Modern Optimization Techniques

Lucas Rego Drumond

Information Systems and Machine Learning Lab (ISMLL)
Institute of Computer Science
University of Hildesheim, Germany

Equality Constrained Optimization

Outline

Outline

1. Equality Constrained Optimization

2. Newton Methods for Equality Constrained Problems

Equality Constrained Optimization Problems

A **constrained optimization problem** has the form:

$$\begin{array}{rl} \text{minimize} & f_0(\mathbf{x}) \\ \text{subject to} & h_j(\mathbf{x}) = 0, \quad j = 1, \ldots, p \end{array}$$

Where:

- $f_0 : \mathbb{R}^n \to \mathbb{R}$
- $h_1, \ldots, h_p : \mathbb{R}^n \to \mathbb{R}$
- An optimal $\mathbf{x}^*$

Convex Equality Constrained Optimization Problems

An equality constrained optimization problem:

$$\begin{array}{rl} \text{minimize} & f_0(\mathbf{x}) \\ \text{subject to} & h_j(\mathbf{x}) = 0, \quad j = 1, \ldots, p \end{array}$$

is convex iff:

- f_0 is convex
- $h_1, \ldots, h_p$ are affine

$$\begin{array}{rl} \text{minimize} & f_0(\mathbf{x}) \\ \text{subject to} & A\mathbf{x} = \mathbf{b} \end{array}$$

Optimality criterion

Given the following problem:

$$\begin{aligned} \text{minimize} \quad & f_0(\mathbf{x}) \\ \text{subject to} \quad & A\mathbf{x} = \mathbf{b} \end{aligned}$$

The Lagrangian is given by:

$$L(\mathbf{x}, \nu) = f_0(\mathbf{x}) + \nu(A\mathbf{x} - \mathbf{b})$$

And it's derivative:

$$\nabla_{\mathbf{x}} L(\mathbf{x}, \nu) = \nabla_{\mathbf{x}} f_0(\mathbf{x}) + A^T \nu$$

Optimality criterion

Given the following problem:

$$\begin{array}{ll} \text{minimize} & f_0(\mathbf{x}) \\ \text{subject to} & A\mathbf{x} = \mathbf{b} \end{array}$$

The optimal solution $\mathbf{x}^*$ must fulfill the KKT Conditions:

1. Primal feasibility: $f_i(\mathbf{x}^*) \leq 0$ and $h_j(\mathbf{x}) = 0$ for all i, j
2. Dual feasibility: $\lambda \succeq 0$
3. Complementary Slackness: $\lambda_i f_i(\mathbf{x}^*) = 0$ for all i
4. Stationarity: $\nabla f_0(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i \nabla f_i(\mathbf{x}^*) + \sum_{i=1}^{p} \nu_i \nabla h_i(\mathbf{x}^*) = 0$

1. Primal feasibility: $f_i(\mathbf{x}^*) \leq 0$ and $h_j(\mathbf{x}) = 0$ for all i, j
2. Dual feasibility: $\lambda \succeq 0$
3. Complementary Slackness: $\lambda_i f_i(\mathbf{x}^*) = 0$ for all i
4. Stationarity: $\nabla f_0(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i \nabla f_i(\mathbf{x}^*) + \sum_{i=1}^{p} \nu_i \nabla h_i(\mathbf{x}^*) = 0$

Since there are no inequality constraints, the conditions in red are

Optimality criterion

Given the following problem:

$$\begin{array}{rl} \text{minimize} & f_0(\mathbf{x}) \\ \text{subject to} & A\mathbf{x} = \mathbf{b} \end{array}$$

The optimal solution $\mathbf{x}^*$ must fulfil the KKT Conditions:

- Primal feasibility: $h_j(\mathbf{x}^*) = 0$
- Stationarity: $\nabla f_0(\mathbf{x}^*) + \sum_{i=1}^{p} \nu_i \nabla h_i(\mathbf{x}^*) = 0$

for $h_j(\mathbf{x}) = \mathbf{a_j x} - b_j$ we have:

- Primal feasibility: $A\mathbf{x}^* = \mathbf{b}$
- Stationarity: $\nabla f_0(\mathbf{x}^*) + A^T \nu^* = 0$

Optimality criterion

Given the following problem:

$$\begin{array}{rl} \text{minimize} & f_0(\mathbf{x}) \\ \text{subject to} & A\mathbf{x} = \mathbf{b} \end{array}$$

$\mathbf{x}^*$ is optimal iff there exists a ν^*:

- Primal feasibility: $A\mathbf{x}^* = \mathbf{b}$
- Stationarity: $\nabla f_0(\mathbf{x}^*) + A^T\nu^* = 0$

Example

Given the following problem:

$$\begin{array}{rl} \text{minimize} & (x_1 - 2)^2 + 2(x_2 - 1)^2 - 5 \\ \text{subject to} & x_1 + 4x_2 = 3 \end{array}$$

- Primal feasibility: $x_1 + 4x_2 = 3$
- Stationarity: $\nabla f_0(\mathbf{x}^*) + [1 \;\; 4]^T \nu^* = 0$

$$\frac{\partial f_0}{\partial x_1} = 2(x_1 - 2) = 2x_1 - 4$$
$$\frac{\partial f_0}{\partial x_2} = 4(x_2 - 1) = 4x_2 - 4$$

Example

From the KKT conditions we have:

- Primal feasibility: $x_1 + 4x_2 = 3$
- Stationarity:

$$\begin{bmatrix} 2x_1 - 4 \\ 4x_2 - 4 \end{bmatrix} + \nu \begin{bmatrix} 1 \\ 4 \end{bmatrix} = 0$$

This gives us the following system of equations:

$$\begin{aligned} 2x_1 + \nu &= 4 \\ 4x_2 + 4\nu &= 4 \\ x_1 + 4x_2 &= 3 \end{aligned} \qquad \begin{bmatrix} 2 & 0 & 1 \\ 0 & 4 & 4 \\ 1 & 4 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \nu \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \\ 3 \end{bmatrix}$$

With solution: $x_1 = \frac{5}{3}$, $x_2 = \frac{1}{3}$, $\nu = \frac{2}{3}$

Example II - Quadratic Programming

Given P positive semi-definite, the following problem:

$$
\begin{array}{ll}
\text{minimize} & \frac{1}{2}\mathbf{x}^T P\mathbf{x} + \mathbf{q}^T\mathbf{x} + r \\
\text{subject to} & A\mathbf{x} = \mathbf{b}
\end{array}
$$

Optimality Condition:

$$
\begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \mathbf{x}^* \\ \nu^* \end{bmatrix} = \begin{bmatrix} -\mathbf{q} \\ \mathbf{b} \end{bmatrix}
$$

- KKT Matrix
- Solution is the inverse of the KKT matrix times the right hand side of the system
- The KKT matrix is nonsingular iff:

$$
A\mathbf{x} = 0, \quad \mathbf{x} \neq 0 \quad \Rightarrow \mathbf{x}^T P\mathbf{x} > 0
$$

Outline

Descent step for equality constrained problems

Given the following problem:

$$\begin{array}{ll} \text{minimize} & f_0(\mathbf{x}) \\ \text{subject to} & A\mathbf{x} = \mathbf{b} \end{array}$$

we want to start with a feasible solution $\mathbf{x}$ and compute a step $\Delta\mathbf{x}$ such that

- $f_0(\mathbf{x} + \Delta\mathbf{x}) \leq f_0(\mathbf{x})$
- $A(\mathbf{x} + \Delta\mathbf{x}) = \mathbf{b}$

Which means solving the following problem for $\Delta\mathbf{x}$:

$$\begin{array}{ll} \text{minimize} & f_0(\mathbf{x} + \Delta\mathbf{x}) \\ \text{subject to} & A(\mathbf{x} + \Delta\mathbf{x}) = \mathbf{b} \end{array}$$

Newton Step

The Newton Step is the solution for the minimization of the second order approximation of f_0:

$$\begin{aligned} \text{minimize} \quad & \hat{f}(\mathbf{x} + \Delta\mathbf{x}) = f_0(\mathbf{x}) + \nabla f_0(\mathbf{x})^T \Delta\mathbf{x} + \frac{1}{2}\Delta\mathbf{x}^T \nabla^2 f_0(\mathbf{x})\Delta\mathbf{x} \\ \text{subject to} \quad & A(\mathbf{x} + \Delta\mathbf{x}) = \mathbf{b} \end{aligned}$$

The equality constraint can be rewritten as

$$\begin{aligned} A(\mathbf{x} + \Delta\mathbf{x}) &= \mathbf{b} \\ A\mathbf{x} + A\Delta\mathbf{x} &= \mathbf{b} \\ A\Delta\mathbf{x} &= \mathbf{b} - A\mathbf{x} \end{aligned}$$

And since we assume $\mathbf{x}$ feasible, we have $A\mathbf{x} = \mathbf{b}$:

$$A\Delta\mathbf{x} = 0$$

Newton Step

The Newton Step is the solution for the minimization of the second order approximation of f_0:

$$\begin{array}{ll} \text{minimize} & \hat{f}(\mathbf{x} + \Delta\mathbf{x}) = f_0(\mathbf{x}) + \nabla f_0(\mathbf{x})^T \Delta\mathbf{x} + \frac{1}{2}\Delta\mathbf{x}^T \nabla^2 f_0(\mathbf{x}) \Delta\mathbf{x} \\ \text{subject to} & A\Delta\mathbf{x} = \mathbf{0} \end{array}$$

$$\begin{array}{ll} \text{minimize} & \hat{f}(\mathbf{x} + \Delta\mathbf{x}) = \color{green}{f_0(\mathbf{x})} + \color{blue}{\nabla f_0(\mathbf{x})}^T \Delta\mathbf{x} + \frac{1}{2}\Delta\mathbf{x}^T \color{red}{\nabla^2 f_0(\mathbf{x})} \Delta\mathbf{x} \\ \text{subject to} & A\Delta\mathbf{x} = \mathbf{0} \end{array}$$

This is a quadratic programming with:

- $\color{red}{P = \nabla^2 f_0(\mathbf{x})}$
- $\color{blue}{\mathbf{q} = \nabla f_0(\mathbf{x})}$
- $\color{green}{r = f_0(\mathbf{x})}$

and optimality conditions:

Newton Step

The Newton Step is the solution for the minimization of the second order approximation of f_0:

$$\begin{array}{ll} \text{minimize} & \hat{f}(\mathbf{x}+\Delta\mathbf{x}) = f_0(\mathbf{x}) + \nabla f_0(\mathbf{x})^T\Delta\mathbf{x} + \frac{1}{2}\Delta\mathbf{x}^T\nabla^2 f_0(\mathbf{x})\Delta\mathbf{x} \\ \text{subject to} & A\Delta\mathbf{x} = \mathbf{0} \end{array}$$

Is computed by solving the following system:

$$\begin{bmatrix} \nabla^2 f_0(\mathbf{x}) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta\mathbf{x} \\ \mathbf{w} \end{bmatrix} = \begin{bmatrix} -\nabla f_0(\mathbf{x}) \\ \mathbf{0} \end{bmatrix}$$

For the unconstrained case:

$$\nabla^2 f_0(\mathbf{x})\Delta\mathbf{x} = -\nabla f_0(\mathbf{x})$$

Newton's method for Equality Constrained Problems

1: **procedure** NEWTONS METHOD
input: f_0, initial feasible point $\mathbf{x} \in \operatorname{dom} f_0$ and $A\mathbf{x} = \mathbf{b}$
2: **repeat**
3: Get $\Delta\mathbf{x}$ by solving $\begin{bmatrix} \nabla^2 f_0(\mathbf{x}) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta\mathbf{x} \\ \mathbf{w} \end{bmatrix} = \begin{bmatrix} -\nabla f_0(\mathbf{x}) \\ \mathbf{0} \end{bmatrix}$
4: Get Step Size μ
5: $\mathbf{x} \leftarrow \mathbf{x} + \mu\Delta\mathbf{x}$
6: **until** convergence
7: **return** $\mathbf{x}$, $f_0(\mathbf{x})$
8: **end procedure**

What if we don't have a feasible $\mathbf{x}$ to start with?

Newton Step at infeasible points

If $\mathbf{x}$ is infeasible, i.e. $A\mathbf{x} \neq \mathbf{b}$, we have the following problem:

$$\begin{aligned} \text{minimize} \quad & \hat{f}(\mathbf{x} + \Delta\mathbf{x}) = f_0(\mathbf{x}) + \nabla f_0(\mathbf{x})^T \Delta\mathbf{x} + \frac{1}{2}\Delta\mathbf{x}^T \nabla^2 f_0(\mathbf{x}) \Delta\mathbf{x} \\ \text{subject to} \quad & A\Delta\mathbf{x} = \mathbf{b} - A\mathbf{x} \end{aligned}$$

Which can be solved for $\Delta\mathbf{x}$ by solving the following system of equations:

$$\begin{bmatrix} \nabla^2 f_0(\mathbf{x}) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta\mathbf{x} \\ \mathbf{w} \end{bmatrix} = - \begin{bmatrix} \nabla f_0(\mathbf{x}) \\ A\mathbf{x} - \mathbf{b} \end{bmatrix}$$

After one step in this algorithm, we have a feasible solution

Solving KKT systems of equations

The KKT systems are systems of equations that look like this:

$$\begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \mathbf{v} \\ \mathbf{w} \end{bmatrix} = - \begin{bmatrix} \mathbf{g} \\ \mathbf{h} \end{bmatrix}$$

Standard methods for solving it:

- LDL^T factorization
- Elimination (might require inverting H)