Constrained Optimization I

Asst. Prof. Dr.-Ing. Sudchai Boonto

November 16, 2023

Department of Control System and Instrumentation Engineering
King Mongkut's Unniversity of Technology Thonburi
Thailand

Objective

At the end of this chapter you should be able to:

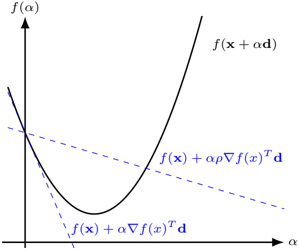
- Describe, implement, and the constrained optimization problems

Reviews: Back Tracking Line Search

- The **backtracking** line search is very simple and quite effective.
- Depend on two constants α, β with $0 < \beta < 0.5, 0 < \rho < 1$.

Require: a descent direction $\mathbf{d}$, $\beta \in (0, 0.5)$ and $\rho \in (0, 1)$

$\alpha = 1$
while $f(\mathbf{x} + \alpha\mathbf{d}) > f(\mathbf{x}) + \beta\alpha\nabla f(\mathbf{x})^T\mathbf{d}$ do
  $\alpha = \rho\alpha$
end while
return α



- The lower dashed line shows the linear extrapolation of $f(\mathbf{x})$
- The upper dashed line ahs a slope of a factor of ρ smaller.
- The backtracking condition is that $f(\mathbf{x})$ is lies below the upper dashed line.

Newton Method with Line Search

Minimizer of second-order approximation

$$f(\mathbf{x}+\Delta\mathbf{x}) \approx f(\mathbf{x}) + \nabla f(\mathbf{x})^T \Delta\mathbf{x} + \frac{1}{2}\Delta\mathbf{x}^T \nabla^2 f(x) \Delta\mathbf{x}$$

$$\nabla f(\mathbf{x}+\Delta\mathbf{x}) = \nabla f(\mathbf{x}) + \nabla^2 f(x)\Delta\mathbf{x} = 0 \qquad \text{(with respect to } \Delta\mathbf{x})$$

The second equation is the optimality condition. We have

$$\Delta\mathbf{x} = -\nabla^2 f(\mathbf{x})^{-1} \nabla f(\mathbf{x}) = \mathbf{d}$$

The positive definiteness of $\nabla^2 f(x)$ implies that

$$\nabla f(x)^T \Delta x = -\nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x) = -\mathbf{g}^T \mathbf{H}^{-1} \mathbf{g} < 0$$

The Newton step is a descent direction (unless $\mathbf{x}$ is optimal).

Newton Method with Line Search

- The local Newton method is

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \nabla^2 f(\mathbf{x})^{-1} \nabla f(\mathbf{x}_k)$$

- Newton method with line search

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k \mathbf{D}_k \nabla f(\mathbf{x}_k) = \mathbf{x}_k - \alpha_k \nabla^2 f(\mathbf{x})^{-1} \nabla f(\mathbf{x}_k)$$

  In general the Hessian matrix $\nabla^2 f(\mathbf{x}_k)$ may not be positive definite, it is necessary to choose another preconditioner.

- One of then involves choosing $\mathbf{D}_k$ diagonal, with entries

$$\mathbf{D}_k(i,i) = \max\left(\varepsilon, \frac{\partial^2 f}{\partial(x)_i^2(\mathbf{x}_k)}\right)^{-1}$$

  with $\varepsilon > 0$. Then, each diagonal element (i.e., each eigenvalue) is greater than or equal to ε, which guarantees the positive definiteness of the matrix.

Newton Method with Line Search

Modified Cholesky factorization

Require: A symmetric matrix $A \in \mathbb{R}^{n \times n}$

$k = 0$

if $\min_i a_{ii} > 0$ **then**

  $\tau_k = 0$

else

  $\tau_k = \frac{1}{2}\|A\|_F$

end if

repeat

  Calculate the Cholesky factorization $\mathbf{L}\mathbf{L}^T$ of $A + \tau\mathbf{I}$

  if the factorization is not successful **then**

    $\tau_{k+1} = \max(2\tau_k, \frac{1}{2}\|A\|_F)$

    $k = k + 1$

  end if

until the factorization is successful

return τ_{k+1}

Newton Method with Line Search

- The most widely used technique is the regularization (damped Newton or Levenberg–Marquardt algorithm).

$$\mathbf{D}_k = \left(\nabla^2 f(\mathbf{x}_k) + \tau \mathbf{I}\right)^{-1}$$

Require: a starting point $\mathbf{x}$, tolerance $\varepsilon > 0$

$k = 0$

repeat

Calculate a lower triangular matrix $\mathbf{L}_k$ and τ such that

$\mathbf{L}_k \mathbf{L}_k^T = \nabla^2 f(\mathbf{x}_k) + \tau \mathbf{I}$

by using the modified Cholesky factorization

Find $\mathbf{z}_k$ by solving the triangular system $\mathbf{L}_k \mathbf{z}_k = \nabla f(\mathbf{x}_k)$

Find $\mathbf{d}_k$ by solving the triangular system $\mathbf{L}_k^T \mathbf{d}_k = -\mathbf{z}_k$

Choose step size α_k by backtracking line search with $\alpha_0 = 1$

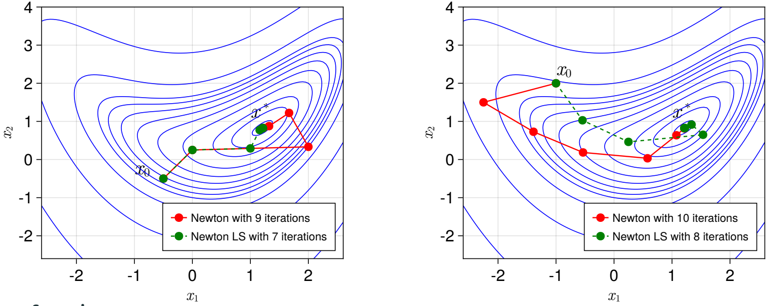
$x_{k+1} = x_k + \alpha_k \mathbf{d}_k$

$k = k + 1$

until $\|\nabla f(\mathbf{x}_k)^T \nabla^2 f(\mathbf{x}_k)^{-1} \nabla f(\mathbf{x}_k)\| \leq \varepsilon$

return x_{k+1}

Newton Method with Line Search



Bean function:

$$f(\mathbf{x}) = (1 - x_1)^2 + (1 - x_2)^2 + 0.5(2x_2 - x_1^2)^2$$

- The Newton decrement is that it is independent of linear changes in the problem variables.
- The backtracking line search especially suitable for a Newton algorithm is that as the iterates approach the solution proin, α_k approaches unity, the quadratic approximation used as the basis fo the newton method becomes increasingly accurate.

Equality Constrained Minimized

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

where $f(\mathbf{x})$ is a convex twice continuously differentiable function and $\mathbf{A}$ is of full row rank. The KKT condition is

$$\begin{aligned} \nabla f(\mathbf{x}) + \mathbf{A}^T \boldsymbol{\lambda} = 0 \\ \mathbf{A}\mathbf{x} = \mathbf{b} \end{aligned}$$

are satisfied for some Lagrange multiplier $\boldsymbol{\lambda}$.

- Assume $\mathbf{x}^*, \boldsymbol{\lambda}^*$ exist, then $\mathbf{x}_k, \boldsymbol{\lambda}_k$ satisfies $\mathbf{A}\mathbf{x} = \mathbf{b}$.

$$\begin{aligned} \mathbf{x}_{k+1} &= \mathbf{x}_k + \mathbf{d}_k \\ \mathbf{A}\mathbf{x}_{k+1} &= \mathbf{A}\mathbf{x}_k + \mathbf{A}\mathbf{d}_k \quad \Rightarrow \quad \mathbf{A}\mathbf{d}_k = 0 \end{aligned}$$

Equality Constrained Minimized

- The linear approximation of the gradient is

$$\begin{aligned} \nabla f(\mathbf{x}_k + \mathbf{d}_k) &\approx \nabla f(\mathbf{x}_k) + \nabla^2 f(\mathbf{x}_k)\mathbf{d}_k \\ &= \nabla f(\mathbf{x}_k) + \mathbf{H}_k \mathbf{d}_k \end{aligned}$$

- KKT Condition becomes

$$\begin{aligned} \nabla f(\mathbf{x}_k) + \nabla^2 f(\mathbf{x}_k)\mathbf{d}_k + \mathbf{A}^T \boldsymbol{\lambda}_{k+1} &= 0 \\ \mathbf{A}\mathbf{d}_k &= 0 \end{aligned}$$

- Matrix Form:

$$\begin{bmatrix} \nabla^2 f(\mathbf{x}_k) & \mathbf{A}^T \\ \mathbf{A} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{d}_k \\ \boldsymbol{\lambda}_{k+1} \end{bmatrix} = \begin{bmatrix} -\nabla f(\mathbf{x}_k) \\ 0 \end{bmatrix}$$

Equality Constrained Minimized

- If there are no constraints, it becomes

$$\begin{aligned} \mathbf{H}_k \mathbf{d}_k &= -\nabla f(\mathbf{x}) \\ \mathbf{d}_k &= -\mathbf{H}_k^{-1} \nabla f(\mathbf{x}) \end{aligned}$$

The vector $\mathbf{d}_k$ satisfies the constraints is regarded as a Newton direction.

- We can calculate $\mathbf{d}_k$ as

$$\mathbf{d}_k = -\mathbf{H}_k^{-1} \left(\mathbf{A}^T \boldsymbol{\lambda}_{k+1} + \nabla f(\mathbf{x}_k) \right)$$

where

$$\begin{aligned} \mathbf{d}_k &= -\mathbf{H}_k^{-1} \mathbf{A}^T \boldsymbol{\lambda}_{k+1} - \mathbf{H}_k^{-1} \nabla f(\mathbf{x}_k) \\ \boldsymbol{\lambda}_{k+1} &= -\left(\mathbf{A} \mathbf{H}_k^{-1} \mathbf{A}^T \right) \mathbf{A} \mathbf{H}_k^{-1} \nabla f(\mathbf{x}_k) \\ \mathbf{A} \mathbf{d}_k &= 0 \end{aligned}$$

- The stop criterion can be formulate as $\|\nabla f(\mathbf{x}_{k+1}) + \mathbf{A}^T \boldsymbol{\lambda}_{k+1}\| < \varepsilon$

Equality Constrained Minimized

Newton Method with Linear Constraints and $\mathbf{x} \in$ dom $f(\mathbf{x})$

Require: a starting point $\mathbf{x} \in$ dom $f(\mathbf{x})$, $\boldsymbol{\lambda}$, tolerance $\varepsilon > 0$
$k = 0$
repeat
Calculate $\mathbf{d}_k$ and $\boldsymbol{\delta}_k$
Find α_k that $\min\text{imize} \quad f(\mathbf{x}_k + \alpha\mathbf{d}_k)$, using the backtracking line search
$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k\mathbf{d}_k$ and compute $\nabla f(\mathbf{x}_{k+1})$
$k = k + 1$
until $\|\nabla f(\mathbf{x}_{k+1}) + \mathbf{A}^T\boldsymbol{\lambda}_{k+1}\| \leq \varepsilon$
return x_{k+1}

Equality Constrained Minimized

- If $\mathbf{x}$ is not in **dom** $f(\mathbf{x})$, $(\mathbf{x}_k, \boldsymbol{\lambda}_k)$ is known but $\mathbf{x}_k$ is not feasible

$$(\mathbf{x}_k, \boldsymbol{\lambda}_k) \quad \rightarrow \quad (\mathbf{x}_{k+1}, \boldsymbol{\lambda}_{k+1})$$

  We must have $\mathbf{A}\mathbf{d}_k = 0 = -(\mathbf{A}\mathbf{x} - b)$

- The Newton direction must satisfy the equations:

$$\begin{bmatrix} \mathbf{H}_k & \mathbf{A}^T \\ \mathbf{A} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{d}_k \\ \boldsymbol{\lambda}_{k+1} \end{bmatrix} = \begin{bmatrix} -\nabla f(\mathbf{x}_k) \\ \mathbf{b} - \mathbf{A}\mathbf{x}_k \end{bmatrix}$$

- Here $\mathbf{A}\mathbf{d}_k \neq 0$ $\mathbf{d}_k$ is not necessarily a descent direction. The line search and the stop criterion are not applicable.
- Let use $\boldsymbol{\lambda}_{k+1} = \boldsymbol{\lambda}_k + \boldsymbol{\delta}_k$

$$\begin{bmatrix} \mathbf{H} & \mathbf{A}^T \\ \mathbf{A} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{d}_k \\ \boldsymbol{\delta}_k \end{bmatrix} = - \begin{bmatrix} \mathbf{r}_d \\ \mathbf{r}_p \end{bmatrix}$$

Equality Constrained Minimized

- $\mathbf{r}_d$ and $\mathbf{r}_p$ is called **dual** and **primal residual**

$$\mathbf{r}_d = \nabla f(\mathbf{x}_k) + \mathbf{A}^T \boldsymbol{\lambda}_k$$
$$\mathbf{r}_p = \mathbf{A}\mathbf{x}_k - \mathbf{b}$$

- We have

$$\mathbf{d}_k = -\mathbf{H}_k^{-1}\left(\mathbf{A}^T\boldsymbol{\delta}_k + \mathbf{r}_d\right)$$
$$\boldsymbol{\delta}_k = -\left(\mathbf{A}\mathbf{H}_k^{-1}\mathbf{A}^T\right)^{-1}\left(\mathbf{A}\mathbf{H}_k^{-1}\mathbf{r}_d - \mathbf{r}_p\right)$$

$$\|\mathbf{r}(\mathbf{x}_k, \boldsymbol{\lambda}_k)\| \le \varepsilon$$

- The backtracking line search can be found by

$$\text{minimize}\, \|\mathbf{r}(\mathbf{x}_k + \alpha_k\mathbf{d}_k, \boldsymbol{\lambda}_k + \alpha_k\boldsymbol{\delta}_k\|$$

Equality Constrained Minimized

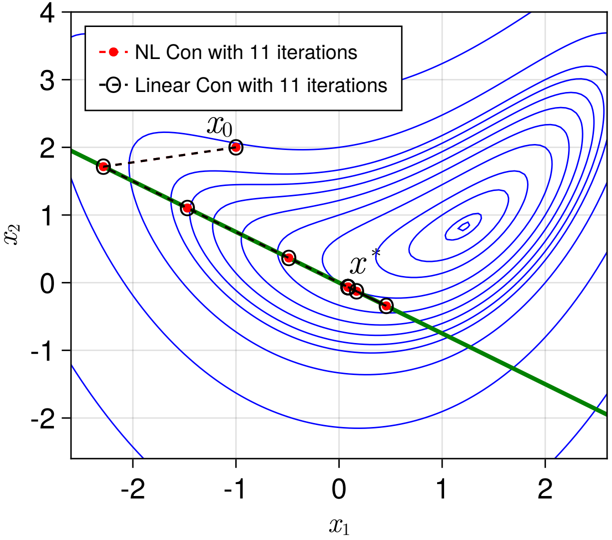
Newton Method with Linear Constraints and $\mathbf{x}$ is not in dom $f(\mathbf{x})$

Require: a starting point $\mathbf{x}, \boldsymbol{\lambda}$, tolerance $\varepsilon > 0$
  $k = 0$
  repeat
    Calculate $\mathbf{d}_k$ and $\boldsymbol{\lambda}_{k+1}$
    Find α_k using backtracking that minimizes $\|\mathbf{r}(\mathbf{x}_k + \alpha\mathbf{d}_k, \boldsymbol{\lambda}_k + \alpha\boldsymbol{\delta}_k)\|$
    $x_{k+1} = x_k + \alpha_k\mathbf{d}_k$
    $k = k + 1$
  until $\|\mathbf{r}(\mathbf{x}_k, \boldsymbol{\lambda}_k)\| < \varepsilon$
  return x_{k+1}

Modified Backtracking

Require: $\gamma \in (0, 0.5)$ and $\rho \in (0, 1)$, $\lambda = 1$
  while $\|\mathbf{r}(\mathbf{x}_k + \alpha\mathbf{d}_k, \boldsymbol{\lambda}_k + \alpha\boldsymbol{\delta}_k\| > (1 - \rho\alpha)\|\mathbf{r}(\mathbf{x}_k, \boldsymbol{\lambda}_k)\|$ **do**
    $\alpha = \gamma\alpha$
  end while
  return x_{k+1}

Equality Constrained Minimized



Nonlinear Equality Constrained Minimized

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & f(\mathbf{x}), \qquad f(\mathbf{x}) : \mathbb{R}^n \mapsto \mathbb{R} \\ \text{subject to} \quad & c(\mathbf{x}) = 0, \qquad c(\mathbf{x}) : \mathbb{R}^n \mapsto \mathbb{R}^m \end{aligned}$$

First-Order Necessary Conditions:

- $\nabla f(\mathbf{x}) = 0$ in free directions
- $c(\mathbf{x}) = 0$ for active constraints.
- Any non-zero component of $\nabla f(\mathbf{x})$ must be normal to the constraint surface/manifold

$$\nabla f(\mathbf{x}) + \lambda \nabla c(\mathbf{x}) = 0,$$

where λ is a Lagrange multiplier(dual variable).

Nonlinear Equality Constrained Minimized

- In general

$$\frac{\partial f}{\partial \mathbf{x}} + \boldsymbol{\lambda}^T \frac{\partial c}{\partial \mathbf{x}} = 0, \qquad \boldsymbol{\lambda} \in \mathbb{R}^m$$

- Based on this gradient condition, we defined:

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \boldsymbol{\lambda}^T c(\mathbf{x}),$$

  where $\mathcal{L}$ is called "Lagrangian"
- The KKT condition is

$$\nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = \nabla f(\mathbf{x}) + \left(\frac{\partial c}{\partial \mathbf{x}}\right)^T \boldsymbol{\lambda} = 0$$

$$\nabla_{\boldsymbol{\lambda}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = c(\mathbf{x}) = 0$$

Nonlinear Equality Constrained Minimized

We can solve the KKT condition with the Newton method:

$$\nabla_{\mathbf{x}}\mathcal{L}(\mathbf{x}+\Delta\mathbf{x},\boldsymbol{\lambda}+\Delta\boldsymbol{\lambda}) \approx \nabla_{\mathbf{x}}\mathcal{L}(\mathbf{x},\boldsymbol{\lambda}) + \frac{\partial^2\mathcal{L}}{\partial\mathbf{x}^2}\Delta\mathbf{x} + \frac{\partial^2\mathcal{L}}{\partial\mathbf{x}\partial\boldsymbol{\lambda}}\Delta\boldsymbol{\lambda} = 0$$

$$\nabla_{\boldsymbol{\lambda}}\mathcal{L}(\mathbf{x}+\Delta\mathbf{x},\boldsymbol{\lambda}+\Delta\boldsymbol{\lambda}) \approx c(\mathbf{x}) + \frac{\partial c}{\partial\boldsymbol{\lambda}}\Delta\mathbf{x} = 0$$

Note:

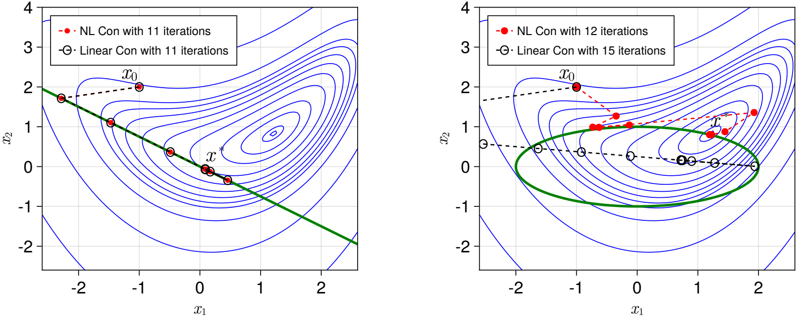
$$\frac{\partial}{\partial\mathbf{x}}\nabla_{\boldsymbol{\lambda}}\mathcal{L}(\mathbf{x},\boldsymbol{\lambda}) = \left(\frac{\partial c}{\partial\mathbf{x}}\right)^T, \qquad \frac{\partial c}{\partial\mathbf{x}}\Delta\mathbf{x} = -c(\mathbf{x})$$

The KKT system is

$$\begin{bmatrix} \frac{\partial^2\mathcal{L}}{\partial\mathbf{x}^2} & \left(\frac{\partial c}{\partial\mathbf{x}}\right)^T \\ \frac{\partial c}{\partial\boldsymbol{\lambda}} & 0 \end{bmatrix} \begin{bmatrix} \Delta\mathbf{x} \\ \Delta\boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} -\nabla_{\mathbf{x}}\mathcal{L}(\mathbf{x},\boldsymbol{\lambda}) \\ -c(\mathbf{x}) \end{bmatrix}$$

In the algorithm we have $\mathbf{x}_{k+1} = \mathbf{x}_k + \Delta\mathbf{x}$ and $\boldsymbol{\lambda}_{k+1} = \boldsymbol{\lambda}_k + \Delta\boldsymbol{\lambda}$.

Nonlinear Equality Constrained Minimized



When the equality constraints are nonlinear, it is obvious that the Newton with Linear constraint method cannot use.

Reference

1. Joaquim R. R. A. Martins, Andrew Ning, ”*Engineering Design Optimization*,” Cambridge University Press, 2021.
2. Mykel J. kochenderfer, and Tim A. Wheeler, ”*Algorithms for Optimization*,” The MIT Press, 2019.
3. Ashok D. Belegundu, Tirupathi R. Chandrupatla, ”Optimization Concepts and Applications in Engineering,” Cambridge University Press, 2019.
4. Stephen Boyd, and Lieven Vandenberghe , ”*Convex Optimization*,” Cambridge University Press, 2009.