# Newton Type Constrained Optimization in a Nutshell

Moritz Diehl

Optimization in Engineering Center (OPTEC)

K.U. Leuven

# Overview

- Equality Constrained Optimization
- Optimality Conditions and Multipliers
- Newton’s Method = SQP
- Inequality Constraints
- Constrained Gauss Newton Method
- Relation to Optimal Control

# General Nonlinear Program (NLP)

In direct methods, we have to solve the discretized optimal control problem, which is a 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.

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

# Lagrange Function and Optimality Conditions

Introduce Lagrangian function

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

Then for an optimal solution $w^*$ exist multipliers $\lambda^*$ such that

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

# Newton's Method on Optimality Conditions

How to solve nonlinear equations

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

Linearize!

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

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

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

# Newton Step = Quadratic Program

Conditions

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

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

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

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

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

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

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

Introduce Lagrangian function

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

# Optimality Conditions with Inequalities

**THEOREM**(Karush-Kuhn-Tucker (KKT) conditions) For an optimal solution $w^*$ exist multipliers $\lambda^*$ and $\mu^*$ such that

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

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 a Quadratic Program (QP) (we omit these conditions). This QP is

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

with

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

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

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

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

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

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

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

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

# Discrete Time Optimal Control Problem

$$\underset{s,\,q}{\text{minimize}} \quad \sum_{i=0}^{N-1} l_i(s_i, q_i) \; + \; E\left(s_N\right)$$

subject to

$$\begin{aligned}
s_0 - x_0 &= 0, & & \text{(initial value)} \\
s_{i+1} - f_i(s_i, q_i) &= 0, & i = 0, \ldots, N-1, \quad & \text{(discrete system)} \\
h_i(s_i, q_i) &\geq 0, & i = 0, \ldots, N, \quad & \text{(path constraints)} \\
r\left(s_N\right) &\geq 0. & & \text{(terminal constraints)}
\end{aligned}$$

Can arise also from direct multiple shooting parameterization of continous optimal control problem. This NLP can be solved by SQP or Constrained Gauss-Newton method.

# Summary

- Nonlinear Programs (NLP) have nonlinear but differentiable problem functions.
- Sequential Quadratic Programming (SQP) is a Newton type method to solve NLPs that
  - solves in each iteration a Quadratic Program (QP)
  - obtains this QP by linearizing all nonlinear problem functions
- an important SQP variant is the Constrained Gauss-Newton Method
- SQP can be generalized to Sequential Convex Programming (SCP)
- Discrete time optimal control problems are a special case of NLPs.

# Literature

- J. Nocedal and S. Wright: Numerical Optimization, Springer, 2006 (2nd edition)