African Institute for Mathematical Sciences
African Master's in Machine Intelligence

Group 08
Equality Constrained Minimization

Members :
Volviane Saphir MFOGO
Moteu Ngoli Tatiana
Salomon Kabongo KABENAMUALU
Jean Paul ISHIMWE

January 3, 2020

Content

Equality Constrained Minimization
Formulation

Standard form of equality constrained

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Ax = b \end{array} \tag{1}$$

Equality Constrained Minimization
Formulation

Standard form of equality constrained

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Ax = b \end{array} \tag{1}$$

- f convex, twice continuously differentiable

Standard form of equality constrained

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Ax = b \end{array} \tag{1}$$

- f convex, twice continuously differentiable
- $A \in \mathbb{R}^{p \times n}$ with rank $A = p < n$

Standard form of equality constrained

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Ax = b \end{array} \tag{1}$$

- f convex, twice continuously differentiable
- $A \in \mathbb{R}^{p \times n}$ with rank $A = p < n$
- We will assume that an optimal solution $x^\star$ exists, and use $p^\star$ to denote the optimal value, $p^\star = \inf\{f(x) | Ax = b\} = f(x^\star)$

Equality Constrained Minimization
Formulation

Standard form of equality constrained

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Ax = b \end{array} \tag{1}$$

- f convex, twice continuously differentiable
- $A \in \mathbb{R}^{p \times n}$ with rank $A = p < n$
- We will assume that an optimal solution $x^\star$ exists, and use $p^\star$ to denote the optimal value, $p^\star = \inf\{f(x) | Ax = b\} = f(x^\star)$

$x^\star$ is optimal iff there exists a $\nu^\star$ such that

$$Ax^\star = b, \quad \nabla f(x^\star) + A^T \nu^\star = 0 \tag{2}$$

Equality Constrained Quadratic Minimization

With $P \in \mathbf{S}_+^n$

$$\begin{array}{ll} \text{minimize} & (1/2)x^T P x + q^T x + r \\ \text{subject to} & Ax = b \end{array} \tag{3}$$

Equality Constrained Quadratic Minimization

With $P \in \mathbf{S}_+^n$

$$\begin{array}{ll} \text{minimize} & (1/2)x^T P x + q^T x + r \\ \text{subject to} & Ax = b \end{array} \tag{3}$$

Optimality condition:

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

- coefficient matrix is called KKT matrix
- KKT matrix is nonsingular if and only if

$$Ax = 0, \quad x \neq 0 \quad \Longrightarrow \quad x^T P x > 0$$

Elimination Equality Constrained

- One general approach to solving the equality constrained problem (1) is to eliminate the equality constraints and then solve the resulting unconstrained problem using methods for unconstrained minimization.

Elimination Equality Constrained

- One general approach to solving the equality constrained problem (1) is to eliminate the equality constraints and then solve the resulting unconstrained problem using methods for unconstrained minimization.

- We first find a matrix $F \in \mathbb{R}^{n \times (n-p)}$ and vector $\hat{x} \in \mathbb{R}^n$ that parametrize the (affine) feasible set:

$$\{x | Ax = b\} = \{Fz + \hat{x} | z \in \mathbb{R}^{n-p}\}$$

Elimination Equality Constrained

- $\hat{x}$ can be chosen as any particular solution of $Ax = b$

Elimination Equality Constrained

- $\hat{x}$ can be chosen as any particular solution of $Ax = b$
- $F \in \mathbb{R}^{n \times (n-p)}$ is any matrix whose range is the nullspace of A ie ($rank\ F = n - p\ and\ AF = 0$)

Elimination Equality Constrained

- $\hat{x}$ can be chosen as any particular solution of $Ax = b$
- $F \in \mathbb{R}^{n\times(n-p)}$ is any matrix whose range is the nullspace of A ie ($rank\ F = n - p\ and\ AF = 0$)
- Then the reduced or eliminated optimization problem will be:

$$\text{minimize } \hat{f}(z) = f(Fz + \hat{x})$$

which is an unconstrained problem with variable $z \in \mathbb{R}^{n-p}$

Elimination Equality Constrained

- $\hat{x}$ can be chosen as any particular solution of $Ax = b$
- $F \in \mathbb{R}^{n \times (n-p)}$ is any matrix whose range is the nullspace of A ie ($rank\ F = n - p\ and\ AF = 0$)
- Then the reduced or eliminated optimization problem will be:

$$\text{minimize } \hat{f}(z) = f(Fz + \hat{x})$$

which is an unconstrained problem with variable $z \in \mathbb{R}^{n-p}$

- From its solution $z^\star$,we can find the solution of the equality constrained problem as $x^\star = Fz^\star + \hat{x}$

Elimination Equality Constrained

- $\hat{x}$ can be chosen as any particular solution of $Ax = b$
- $F \in \mathbb{R}^{n\times(n-p)}$ is any matrix whose range is the nullspace of A ie ($rank\ F = n - p\ and\ AF = 0$)
- Then the reduced or eliminated optimization problem will be:

$$\text{minimize } \hat{f}(z) = f(Fz + \hat{x})$$

  which is an unconstrained problem with variable $z \in \mathbb{R}^{n-p}$
- From its solution $z^\star$,we can find the solution of the equality constrained problem as $x^\star = Fz^\star + \hat{x}$
- We can also construct an optimal dual variable $\nu^\star$ for the equality constrained problem, as $\nu^\star = -(AA^T)^{-1}A \quad \nabla f(x^\star)$

Elimination Equality Constrained

Example:

$$\begin{array}{ll} \text{minimize} & f_1(x_1) + f_2(x_2) + \ldots + f_n(x_n) \\ \text{subject to} & x_1 + x_2 + \ldots + x_n = b \end{array}$$

We can eliminate x_n (for example) using the parametrization:
$x_n = b - x_1 - \ldots x_{n-1}$
which corresponds to the choices:

$$\hat{x} = be_n, \quad F = \begin{bmatrix} I \\ -\mathbf{1}^T \end{bmatrix} \in \mathbf{R}^{n \times (n-1)}$$

Elimination Equality Constrained

Example:

$$\begin{array}{ll} \text{minimize} & f_1(x_1) + f_2(x_2) + \ldots + f_n(x_n) \\ \text{subject to} & x_1 + x_2 + \ldots + x_n = b \end{array}$$

We can eliminate x_n (for example) using the parametrization:
$x_n = b - x_1 - \ldots x_{n-1}$
which corresponds to the choices:

$$\hat{x} = be_n, \quad F = \begin{bmatrix} I \\ -\mathbf{1}^T \end{bmatrix} \in \mathbf{R}^{n \times (n-1)}$$

The reduced problem is then

minimize $f_1(x_1) + f_2(x_2) + \ldots + f_{n-1}(x_{n-1}) + f_n(b - x_1 - \ldots x_{n-1})$
with variables $x_1, \ldots, x_{n-1}$.

Motivation

Here we describe an extension of Newton's method to include linear equality constraint. The methods are almost the same except for two differences:

Motivation

Here we describe an extension of Newton's method to include linear equality constraint. The methods are almost the same except for two differences:

- The initial point must be feasible (satisfy $x \in \textit{dom } f$ and Ax = b).

Motivation

Here we describe an extension of Newton's method to include linear equality constraint. The methods are almost the same except for two differences:

- The initial point must be feasible (satisfy $x \in \textit{dom } f$ and Ax = b).
- the Newton step must be a feasible direction ($A\Delta x_{nt} = 0$). i.e. the definition of Newton step is modified to take the equality constraints into account.

To derive the Newton step Δx_{nt} for the standard equality constrained problem

Equality constrained minimization

$$\begin{array}{ll} \textit{Minimize} & f(x) \\ \textit{Subject to} & Ax = b \end{array}$$

at a feasible point x, we replace the objective with its second-order Taylor approximation near x, to form the problem

To derive the Newton step Δx_{nt} for the standard equality constrained problem

Equality constrained minimization

$$\begin{array}{ll} \textit{Minimize} & f(x) \\ \textit{Subject to} & Ax = b \end{array}$$

at a feasible point x, we replace the objective with its second-order Taylor approximation near x, to form the problem

Newton with Equality Constrained

$$\begin{array}{lc} \textit{Minimize} & \hat{f}(x+v) = f(x) + \nabla f(x)^T v + (1/2) v^T \nabla^2 f(x) v \\ \textit{Subject to} & A(x+v) = b \end{array}$$

with variable v.

Newton Step cont'd

- This is a (convex) quadratic minimization problem with equality constraints

Optimal condition

$$\begin{aligned} Ax^* &= b \\ \nabla \hat{f}(x^*) + A^T \nu^* &= 0 \end{aligned}$$

We substitute $x + \Delta x_{nt}$ for x^* and w for ν^* , and replace the gradient term in the second equation by its linearized approximation near x

Newton Step cont'd

- This is a (convex) quadratic minimization problem with equality constraints

Optimal condition

$$\begin{aligned} Ax^* &= b \\ \nabla \hat{f}(x^*) + A^T \nu^* &= 0 \end{aligned}$$

We substitute $x + \Delta x_{nt}$ for x^* and w for ν^* , and replace the gradient term in the second equation by its linearized approximation near x

Newton Optimal condition

$$A(x + \Delta x_{nt}) = b, \nabla f(x + \Delta x_{nt}) + A^T w \approx \nabla f(x) + \nabla^2 f(x) \Delta x_{nt} + A^T w$$

Using Ax=b, these become

$$\text{A } \Delta x_{nt} = 0, \ \nabla^2 f(x) \Delta x_{nt} + A^T w = -\nabla f(x)$$

Newton Step cont'd

- From our analysis of the equality constrained quadratic problem, the Newton step Δx_{nt} is characterized by

Newton Step cont'd

- From our analysis of the equality constrained quadratic problem, the Newton step Δx_{nt} is characterized by

Newton Optimal condition

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

where w is the associated optimal dual variable for the quadratic problem.

- The Newton step is defined only at points for which the KKT matrix is nonsingular.

The Newton Decrement

At a certain point x, we can define newton decrement ($\lambda(x)$) for the equality constrained problem as:

$$\lambda(x) = \left(\Delta x_{nt}^T \nabla^2 f(x) \Delta x_{nt}\right)^{1/2} = \left(-\nabla f(x)^T \Delta x_{nt}\right)^{1/2}$$

1) The difference between $f(x)$ and the minimum of its approximate quadratic form is given by:

$$f(x) - \inf_{Ax=b} \hat{f}(y) = \frac{1}{2}\lambda(x)^2$$

2) Directional derivative in Newton direction:

$$\frac{d}{dt} f(x + t\Delta x_{nt})|_{t=0} = \nabla f(x)^T \Delta x_{nt} = -\lambda(x)^2$$

Algorithm

Newton's Method with Equality Constraints

given a starting point $x \in \mathbf{dom}\, f$, tolerance $\epsilon > 0$.

repeat

1. *Compute the Newton step and decrement.*
   $\Delta x_{\mathrm{nt}} := -\nabla^2 f(x)^{-1} \nabla f(x); \quad \lambda^2 := \nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x).$
2. *Stopping criterion.* **quit** if $\lambda^2/2 \leq \epsilon$.
3. *Line search.* Choose step size t by backtracking line search.
4. *Update.* $x := x + t\Delta x_{\mathrm{nt}}$.

Algorithm

Newton's Method with Equality Constraints

given a starting point $x \in \mathbf{dom}\, f$, tolerance $\epsilon > 0$.

repeat

1. *Compute the Newton step and decrement.*
   $\Delta x_{\rm nt} := -\nabla^2 f(x)^{-1} \nabla f(x); \quad \lambda^2 := \nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x)$.
2. *Stopping criterion.* **quit** if $\lambda^2/2 \leq \epsilon$.
3. *Line search.* Choose step size t by backtracking line search.
4. *Update.* $x := x + t\Delta x_{\rm nt}$.

- **A feasible descent method:** Meaning that for every step $x^{(k)}$ is feasible and your objective goes down unless you are at the optimal point (i.e $f(x^{(k+1)} < f(x^{(k)}))$.

Algorithm

Newton's Method with Equality Constraints

given a starting point $x \in \mathbf{dom}\, f$, tolerance $\epsilon > 0$.
repeat

1. *Compute the Newton step and decrement.*
   $\Delta x_{\mathrm{nt}} := -\nabla^2 f(x)^{-1} \nabla f(x); \quad \lambda^2 := \nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x).$
2. *Stopping criterion.* **quit** if $\lambda^2/2 \leq \epsilon$.
3. *Line search.* Choose step size t by backtracking line search.
4. *Update.* $x := x + t\Delta x_{\mathrm{nt}}$.

- **A feasible descent method:** Meaning that for every step $x^{(k)}$ is feasible and your objective goes down unless you are at the optimal point (i.e $f(x^{(k+1)} < f(x^{(k)}))$.
- **Affine Invariant:** Any change in coordinates, it doesn't matter If we defined $g(y) = f(Ay)$ for nonsingular A, then $\lambda_g(y)$ would match $\lambda_f(x)$ at $x = Ay$

Implementation

To illustrate the algorithms, we introduce two toy functions to minimize:

- **simple quadratic problem:**

$$f(x_1, x_2) = \frac{1}{2}(x_1^2 + \gamma x_2^2),$$

where γ determines the condition number.

- **a non-quadratic function:**

$$f(x_1, x_2) = \log(e^{x_1+3x_2-0.1} + e^{x_1-3x_2-0.1} + e^{-x_1-0.1}).$$

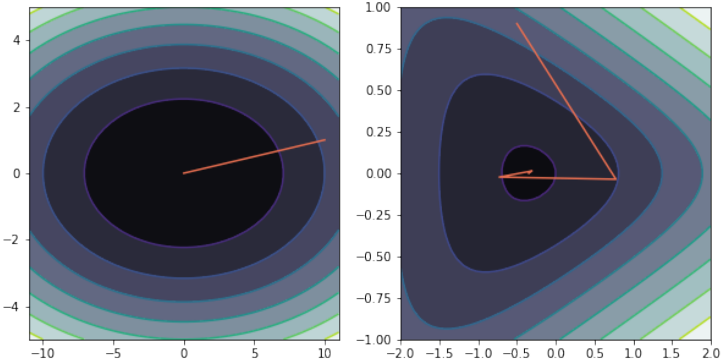

Figure: Number of steps quadratic function: 1 and Number of steps non-quadratic function: 5

The code can be seen here

Elimination of the equality constraint

- Newton's method for the reduced problem:

$$\text{minimize} \quad \hat{f}(z) = f(Fz + \hat{x})$$

starting at $z^{(0)}$, generates iterates $z^{(k)}$.

Elimination of the equality constraint

- Newton's method for the reduced problem:

$$\text{minimize} \quad \hat{f}(z) = f(Fz + \hat{x})$$

starting at $z^{(0)}$, generates iterates $z^{(k)}$.

- Newton's method with equality constraints : when started at $x^{(0)} = Fz^{(0)} + \hat{x}$ are :

$$x^{(k)} = Fz^{(k)} + \hat{x}$$

$\longrightarrow$ the iterates in Newton's method for the equality constrained problem coincide with the iterates in Newton's method applied to the unconstrained reduced problem. All convergence analysis therefore remains valid.

Summary

- The Newton method for equality constrained optimization problems is the most natural extension of the Newton's method for unconstrained problem: it solves the problem on the affine subset of constraints.
- All results valid for the Newton's method on unconstrained problems remain valid, in particular it is a **good method**.
- Drawback: we need a feasible initial point. **What if it not?**

Motivation

- Newton's method for constrained problem is a descent method that generates a sequence of feasible points.
- This requires in particular a feasible point as a starting point.
- Here we generalize Newton's method to work with initial points and iterates that are not feasible.
- A price to pay is that it is not necessarily a descent method.

- We start with the optimality conditions for the equality constrained minimization problem:

$$Ax^\star = b, \ \nabla f(x^\star) + A^T \nu^\star = 0$$

x denote the current point, which we do not assume to be feasible, but we do assume satisfies $x \in \mathit{dom} f$.

- We start with the optimality conditions for the equality constrained minimization problem:

$$Ax^\star = b,\ \nabla f(x^\star) + A^T \nu^\star = 0$$

x denote the current point, which we do not assume to be feasible, but we do assume satisfies $x \in \mathit{dom} f$.

- This is to find a step Δx so that $x + \Delta x$ satisfies the optimality conditions i.e $x + \Delta x \approx x^\star$

Newton Step at Infeasible Points

To do this, we substitute $x + \Delta x$ for $x^\star$ and for $v^\star$ in the optimality conditions, and use the first-order approximation

$$\nabla f(x + \Delta x) \approx \nabla f(x) + \nabla^2 f(x) \Delta x$$

for the gradient to obtain

$$A(x + \Delta x) = b, \ \nabla f(x) + \nabla^2 f(x) \Delta x + A^T w = 0$$

This is a set of linear equations for Δx and w,

Linear equation

$$\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ w \end{bmatrix} = - \begin{bmatrix} \nabla f(x) \\ Ax - b \end{bmatrix} \quad (1)$$

Infeasible start Newton method

given starting point $x \in \mathbf{dom}\, f$, ν, tolerance $\epsilon > 0$, $\alpha \in (0, 1/2)$, $\beta \in (0, 1)$.

repeat

1. Compute primal and dual Newton steps Δx_{nt}, $\Delta \nu_{\mathrm{nt}}$.
2. *Backtracking line search on* $\|r\|_2$.

   $t := 1.$

   while $\|r(x + t\Delta x_{\mathrm{nt}}, \nu + t\Delta \nu_{\mathrm{nt}})\|_2 > (1 - \alpha t)\|r(x, \nu)\|_2, \quad t := \beta t.$
3. *Update.* $x := x + t\Delta x_{\mathrm{nt}}$, $\nu := \nu + t\Delta \nu_{\mathrm{nt}}$.

until $Ax = b$ and $\|r(x, \nu)\|_2 \le \epsilon$.

References

- Boyd, Stephen, and Lieven Vandenberghe. Convex optimization. Cambridge university press, 2004.

Thank You!