Newton Type Optimization in a Nutshell

Moritz Diehl

March 27, 2015

Overview

- Equality Constrained Optimization
- Optimality Conditions and Multipliers
- Newton's Method = SQP
- Inequality Constraints
- Constrained Gauss Newton Method
- How to solve QP subproblems?
- Interior Point Methods

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 λ^* 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

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

Linearize!

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

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

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

Newton Step = Quadratic Program

Conditions

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

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

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

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

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

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

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

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 λ^* and μ^* 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

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

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

with

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

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:

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

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

Overview

- Equality Constrained Optimization
- Optimality Conditions and Multipliers
- Newton's Method = SQP
- Inequality Constraints
- Constrained Gauss Newton Method
- **How to solve QP subproblems?**
- Interior Point Methods

How to solve QP subproblems?

For an equality constrained QP

$$\min_{w} g^T w + \frac{1}{2} w^T A w \;\text{ s.t. }\quad b + Bw \;=\; 0,$$

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

$$\begin{array}{cccl} g & +Aw & -B^T\lambda & = 0, \\ b & +Bw & & = 0, \end{array}$$

In case of inequalities, two variants exist:

- Active Set Methods (similar to simplex for LP)
- Interior Point Methods

Interior Point Methods

Regard inequality constrained QP in standard form

$$\min_w g^T w + \frac{1}{2} w^T A w \text{ s.t. } \begin{array}{rcl} b + Bw & = & 0, \\ w & \geq & 0, \end{array}$$

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

$$\min_w g^T w + \frac{1}{2} w^T A w - \tau \sum_i \log(w_i) \text{ s.t. } \quad b + Bw \; = \; 0,$$

Solve each τ-problem with Newton type method. 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!)

Non-Linear Systems in Interior Point Methods

Optimality conditions for

$$\min_w g^T w + \frac{1}{2} w^T A w - \tau \sum_i \log(w_i) \text{ s.t. } \quad b + Bw = 0,$$

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

$$\begin{aligned} g + Aw - B^T \lambda - \mu &= 0, \\ b + Bw &= 0, \\ w_i \mu_i &= \tau, i = 1, \ldots, n. \end{aligned}$$

Only last condition is non-linear, it replaces the last KKT condition. The system can be solved by Newton's method.

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)

Literature

- J. Nocedal and S. Wright: Numerical Optimization, Springer, 2006 (2nd edition)
- S. Boyd and L. Vandenberghe: Convex Optimization, Cambridge Univ. Press, 2004