQP iterate

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

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.

Gauss-Newton QP = Constrained Linear Least Squares Problem

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

Linear convergence. Fast, if objective value $\|R(w^*)\|$ small or nonlinearity of R, G, H small

Interior Point Methods

(without equalities for simplicity of exposition)

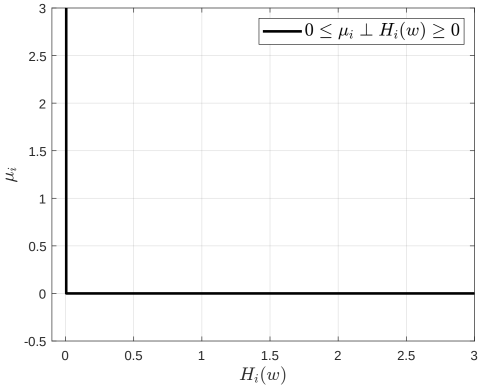
NLP with inequalites

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

KKT conditions

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

Main difficulty: nonsmoothness of complementarity conditions



Barrier Problem in Interior Point Method

NLP with inequalites

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

Idea: put inequality constraint into objective

Barrier Problem in Interior Point Method

NLP with inequalites

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

Idea: put inequality constraint into objective

Barrier Problem

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

Barrier Problem in Interior Point Method

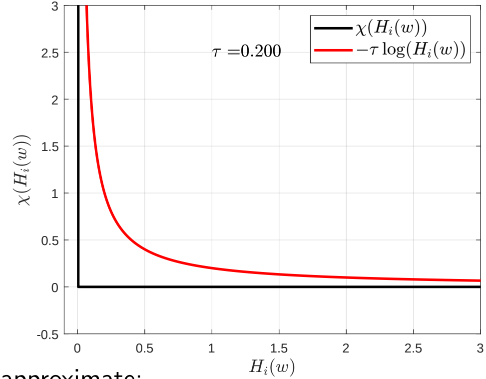
NLP with inequalites

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

Idea: put inequality constraint into objective

Barrier Problem

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



approximate:

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

Barrier Problem in Interior Point Method

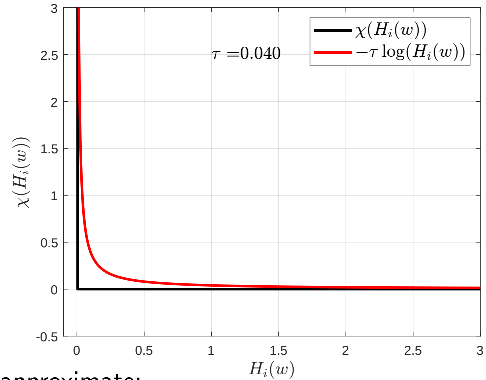
NLP with inequalites

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

Idea: put inequality constraint into objective

Barrier Problem

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



approximate:

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

Barrier Problem in Interior Point Method

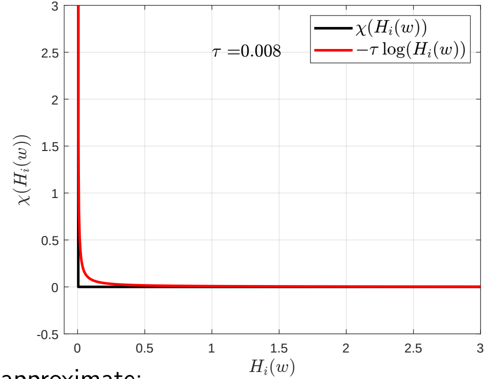
NLP with inequalites

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

Idea: put inequality constraint into objective

Barrier Problem

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



approximate:

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

Barrier Problem in Interior Point Method

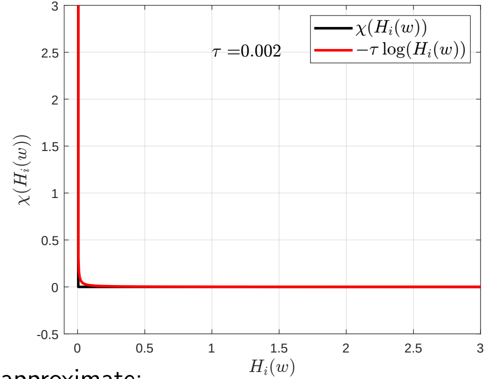
NLP with inequalites

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

Idea: put inequality constraint into objective

Barrier Problem

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



approximate:

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

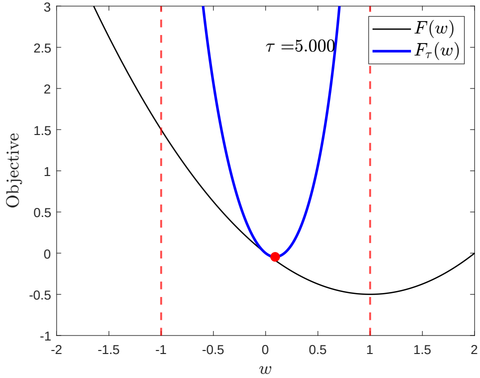
Example Barrier Problem

Example NLP

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

Barrier problem

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



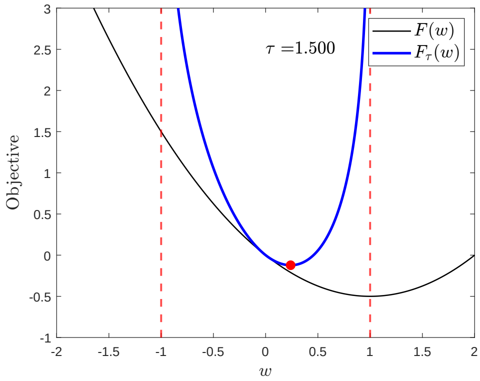
Example Barrier Problem

Example NLP

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

Barrier problem

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



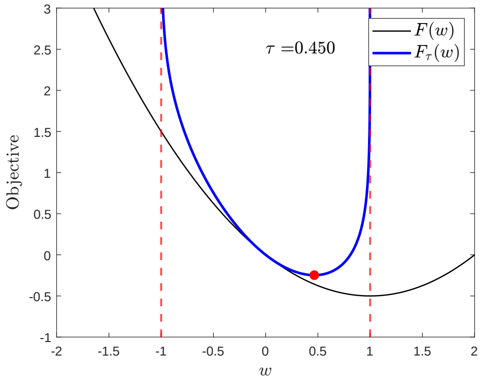
Example Barrier Problem

Example NLP

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

Barrier problem

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



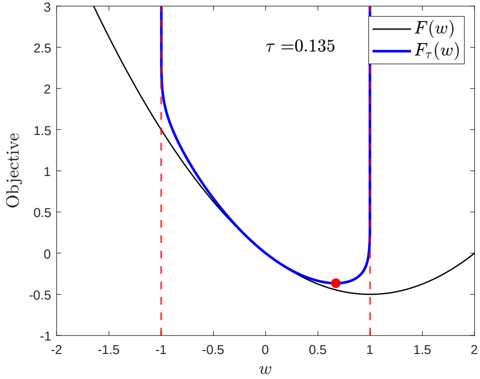
Example Barrier Problem

Example NLP

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

Barrier problem

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



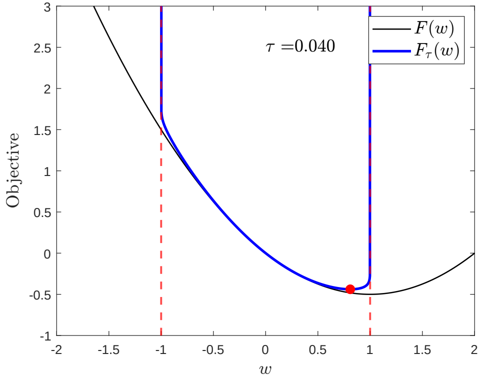
Example Barrier Problem

Example NLP

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

Barrier problem

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



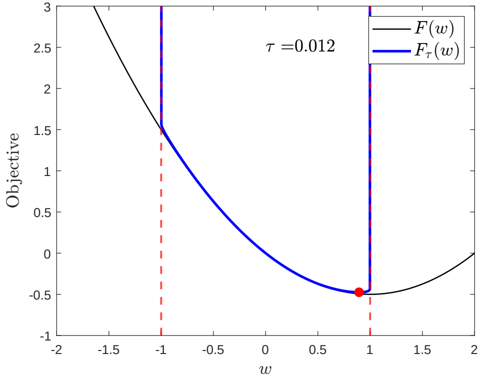
Example Barrier Problem

Example NLP

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

Barrier problem

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



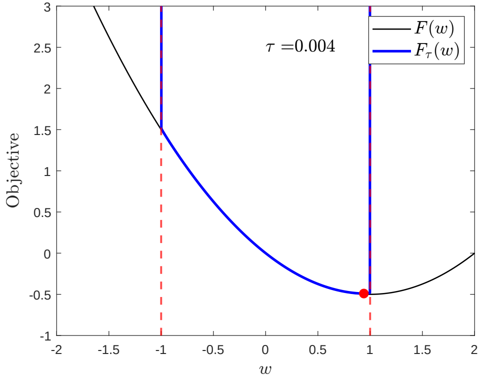
Example Barrier Problem

Example NLP

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

Barrier problem

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



Barrier problem

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

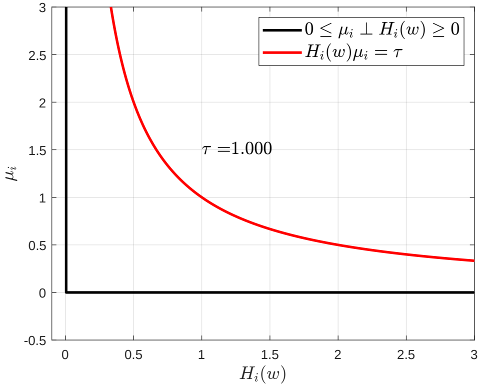
KKT conditions

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

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

Smoothed KKT conditions

$$\begin{aligned} &\nabla F(w) - \nabla H(w)^\top \mu = 0 \\ &H_i(w)\mu_i = \tau \\ &(H_i(w) > 0, \mu_i > 0) \end{aligned}$$



Primal-dual interior point methods

Alternative interpretation

Barrier problem

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

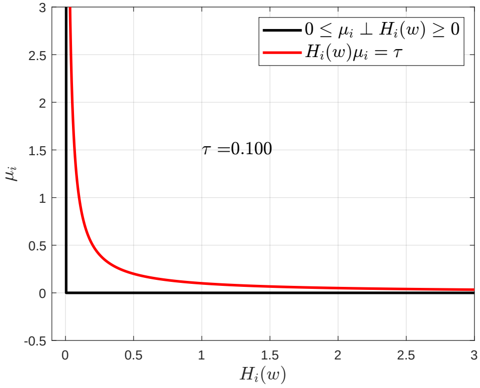
KKT conditions

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

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

Smoothed KKT conditions

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



Barrier problem

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

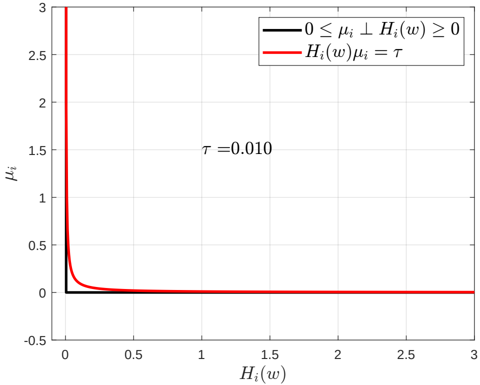
KKT conditions

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

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

Smoothed KKT conditions

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



Primal-dual interior point methods
Alternative interpretation

Barrier problem

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

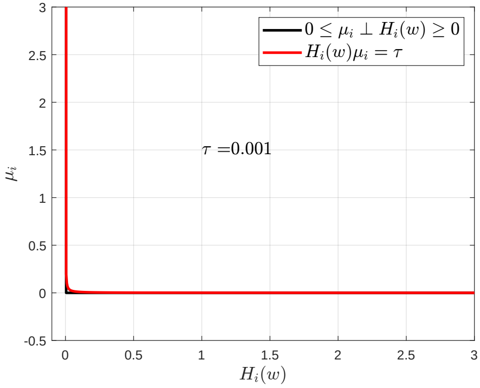
KKT conditions

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

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

Smoothed KKT conditions

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



Primal-dual interior point method

Nonlinear programming problem

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

Smoothed KKT conditions

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

$(s, \mu > 0)$

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

Fix τ, perform Newton iterations

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

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

Line-search

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

$$
\begin{aligned}
w^{k+1} &= w^k + \alpha \Delta w \\
s^{k+1} &= s^k + \alpha \Delta s \\
\lambda^{k+1} &= \lambda^k + \alpha \Delta \lambda \\
\mu^{k+1} &= \mu^k + \alpha \Delta \mu
\end{aligned}
$$

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

Reduce τ, and perform next Newton iterations solve, etc

Summary Nonlinear Optimization

- optimization problem come in many variants (LP, QP, NLP, MPCC, MINLP, OCP, ...)
- each problem class be addressed with suitable software
- nonlinear MPC needs to solve nonlinear programs (NLP)
- Lagrangian function, duality, and KKT conditions are important concepts
- for convex problems holds strong duality, KKT conditions sufficient for global optimality
- Newton-type optimization for NLP solves the nonsmooth KKT conditions via Sequential Quadratic Programming (SQP, e.g. acados) or via Interior Point Method (e.g. ipopt)
- NLP solvers need to evaluate first and second order derivatives (e.g. via CasADi)

Some References

- Jorge Nocedal, Stephen J. Wright, Numerical optimization. Springer, 2006.
- S. Boyd, L. Vandenberghe, Convex Optimization, Cambridge University Press, 2004
- Moritz Diehl, Sébastien Gros, Numerical Optimal Control (Draft), 2024
- Moritz Diehl, Lecture Notes on Numerical Optimization (Draft), 2017