Equality Constrained Optimization

- Problem Formulation
- Optimization Problems with Linear Equality Constraints
- Optimization Problems with Nonlinear Equality Constraints

Contents

Problem Formulation

Let $F : \mathbb{R}^n \to \mathbb{R}$ and $G : \mathbb{R}^n \to \mathbb{R}^m$ be given twice continuously differentiable functions. We are searching for solutions of the minimization problem

$$\min_x F(x) \qquad \text{s.t.} \qquad G(x) = 0$$

Simple examples:

- For $F(x) = x_1^2 + x_2^2$ and $G(x) = x_1 + x_2 - 1$ the solution is $x_1 = x_2 = \frac{1}{2}$.
- More generally, for $F(x) = \|x\|_2^2$ and $G(x) = Ax - b$ the aim is to find the solution x of the equation $Ax = b$ whose norm is minimal.
- For $F(x) = 0$ the aim is to find points x that satisfy $G(x) = 0$.

Contents

Optimization Problems with Linear Equality Constraints

Let us first consider the special case that the equality constraints are linear,

$$\min_x \; F(x) \qquad \text{s.t.} \qquad Ax - b = 0 \; .$$

In the following, we assume that the matrix $A \in \mathbb{R}^{m \times n}$ has full rank (and $m \leq n$).

Elimination of Linear Equality Constraints

Since we assume that A has full rank we can compute a QR factorization as

$$A = Q[R_1 R_2]$$

with R_1 being an upper triangular and invertible matrix. (we permute the components of x if necessary)

This implies that the components $x = (y^T, z^T)^T$ of the vector x satisfy

$$y = R_1^{-1}\left[Q^T b - R_2 z\right] \ .$$

Now, we can eliminate y explicitly and solve

$$\min_z \tilde{F}(z) \qquad \text{with} \qquad \tilde{F}(z) = F\left(\begin{pmatrix} R_1^{-1}\left[Q^T b - R_2 z\right] \\ z \end{pmatrix}\right) \ ,$$

by applying Newton's method (or Newton-type methods).

A Closer Look at the Newton iterates

If we apply a Newton type method with $\tilde{M}(z) = \tilde{F}''(z)$ to the reduced optimization problem

$$\min_{z} \; \tilde{F}(z)$$

the step direction Δz_k is found by solving the quadratic optimization optimization problems

$$\min_{\Delta z} \; \tilde{F}(z) + \tilde{F}'(z)\Delta z + \frac{1}{2}\Delta z^T \tilde{M}(z)\Delta z$$

as long as $\tilde{F}''(z)$ is positive definite; the next iterate is $z^+ = z + \alpha\Delta z$.

This is the same as solving the quadratic optimization problem

$$\min_{\Delta x} \; F(x) + F'(x)\Delta x + \frac{1}{2}\Delta x^T M(x)\Delta x \qquad \text{s.t.} \qquad A(x+\Delta x) - b = 0$$

with $M(x) = \tilde{F}''(x)$ associating $\Delta x = \left(\Delta y^T \Delta z^T\right)^T$.

A Closer Look at the Newton iterates

Proof:

The derivatives of the function

$$\tilde{F}(z) = F\left(Wz + w\right) \quad \text{with} \quad W = \begin{pmatrix} -R_1^{-1}R_2 \\ I \end{pmatrix}, \quad w = \begin{pmatrix} R_1^{-1}Q^T b \\ 0 \end{pmatrix}$$

can be written in the form

$$\tilde{F}'(z) = F'(x)\,W \quad \text{and} \quad \tilde{F}''(z) = W^T F''(x)\,W$$

The statement from the previous slide follows then by substituting $\Delta x = W\Delta z$. (Also notice the relations $AW = 0$ and $Aw = b$.)

A Closer Look at the Newton iterates

The Newton iterates for the equality constrained problem also satisfy

$$\begin{pmatrix} M(x) & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ \lambda \end{pmatrix} = \begin{pmatrix} -F'(x)^T \\ b - Ax \end{pmatrix}$$

The auxiliary variable λ is called the dual iterate.

Karush-Kuhn-Tucker (KKT) Matrix

The matrix

$$\begin{pmatrix} F''(x) & A^T \\ A & 0 \end{pmatrix}$$

is called Karush-Kuhn-Tucker (KKT) matrix.

- The KKT matrix is invertible if for all $x \neq 0$ with $Ax = 0$ the inequality $x^T F''(x)x > 0$ is satisfied. (i.e., if the Hessian matrix is positive definite on the null-space of the linear equality constraints)
- This is condition is equivalent to stating that the matrix $\tilde{F}''(z) = W^T F''(x) W$ is positive definite.

  (this is the case if z is a "regular" local minimizer of $\tilde{F}$.)

Summary: Newton-type methods for Optimization Problems with Linear Equality Constraints

1. Choose an initial guess $x_0 \in \mathbb{R}^n$ and termination tolerance $\epsilon > 0$.
2. **Repeat:**
   2.1 Choose a positive definite Hessian approximation $M(x_k) \approx F''(x_k)$.
   2.2 Solve the KKT system
   $$\begin{pmatrix} M(x_k) & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} \Delta x_k \\ \lambda \end{pmatrix} = \begin{pmatrix} -F'(x_k)^T \\ b - Ax_k \end{pmatrix}$$
   2.3 Terminate if $|F'(x_k)\Delta x_k| + \sum_{i=0}^{m} |\lambda_i||A_i x_k + b_i| < \epsilon$.
   2.4 Choose a line-search parameter $\alpha \in [0,1]$ and set $x_{k+1} = x_k + \alpha \Delta x_k$, $k \leftarrow k+1$.
3. Output $x^* \approx x_k$ and $\lambda^* \approx \lambda$ as an approximate optimal solution.

Contents

First Order Optimality Conditions

Let us come back to the general nonlinear equality constrained optimization problem

$$\min_x F(x) \qquad \text{s.t.} \qquad G(x) = 0$$

First Order Optimality Conditions: If x^* is a minimizer at which the matrix $A^* = G'(x^*)$ has full rank, then there exists a multiplier $\lambda \in \mathbb{R}^m$ such that

$$\begin{aligned} 0 &= F'(x^*)^T + G'(x^*)^T \lambda^* \\ 0 &= G(x^*) \end{aligned}$$

This condition is called the first order necessary KKT (Karush-Kuhn-Tucker) condition.

Example

Unlike in the linear case, if the linear linear independence constraint qualification is not satisfied, an opimization problem might violate the first order KKT conditions: the problem

$$\min x \quad \text{s.t.} \quad x^2 = 0$$

has a solution at $x^* = 0$, but we have $G'(x^*) = 0$.

In this example, we cannot find a multiplier λ since the equation

$$0 = F'(x^*) + G'(x^*)\lambda = 1$$

is wrong. In practice such degeneracies can often be avoided: in the above example, we could replace the constraint $x^2 = 0$ by the equivalent constraint $x = 0$.

Notation

The first order optimality conditions is sometimes written in the more compact form

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

where L denotes the so-called Lagrange function,

$$L(x, \lambda) = F(x) + G(x)^T \lambda \; .$$

The condition that the matrix $G'(x^*)$ has full rank, is in the literature often called "linear independence constraint qualification" (LICQ).

Newton's Method for Equality Constrained Optimization

An important algorithm for solving equality constrained optimization problems is obtained by applying Newton's method to the equation

$$R(y) = \begin{pmatrix} \nabla_x L(x,\lambda) \\ G(x) \end{pmatrix} = 0 \qquad \text{with} \qquad y = \begin{pmatrix} x \\ \lambda \end{pmatrix} .$$

Here, the iterates have the form

$$y^+ = y - \alpha M(y)^{-1} R(y) \qquad \text{with} \qquad M(y) = \begin{pmatrix} H(x) & A(x)^T \\ A(x) & 0 \end{pmatrix} ,$$

where $H(x) \approx \nabla_x^2 L(x,\lambda)$ and $A(x) \approx G'(x)$ are the Hessian and constraint Jacobian approximation, respectively.

Newton's Method for Equality Constrained Optimization

The term

$$R(y) = \begin{pmatrix} \nabla_x L(x,\lambda) \\ G(x) \end{pmatrix} = \begin{pmatrix} \nabla_x F(x,\lambda) + \lambda^T G'(x) \\ G(x) \end{pmatrix}$$

is called the KKT-residuum. If we apply Newton's method,

$$y^+ = y - \alpha M(y)^{-1} R(y)$$

we need to compute $R(y)$ at every step.

Since we have to compute $G'(x)$ anyhow, we usually choose $A(x) = G'(x)$, i.e., the first order derivative of the function G is evaluated with high accuracy. (Expceptions are the so-called inexact sequential quadratic programming methods; not discussed in this lecture)

Globalization

How do we measure progress towards a solution?

Recall: in unconstrained minimization, the main idea was to accept the next iterate "x^+" if $F(x^+)$ is sufficiently smaller than $F(x)$.

In equality constrained optimization we need to measure two things:

1. The objective value $F(x)$ and
2. the constraint violation $\|G(x)\|$.

L1-penalty functions

One way to measure progress towards a solution is to introduce the L1-penalty function

$$\Phi(x) = F(x) + \sum_{i=1}^{m} \overline{\lambda}_i |G_i(x)| \ .$$

with $\overline{\lambda}_i$ being sufficiently large constants.

An important property of the function $\Phi(x)$ is that (under mild conditions) we have

$$F(x^*) = \Phi(x^*) \qquad \text{but also} \qquad \Phi(x) \geq F(x)$$

for all $x \in C$ for a compact domain $C \subseteq \mathbb{R}^n$ as long as the coefficients $\overline{\lambda}_i$ are sufficiently large.

Armijo Line Search

Similar to unconstrained optimization, the line search parameter α can be found by using back-tracking until the Armijo condition

$$\Phi(x_k + \alpha_k \Delta x_k) \leq \Phi(x_k) + c\alpha_k D(\Phi(x_k), \Delta x_k)$$

for a constant $c \ll 1$ is satisfied. (This condition ensures that the line search parameter is not excessively large, although it is not sufficient to prove convergence in general.) Here, $D(\Phi(x_k), \Delta x_k)$ denotes the directional derivative

$$D(\Phi(x_k), \Delta x_k) = \lim_{\alpha \to 0^+} \frac{\Phi(x_k + \alpha \Delta x_k) - \Phi(x_k)}{\alpha} .$$

Summary: Newton-type methods for Optimization Problems with Equality Constraints

1. Choose initial guesses $x_0 \in \mathbb{R}^n$ and $\lambda_0 \in \mathbb{R}^m$, tolerance $\epsilon > 0$.
2. **Repeat:**
   2.1 Choose a positive definite Hessian approximation
   $M(x_k) \approx F''(x_k) + \lambda_k^T G''(x_k)$ and set $A(x) = G'(x)$.
   2.2 Solve the KKT system
   $$\begin{pmatrix} M(x_k) & A(x_k)^T \\ A(x_k) & 0 \end{pmatrix} \begin{pmatrix} \Delta x_k \\ \tilde{\lambda}_{k+1} \end{pmatrix} = \begin{pmatrix} -\nabla_x L(x_k, \lambda_k) \\ -G(x_k) \end{pmatrix}$$
   2.3 Terminate if $|F'(x_k)\Delta x_k| + \sum_{i=0}^m |\lambda_i||G_i(x_k)| < \epsilon$.
   2.4 Choose a line-search parameter $\alpha \in [0,1]$ and set $x_{k+1} = x_k + \alpha \Delta x_k$ as well as $\lambda_{k+1} = \lambda_k + \alpha(\lambda_{k+1} - \lambda_k)$, $k \leftarrow k+1$.
3. Output $x^* \approx x_k$ and $\lambda^* \approx \lambda_k$ as an approximate optimal solution.