Model Predictive Control and Reinforcement Learning
– Lecture 3: Numerical Optimization –

Joschka Boedecker and Moritz Diehl

University Freiburg

July 26, 2021

Acknowledgement

Slides contain some figures provided by W. Bangerth and K. Mombaur.

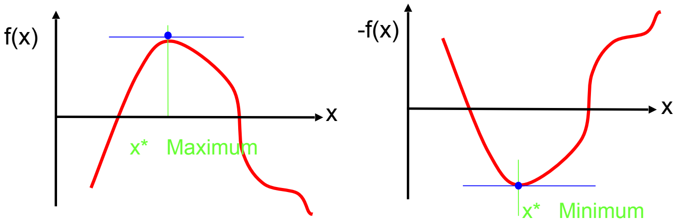
Overview

Overview

What is optimization?

- Optimization = search for the best solution
- in mathematical terms:
  minimization or maximization of an objective function f (x)
  depending on variables x subject to constraints

Equivalence of maximization and minimization problems:
(from now on only minimization)



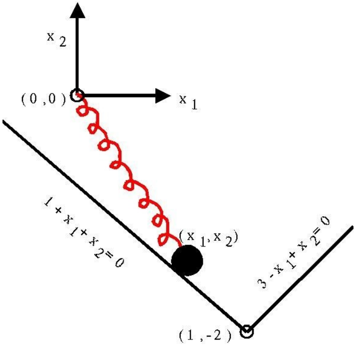
Constrained optimization

- Often variable x shall satisfy certain constraints, e.g.:
  - $x \geq)$
  - $x_1{}^2 + x_2{}^2 = C$
- General formulation:

$$
\begin{aligned}
&\min f(x) \\
&\text{subject to (s.t.)} \\
&\quad g(x) = 0 \\
&\quad h(x) \geq 0
\end{aligned}
$$

f objective function / cost function
g equality constraints
h inequality constraints

Simple example: Ball hanging on a spring



To find position at rest, minimize potential energy!

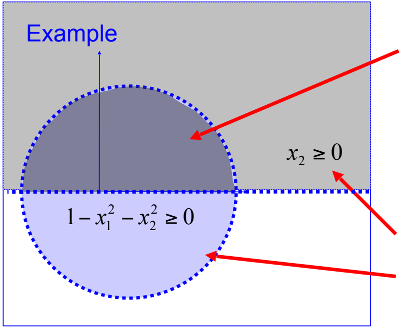
$$\min \underbrace{x_1^2 + x_2^2}_{\text{spring}} + \underbrace{mx_2}_{\text{gravity}}$$

$$1 + x_1 + x_2 \geq 0$$

$$3 - x_1 + x_2 \geq 0$$

Feasible set

Feasible set = collection of all points that satisfy all constraints:



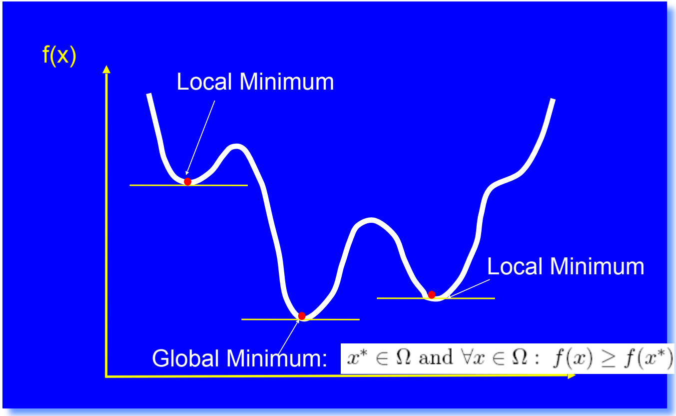
feasible set is intersection of grey and blue area

$$h_1(x) := \qquad x_2 \geq 0$$

$$h_2(x) := \quad 1 - x_1^2 - x_2^2 \geq 0$$

The "feasible set" Ω is $\{x \in \mathbb{R}^n | g(x) = 0,\ h(x) \geq 0\}$.

Local and global minima



The point $x^* \in \mathbb{R}^n$ is a "local minimizer" iff $x^* \in \Omega$ and there exists a neighborhood $\mathcal{N}$ of x^* (e.g an open ball around x^*) so that $\forall x \in \Omega \cap \mathcal{N}: \; f(x) \geq f(x^*)$.

Derivatives

- First and second derivatives of the objective function or the constraints play an important role in optimization

- The first order derivatives are called the gradient (of the resp. fct)

$$\nabla f(x) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, ..., \frac{\partial f}{\partial x_n}\right)^T$$

- and the second order derivatives are called the Hessian matrix

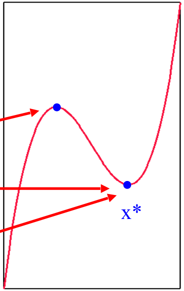
$$\nabla^2 f(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial {x_1}^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial {x_2}^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial {x_n}^2} \end{pmatrix}$$

Optimality conditions (unconstrained)

$$\boxed{\min_{x \in \mathbb{R}^n} f(x)}$$

Assume that f is twice differentiable. We want to test a point *x** for local optimality.

- *necessary condition:*
  $\nabla f(x^*)=0$ (*stationarity*)
- *sufficient condition:*
  *x** stationary and $\nabla^2 f(x^*)$ positive definite

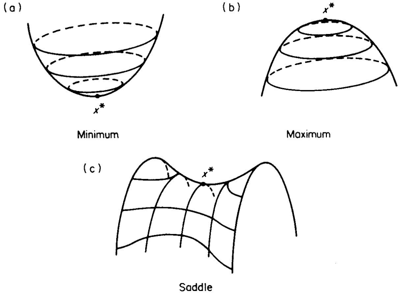


Types of stationary points

(a)-(c) x^* is stationary: $\nabla f(x^*)=0$

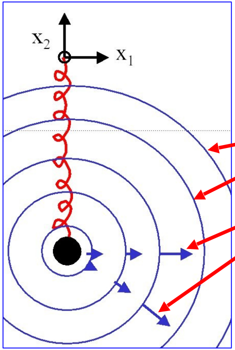
$\nabla^2 f(x^*)$ positive definite: local minimum

$\nabla^2 f(x^*)$ negative definite: local maximum



$\nabla^2 f(x^*)$ indefinite: saddle point

Ball on a spring without constraints



$$\min_{x \in R^2} x_1^2 + x_2^2 + m x_2$$

contour lines of *f(x)*

gradient vector

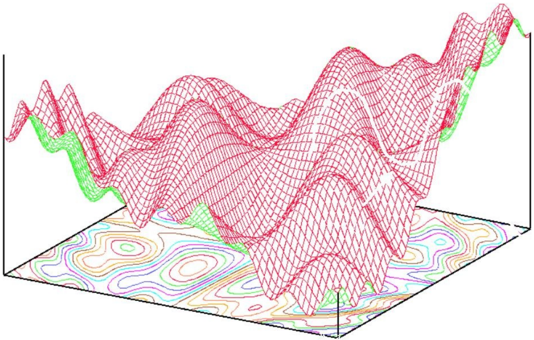
$$\nabla f(x) = (2x_1, 2x_2 + m)$$

unconstrained minimum:

$$0 = \nabla f(x^*) \Leftrightarrow (x_1^*, x_2^*) = (0, -\frac{m}{2})$$

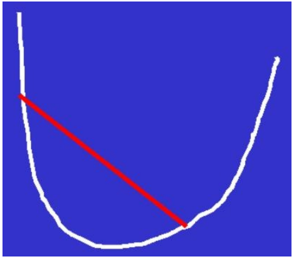
Sometimes there are many local minima

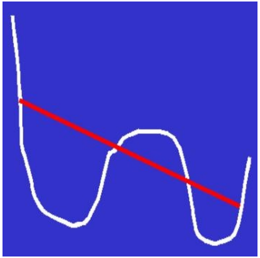
e.g. potential energy
of macromolecule



Global optimization is a very hard issue - most algorithms find only the next local minimum. But there is a favourable special case...

Convex functions



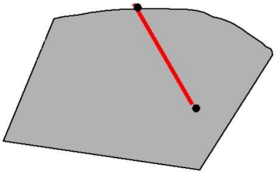


Convex: all connecting lines are above graph

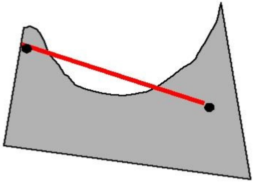
Non-convex: some connecting lines are not above graph

$$\boxed{f : \Omega \to \mathbb{R} \text{ convex} \quad \Leftrightarrow \quad \forall x, y \in \Omega, t \in [0,1] : f(x + t(y-x)) \le f(x) + t(f(y) - f(x))}$$

Convex feasible sets



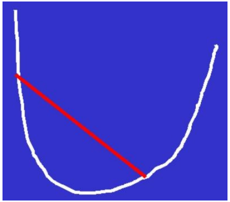
Convex: all connecting lines between feasible points are in the feasible set

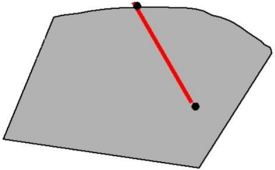


Non-convex: some connecting line between two feasible points is not in the feasible set

$$\boxed{\Omega \text{ convex} \quad \Leftrightarrow \quad \forall x, y \in \Omega, t \in [0,1] : x + t(y - x) \in \Omega}$$

Convex optimization problems





Convex problem if

f(x) is convex **<u>and</u>** the feasible set is convex

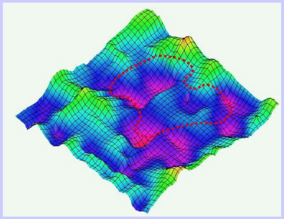
One can show:
For convex problems, every local minimum is also a global minimum.
It is sufficient to find local minima!

Characteristics of optimization problems 1

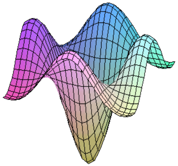
- size / dimension of problem n ,
  i.e. number of free variables

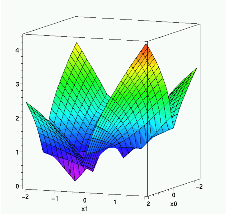
- continuous or discrete search space

- number of minima



- Properties of the objective function:
  - type: linear, nonlinear, quadratic ...
  - smoothness: continuity, differentiability
- Existence of constraints
- Properties of constraints:
  - equalities / inequalities
  - type: „simple bounds“, linear, nonlinear, dynamic equations (optimal control)
  - smoothness





Summary Basic Definitions and Concepts

- Optimization problems can be:
  - unconstrained or constrained
  - convex or non-convex
  - linear or non-linear
  - differentiable or non-smooth
  - continuous or (mixed-)integer
  - finite or infinite dimensional
  - ...

Overview

Problem Class 1: Linear Programming (LP)

- Linear objective,
  linear constraints:
  Linear Optimization Problem
  (convex)

$$\begin{aligned} \min_{x} \quad & c^T x \\ \text{s. t.} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

- Example: **Logistics Problem**
  - shipment of quantities $a_1, a_2, \ldots a_m$ of a product from m locations
  - to be received at n detinations in quantities $b_1, b_2, \ldots b_n$
  - shipping costs c_{ij}
  - determine amounts x_{ij}

⇨ Origin of linear programming in 2nd world war

Problem Class 2: Quadratic Programming (QP)

- Quadratic objective and linear constraints:
  Quadratic Optimization Problem (convex, if Q pos. def.)

$$\begin{aligned} \min_{x} \quad & c^T x + \frac{1}{2} x^T Q x \\ \text{s. t.} \quad & Ax = b \\ & Cx \geq d \end{aligned}$$

- Example: Markovitz mean variance portfolio optimization
  - quadratic objective: portfolio variance (sum of the variances and covariances of individual securities)
  - linear constraints specify a lower bound for portfolio return

- QPs play an important role as **subproblems in nonlinear optimization**

Important: Linear MPC is based on online solution of QP for changing data

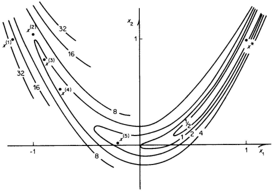
Problem Class 3: Nonlinear Programming (NLP)

- Nonlinear Optimization Problem (in general nonconvex)

$$\begin{aligned} \min_{x} \quad & f(x) \\ \text{s. t.} \quad & h(x) = 0 \\ & g(x) \geq 0 \end{aligned}$$

- E.g. the famous nonlinear Rosenbrock function

$$f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$



Nonlinear MPC is based on online solution of NLP via Newton-type methods

Problem Class 4: Non-smooth optimization

- objective function or constraints are non-differentiable or not continuous e.g.

$$f(x) = |x|$$

$$f(x) = \max_i f_i(x), \quad i = 1, ..n$$

$$f(x) = \begin{cases} \cos x & \text{für } x \leq \frac{\pi}{2} \\ 0 & \text{für } x > \frac{\pi}{2} \end{cases}$$

$$f(x) = i \quad \text{for} \quad i \leq x < i+1, \; i = 0, 1, 2, ...$$

derivative-based methods can still be useful e.g. stochastic gradient descent (SGD) or penalty methods for mathematical programs with complementarity constraints (MPCC)

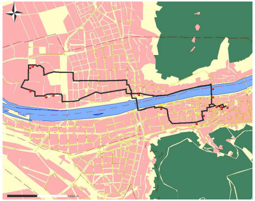
Problem Class 5: (Mixed) Integer Programming (MIP)

- Some or all variables are integer (e.g. linear integer problems)

$$\begin{aligned} \min_{x} \quad & c^T x \\ \text{s. t.} \quad & Ax = b \\ & x \in Z_+^n \end{aligned}$$

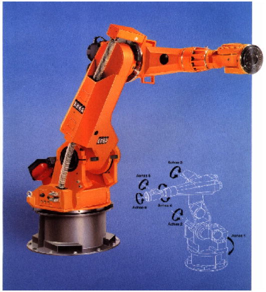
- Special case: combinatorial optimization problems -- feasible set is finite

- Example: traveling salesman problem
  - determine fastest/shortest round trip through n locations



Problem Class 6: Continuous Optimal Control

- Optimization problems including dynamics in form of **differential equations** (infinite dimensional)



Variables $x(t), u(t), p$ (partly ∞-dim.)

$$\min_{,x,u,p} \quad \int_0^T \phi(t, x(t), u(t), p)dt$$

$$\text{s. t.} \quad \dot{x} = f(t, x(t), u(t), p)$$

$$\ldots.$$

THIS COURSE'S MAIN TOPIC!

Nonlinear MPC solves Nonlinear Programs (NLP)

Continuous Time NMPC Problem

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

Direct methods like multiple shooting first discretize, then optimize.

Nonlinear MPC solves Nonlinear Programs (NLP)

Continuous Time NMPC Problem

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

Direct methods like multiple shooting first discretize, then optimize.

Discrete time NMPC Problem (an NLP)

$$\begin{aligned} \min_{s,a} \quad & \sum_{k=0}^{N-1} c(s_k,a_k) + E(s_N) \\ \text{s.t.} \quad & s_0 = \bar{s}_0 \\ & s_{k+1} = f(s_k,a_k) \\ & 0 \geq h(s_k,a_k),\ k = 0,\dots,N-1 \\ & 0 \geq r(s_N) \end{aligned}$$

Variables $s = (s_0,\dots)$ and $a = (a_0,\dots,a_{N-1})$ can be summarized in vector $x = (s,a) \in \mathbb{R}^n$.

Nonlinear MPC solves Nonlinear Programs (NLP)

Discrete time NMPC Problem (an NLP)

$$
\begin{aligned}
\min_{s,a} \quad & \textstyle\sum_{k=0}^{N-1} c(s_k, a_k) + E(s_N) \\
\text{s.t.} \quad & s_0 = \bar{s}_0 \\
& s_{k+1} = f(s_k, a_k) \\
& 0 \geq h(s_k, a_k), \; k = 0, \ldots, N-1 \\
& 0 \geq r(s_N)
\end{aligned}
$$

Variables $s = (s_0, \ldots)$ and $a = (a_0, \ldots, a_{N-1})$ can be summarized in vector $x = (s, a) \in \mathbb{R}^n$.

Nonlinear MPC solves Nonlinear Programs (NLP)

Discrete time NMPC Problem (an NLP)

$$
\begin{aligned}
\min_{s,a} \ & \textstyle\sum_{k=0}^{N-1} c(s_k, a_k) + E(s_N) \\
\text{s.t.} \quad & s_0 = \bar{s}_0 \\
& s_{k+1} = f(s_k, a_k) \\
& 0 \geq h(s_k, a_k), \ k = 0, \dots, N-1 \\
& 0 \geq r(s_N)
\end{aligned}
$$

Variables $s = (s_0, \dots)$ and $a = (a_0, \dots, a_{N-1})$ can be summarized in vector $x = (s, a) \in \mathbb{R}^{n_x}$.

Nonlinear MPC solves Nonlinear Programs (NLP)

Discrete time NMPC Problem (an NLP)

$$
\begin{aligned}
\min_{s,a} \; & \textstyle\sum_{k=0}^{N-1} c(s_k, a_k) + E(s_N) \\
\text{s.t.} \quad & s_0 = \bar{s}_0 \\
& s_{k+1} = f(s_k, a_k) \\
& 0 \geq h(s_k, a_k), \; k = 0, \ldots, N-1 \\
& 0 \geq r(s_N)
\end{aligned}
$$

Variables $s = (s_0, \ldots)$ and $a = (a_0, \ldots, a_{N-1})$ can be summarized in vector $x = (s, a) \in \mathbb{R}^{n_x}$.

Nonlinear Program (NLP)

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

The great watershed

"The great watershed in optimization isn't
between linearity and nonlinearity,
but convexity and nonconvexity"

R. Tyrrell Rockafellar

- For convex optimization problems we can efficiently find global minima.
- For non-convex, but smooth problems we can efficiently find local minima.

Overview

Newton-type methods based on linearizations of nonlinear functions

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

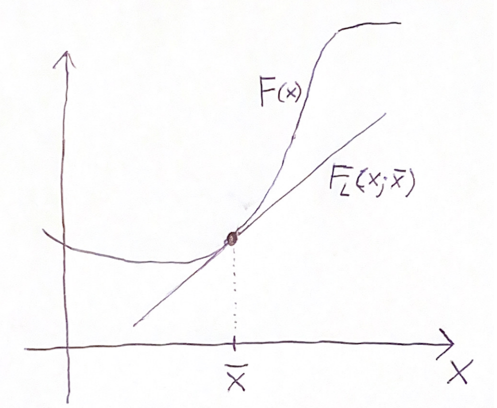
equals

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

equals

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

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



Newton-type methods based on linearizations of nonlinear functions

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

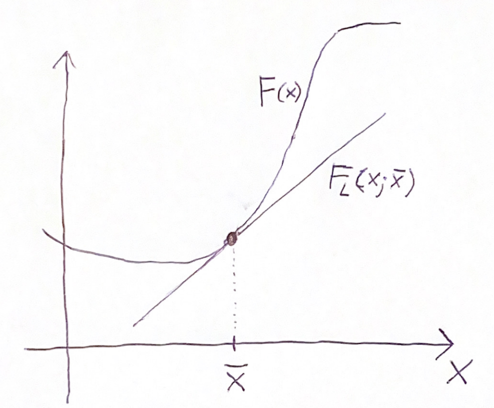
equals

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

equals

$$\boxed{F_{\mathrm{L}}(x;\bar{x}) := F(\bar{x}) + \nabla_x F(\bar{x})^\top (x-\bar{x})}$$

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



Overview

General Nonlinear Program (NLP)

We want to solve the discretized optimal control problem, which is a Nonlinear Program (NLP)

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

We first treat the case **without inequalities**

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

Lagrange Function and Optimality Conditions

Introduce Lagrangian function

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

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

$$\begin{array}{rcl} \nabla_x \mathcal{L}(x^*, \lambda^*) & = & 0, \\ G(x^*) & = & 0, \end{array}$$

Newton's Method on Optimality Conditions

How to solve nonlinear equations

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

Linearize.

$$\begin{array}{llllll} \nabla_x \mathcal{L}(x^k, \lambda^k) & +\nabla_x^2 \mathcal{L}(x^k, \lambda^k)\Delta x & -\nabla_x G(x^k)\Delta\lambda & = & 0, \\ G(x^k) & +\nabla_x G(x^k)^\top \Delta x & & = & 0. \end{array}$$

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

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

Newton Step = Quadratic Program

Conditions

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

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

$$\begin{array}{ll} \min\limits_{\Delta x} & \nabla F(x^k)^\top \Delta x + \frac{1}{2}\Delta x^\top B_k^{\mathrm{ex}} \Delta x \\ \text{s.t.} & G(x^k) + \nabla G(x^k)^\top \Delta x \;=\; 0, \end{array}$$

with

$$B_k^{\mathrm{ex}} = \nabla_x^2 \mathcal{L}(x^k, \lambda^k)$$

Newton's Method

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

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

with $B_k^{\mathrm{ex}} = \nabla_x^2 \mathcal{L}(x^k, \lambda^k)$.

Obtain as QP solution step Δx^k and new multiplier λ_{QP}^+, and iterate:

$$\begin{array}{rcl} x^{k+1} & = & x^k + \Delta x^k \\ \lambda^{k+1} & = & \lambda_{\mathrm{QP}}^+ \end{array}$$

Newton's Method - Simplified Notation

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

$$\begin{aligned} \min_{x} \quad & F_{\mathrm{L}}(x; x^k) + \frac{1}{2}(x - x^k)^\top B_k^{\mathrm{ex}}(x - x^k) \\ \text{s.t.} \quad & G_{\mathrm{L}}(x; x^k) \quad = \quad 0, \end{aligned}$$

with $B_k^{\mathrm{ex}} = \nabla_x^2 \mathcal{L}(x^k, \lambda^k)$.

Obtain new iterate x^+ and new multiplier λ_{QP}^+ and iterate

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

Can be called *Sequential Quadratic Programming (SQP)* with exact Hessian and full steps

Overview

NLP with Inequalities

Regard again NLP with both, equalities and inequalities:

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

Introduce Lagrangian function

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

Optimality Conditions with Inequalities

THEOREM(Karush-Kuhn-Tucker (KKT) conditions) For an optimal solution x^* exist multipliers λ^* and μ^* such that

$$\begin{aligned} \nabla_x \mathcal{L}(x^*, \lambda^*, \mu^*) &= 0, \\ G(x^*) &= 0, \\ H(x^*) &\geq 0, \\ \mu^* &\geq 0, \\ H(x^*)^\top \mu^* &= 0, \end{aligned}$$

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 the following Quadratic Program (QP):

$$\min_{x} \quad \nabla F(x^k)^\top \Delta x + \frac{1}{2} \Delta x^\top B_k^{\text{ex}} \Delta x$$
$$\text{s.t.} \quad \begin{cases} G(x^k) + \nabla G(x^k)^\top \Delta x &= 0, \\ H(x^k) + \nabla H(x^k)^\top \Delta x &\geq 0, \end{cases}$$

with

$$B_k^{\text{ex}} = \nabla_x^2 \mathcal{L}(x^k, \lambda^k, \mu^k)$$

and its solution delivers

$$\Delta x^k, \quad \lambda_{\text{QP}}^+, \quad \mu_{\text{QP}}^+$$

Sequential Quadratic Programming (SQP) - Simplified Notation

In each SQP iteration, solve the following QP:

$$\begin{array}{ll} \displaystyle\min_{x} & F_{\mathrm{L}}(x; x^k) + \frac{1}{2}(x - x^k)^\top B_k^{\mathrm{ex}}(x - x^k) \\ \text{s.t.} & \left\{ \begin{array}{lll} G_{\mathrm{L}}(x; x^k) & = & 0, \\ H_{\mathrm{L}}(x; x^k) & \geq & 0, \end{array} \right. \end{array}$$

with

$$B_k^{\mathrm{ex}} = \nabla_x^2 \mathcal{L}(x^k, \lambda^k, \mu^k)$$

and QP solution delivers new iterate

$$x^{k+1}, \quad \lambda^{k+1}, \quad \mu^{k+1}$$

Constrained Gauss-Newton Method

In special case of least squares objectives

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

can approximate Hessian $\nabla_x^2 \mathcal{L}(x^k, \lambda^k, \mu^k)$ by cheaper and always semidefinite matrix

$$B_k^{\mathrm{GN}} = \nabla R(x) \nabla R(x)^\top.$$

Need no multipliers to compute B_k^{GN}. Obtain convex QP subproblem:

$$\begin{aligned}
\min_{\Delta x} \quad & R(x^k)^\top \nabla R(x^k)^\top \Delta x + \frac{1}{2} \Delta x^\top B_k^{\mathrm{GN}} \Delta x \\
\text{s.t.} \quad & G(x^k) + \nabla G(x^k)^\top \Delta x = 0, \\
& H(x^k) + \nabla H(x^k)^\top \Delta x \geq 0,
\end{aligned}$$

Constrained Gauss-Newton Method

In special case of least squares objectives

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

can approximate Hessian $\nabla_x^2 \mathcal{L}(x^k, \lambda^k, \mu^k)$ by cheaper and always semidefinite matrix

$$B_k^{\mathrm{GN}} = \nabla R(x) \nabla R(x)^\top .$$

Need no multipliers to compute B_k^{GN}. Obtain convex QP subproblem:

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

Constrained Gauss-Newton Method

In special case of least squares objectives

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

can approximate Hessian $\nabla_x^2 \mathcal{L}(x^k, \lambda^k, \mu^k)$ by cheaper and always semidefinite matrix

$$B_k^{\mathrm{GN}} = \nabla R(x) \nabla R(x)^\top.$$

Need no multipliers to compute B_k^{GN}. Obtain convex QP subproblem:

$$\begin{aligned} \min_{x} \quad & \frac{1}{2}\|R_{\mathrm{L}}(x; x^k)\|_2^2 \\ \text{s.t.} \quad & G_{\mathrm{L}}(x; x^k) = 0, \\ & H_{\mathrm{L}}(x; x^k) \geq 0, \end{aligned}$$

Constrained Gauss-Newton Method in a Nutshell

Constrained Least-Squares Problem

$$\begin{aligned} \min_{x} \quad & \frac{1}{2}\|R(x)\|_2^2 \\ \text{s.t.} \quad & G(x) = 0, \\ & H(x) \geq 0, \end{aligned}$$

Constrained Gauss-Newton Subproblem

$$\begin{aligned} x^{k+1} = \arg\min_{x} \quad & \frac{1}{2}\|R_{\mathrm{L}}(x;x^k)\|_2^2 \\ \text{s.t.} \quad & G_{\mathrm{L}}(x;x^k) = 0, \\ & H_{\mathrm{L}}(x;x^k) \geq 0, \end{aligned}$$

Linear convergence, i.e.

$$\|x^{k+1} - x^*\| \leq \kappa \|x^k - x^*\|$$

Contraction rate $\kappa < 1$ small if $\|R(x^*)\|_2^2$ small.

[Bock 1987]

Overview

How to solve QP subproblems?

For an equality constrained QP

$$\min_x g^\top x + \frac{1}{2} x^\top A x \;\text{ s.t.}\quad b + Bx \;=\; 0,$$

the solution (x, λ) is just solution of one linear system:

$$\begin{array}{cccccc} g & +Ax & -B^\top \lambda & = & 0, \\ b & +Bx & & = & 0, \end{array}$$

In case of inequalities, two variants exist:

- Active Set Methods (similar to simplex for LP, not explained here)
- Interior Point Methods (next slide)

Interior Point (IP) Methods

For notational convenience, regard inequality constrained QP in following form

$$\min_x g^\top x + \frac{1}{2} x^\top A x \quad \text{s.t.} \quad \begin{array}{rcl} b + Bx & = & 0, \\ x & \geq & 0, \end{array}$$

Idea: replace inequalities by barrier function $-\tau \log(x_i)$, let τ go to zero.

Convex Barrier Subproblem in IP Method

$$\min_x g^\top x + \frac{1}{2} x^\top A x - \tau \sum_i \log(x_i) \quad \text{s.t.} \quad b + Bx \;=\; 0,$$

Solve each τ-problem with Newton-type method for equality constrained optimization.
Can show

- error goes to zero for $\tau \to 0$
- if τ is reduced each time by a constant factor, and each new problem is initialized at old solution, the number of Newton iterations is bounded (polynomial complexity)

Nonlinear Systems in Interior Point Methods

Optimality conditions for

Convex Barrier Subproblem in IP Method

$$\min_x g^\top x + \frac{1}{2} x^\top A x - \tau \sum_i \log(x_i) \quad \text{s.t.} \quad b + Bx = 0,$$

can be shown to be equivalent to system in variables (x, λ, μ)

$$\begin{aligned} g + Ax - B^\top \lambda - \mu &= 0, \\ b + Bx &= 0, \\ x_i \mu_i &= \tau, \quad i = 1, \ldots, n. \end{aligned}$$

Only last condition is nonlinear, it replaces the last KKT condition. The system can be solved by Newton's method, e.g. in QP solver HPIPM for fast MPC.

Note: IP method can also directly address nonlinear programs, e.g. in NLP solver IPOPT.

Summary

- Optimization problems can be:
  - unconstrained or constrained
  - convex or non-convex
  - linear or non-linear
  - differentiable or non-smooth
  - continuous or (mixed-)integer
  - finite or infinite dimensional
  - ...
- Important classes are
  - linear programs (LP)
  - quadratic programs (QP)
  - nonlinear programs (NLP)
- Newton-type algorithms linearize nonlinear functions and solve convex subproblems

General NLP

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

For least-squares: $F(x) = \|R(x)\|_2^2$ get

Gauss-Newton QP subproblem

$$\begin{aligned}\min_{x \in \mathbb{R}^{n_x}} \quad & \|R_{\mathrm{L}}(x; \bar{x})\|_2^2 \\ \text{s.t.} \quad & G_{\mathrm{L}}(x; \bar{x}) = 0 \\ & H_{\mathrm{L}}(x; \bar{x}) \geq 0\end{aligned}$$

Literature

- J. Nocedal, S. Wright: Numerical Optimization, Springer, 1999/2006
- P. E. Gill, W. Murray, M. H. Wright: Practical Optimization, Academic Press, 1981
- R. Fletcher, Practical Methods of Optimization, Wiley, 1987
- D. E. Luenberger: Linear and Nonlinear Programming, Addison Wesley, 1984
- S. Boyd, L. Vandenberghe: Convex Optimization, Cambridge University Press, 2004 (PDF freely available at:

  http://web.stanford.edu/~boyd/cvxbook/