Lecture 15. Newton's Method with Equality Constraints

15.1 Lagrangian function

Consider an optimization problem

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathbb{R}^n} \quad & f(\boldsymbol{x}) \\ \text{subject to} \quad & g_i(\boldsymbol{x}) = 0\,, \quad \forall\, i = 1, 2, \ldots, m\,. \end{aligned}$$

We can reformulate the Lagrange condition in a compact form. We define the *Lagrangian function* as

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \boldsymbol{\lambda}^\mathsf{T} \boldsymbol{g}(\boldsymbol{x}) = f(\boldsymbol{x}) + \lambda_1 g_1(\boldsymbol{x}) + \cdots + \lambda_m g_m(\boldsymbol{x})\,.$$

The Lagrange condition shows that if $\boldsymbol{x}^*$ is regular and local minimum, then there exists $\boldsymbol{\lambda}^*$ such that

$$\nabla f(\boldsymbol{x}^*) + (\boldsymbol{\lambda}^*)^\mathsf{T} \boldsymbol{g}'(\boldsymbol{x}^*) = \nabla f(\boldsymbol{x}^*) + \lambda_1^* \nabla g_1(\boldsymbol{x}^*) + \cdots + \lambda_m^* \nabla g_m(\boldsymbol{x}^*) = \mathbf{0}\,.$$

Given the definition of Lagrangian, we can rewrite it as

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

Note that

$$\nabla_{\boldsymbol{\lambda}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = \left(\frac{\partial \mathcal{L}}{\partial \lambda_1}, \ldots, \frac{\partial \mathcal{L}}{\partial \lambda_m}\right)^\mathsf{T} = \big(g_1(\boldsymbol{x}), \ldots, g_m(\boldsymbol{x})\big)^\mathsf{T} = \mathbf{0}\,,$$

if $\boldsymbol{x}$ is feasible. So we can simplify the Lagrange condition as $\nabla \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \mathbf{0}$.

Conversely, for convex optimization problems, the Lagrange condition (combining with the feasibility) is sufficient for optimal solutions, which is exactly $\nabla_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \mathbf{0}$ and $\nabla_{\boldsymbol{\lambda}} \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \mathbf{0}$. Thus, we have the following sufficient and necessary condition for convex optimizations.

Theorem (*Lagrange condition, expressed by Lagrangian*)

For any point $\boldsymbol{x}^* \in \mathbb{R}^n$, it is optimal for the convex optimization problem if and only if there exists $\boldsymbol{\lambda}^* \in \mathbb{R}^m$ such that

$$\nabla \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \mathbf{0} \,.$$

Note that $\mathcal{L}$ is not convex in general (see e.g., $f(x) = x^2$, $g(x) = x - 5$, and thus $\mathcal{L}(x, \lambda) = x^2 + \lambda(x - 5)$), although $\mathcal{L}$ is a convex function of $\boldsymbol{x}$ for any fixed $\boldsymbol{\lambda}$. So $\nabla \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \mathbf{0}$ does not imply $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ is a minimum point of $\mathcal{L}$.
In fact, it is a *saddle point* in a sense. For feasible $\boldsymbol{x}$, i.e., $g_1(\boldsymbol{x}) = \cdots = g_m(\boldsymbol{x}) = 0$, it is easy to see that $f(\boldsymbol{x}) = \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda})$. For infeasible $\boldsymbol{x}$, there exists $g_i(\boldsymbol{x}) \neq 0$, so we have $\max_{\boldsymbol{\lambda}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) \to \infty$. Therefore, we conclude that

$$\mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \min_{\boldsymbol{x}} \max_{\boldsymbol{\lambda}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) \,.$$

Moreover, for *convex optimization problems*, the other direction also holds, namely, we have the following proposition.

Theorem

For the Lagrangian $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda})$ of a convex optimization problem, if $\nabla \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \mathbf{0}$, then

$$f(\boldsymbol{x}^*) = \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \max_{\boldsymbol{\lambda}} \min_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) \,.$$

Proof

For the first equality, note that if $\nabla \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \mathbf{0}$, $g_1(\boldsymbol{x}^*) = \cdots = g_m(\boldsymbol{x}^*) = 0$. Thus $f(\boldsymbol{x}^*) = \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$.
Now we define $\tilde{\mathcal{L}}(\boldsymbol{\lambda}) = \min_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda})$. Our goal is to show that $\mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \max_{\boldsymbol{\lambda}} \tilde{\mathcal{L}}(\boldsymbol{\lambda})$.
We first show that there exists $\tilde{\boldsymbol{\lambda}}$ such that $\mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \tilde{\mathcal{L}}(\tilde{\boldsymbol{\lambda}})$. In particular, we show that $\mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \tilde{\mathcal{L}}(\boldsymbol{\lambda}^*)$. Note that $\tilde{\mathcal{L}}(\boldsymbol{\lambda}^*) = \min_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}^*)$. Since $\boldsymbol{\lambda}^*$ is fixed, $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}^*)$ is a convex function of $\boldsymbol{x}$. By the first order condition for convexity, $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}^*)$ achieves the minimum value if and only if $\nabla_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}^*) = \mathbf{0}$. Since $\nabla_{\boldsymbol{x}^*} \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \mathbf{0}$, we have $\tilde{\mathcal{L}}(\boldsymbol{\lambda}^*) = \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$.
Next, we show that for all $\boldsymbol{\lambda} \in \mathbb{R}^m$, $\tilde{\mathcal{L}}(\boldsymbol{\lambda}) \leq \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$. Denote by Ω the feasible set $\{\boldsymbol{x} \mid g_1(\boldsymbol{x}) = g_2(\boldsymbol{x}) = \cdots = g_m(\boldsymbol{x}) = 0\}$. Then for all $\boldsymbol{x} \in \Omega$ and all $\boldsymbol{\lambda}$,

$f(\boldsymbol{x}) = \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda})$. So for all $\boldsymbol{\lambda} \in \mathbb{R}^m$,

$$\tilde{\mathcal{L}}(\boldsymbol{\lambda}) = \min_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) \leq \min_{\boldsymbol{x} \in \Omega} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = \min_{\boldsymbol{x} \in \Omega} f(\boldsymbol{x}) = f(\boldsymbol{x}^*) = \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) \,.$$

Later, we will revisit this theorem in the context of general convex optimization with inequality constraints.

15.2 Karush-Kuhn-Tucker System

We now consider how to solve convex optimization problems with equality constraints. Recall the problems with no constraints, where we approximate the objective function by a quadratic function and minimize the quadratic function. We would like to apply the same idea. So the first step is to minimize quadratic functions with equality constraints.

Given $\boldsymbol{Q} \in \mathbb{R}^{n \times n} \succeq \boldsymbol{0}$, $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, $\boldsymbol{w} \in \mathbb{R}^n$ and $\boldsymbol{b} \in \mathbb{R}^m$, consider the following quadratic problem with equality constrains

$$\begin{aligned} \min_{\boldsymbol{x} \in \mathbf{R}^n} \quad & \frac{1}{2} \boldsymbol{x}^\mathsf{T} \boldsymbol{Q} \boldsymbol{x} + \boldsymbol{w}^\mathsf{T} \boldsymbol{x} \\ \text{subject to} \quad & \boldsymbol{A} \boldsymbol{x} = \boldsymbol{b} \,. \end{aligned}$$

We can solve this problem by the Lagrange multiplier method. The Lagrangian function is

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = \frac{1}{2} \boldsymbol{x}^\mathsf{T} \boldsymbol{Q} \boldsymbol{x} + \boldsymbol{w}^\mathsf{T} \boldsymbol{x} + \boldsymbol{\lambda}^\mathsf{T} (\boldsymbol{A} \boldsymbol{x} - \boldsymbol{b}) \,,$$

and the Lagrange condition is

$$\begin{cases} \nabla_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = \boldsymbol{Q} \boldsymbol{x} + \boldsymbol{w} + \boldsymbol{A}^\mathsf{T} \boldsymbol{\lambda} = \boldsymbol{0} \\ \nabla_{\boldsymbol{\lambda}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = \boldsymbol{A} \boldsymbol{x} - \boldsymbol{b} = \boldsymbol{0} \end{cases} .$$

We know that $\boldsymbol{x}^*$ is a global minimum if and only $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ is a solution to the above system of equations for some $\boldsymbol{\lambda}^*$. The above system of equations can be expressed in the following matrix form:

$$\underbrace{\begin{pmatrix} \boldsymbol{Q} & \boldsymbol{A}^\mathsf{T} \\ \boldsymbol{A} & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{\lambda} \end{pmatrix} = \begin{pmatrix} -\boldsymbol{w} \\ \boldsymbol{b} \end{pmatrix}}_{\text{KKT system}} ,$$

which is called the *KKT system* and the coefficient matrix $\begin{pmatrix} \boldsymbol{Q} & \boldsymbol{A}^\top \\ \boldsymbol{A} & \mathbf{0} \end{pmatrix}$ is called the *KKT matrix*.

Block Gaussian elimination

If we further assume $\boldsymbol{Q}$ is invertible (nonsingular), we can solve the KKT system by *Gaussian elimination*. Because $\boldsymbol{Q} \succeq \mathbf{0}$, it is equivalent to $\boldsymbol{Q} \succ \mathbf{0}$.

Left multiplying $\boldsymbol{A}\boldsymbol{Q}^{-1}$ to the first row and subtracting the second row, we obtain that

$$\begin{pmatrix} \mathbf{0} & \boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^\top \end{pmatrix} \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{\lambda} \end{pmatrix} = -\boldsymbol{b} - \boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{w} \,.$$

So

$$\boldsymbol{\lambda} = -\left(\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^\top\right)^{-1} \left(\boldsymbol{b} + \boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{w}\right) .$$

Plugging it into the first row of the original KKT system

$$\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{A}^\top\boldsymbol{\lambda} = -\boldsymbol{w} \,,$$

we have

$$\boldsymbol{x} = -\boldsymbol{Q}^{-1}\boldsymbol{w} + \boldsymbol{Q}^{-1}\boldsymbol{A}^\top\left(\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^\top\right)^{-1} \left(\boldsymbol{b} + \boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{w}\right) .$$

Remark

Both $\boldsymbol{Q}^{-1}$ and $\left(\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^\top\right)^{-1}$ must exist in the above calculation. If we assume $\boldsymbol{Q} \succ 0$ and $\operatorname{rank}(\boldsymbol{A}) = m$ ($\boldsymbol{A}$ has full row rank), then both of them exist.
First, $\boldsymbol{Q} \succ \mathbf{0}$ implies that $\boldsymbol{Q}$ is invertible. Next, if $\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^\top$ is not invertible, there exists $\boldsymbol{v} \in \mathbb{R}^m \neq \mathbf{0}$ such that $\left(\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^\top\right)\boldsymbol{v} = \mathbf{0}$. If we left multiple $\boldsymbol{v}^\top$ to both sides, we can get

$$\boldsymbol{v}^\top \left(\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^\top\right)\boldsymbol{v} = (\boldsymbol{A}^\top\boldsymbol{v})^\top\boldsymbol{Q}^{-1}(\boldsymbol{A}^\top\boldsymbol{v}) = 0 \,.$$

Since $\operatorname{rank}(\boldsymbol{A}) = m$, its rows are linear independent. So $\boldsymbol{A}^\top\boldsymbol{v} \neq \mathbf{0}$, and $(\boldsymbol{A}^\top\boldsymbol{v})^\top\boldsymbol{Q}^{-1}(\boldsymbol{A}^\top\boldsymbol{v}) = 0$ contradicts to the fact that $\boldsymbol{Q}^{-1} \succ \mathbf{0}$.
The hypothesis of $\operatorname{rank}(\boldsymbol{A}) = m$ is reasonable. Otherwise either the problem is infeasible (e.g., constraints are $x_1 + x_2 = 2$ and $2x_1 + 2x_2 = 3$), or there are redundant constraints (e.g., constraints are $x_1 + x_2 = 2$ and $2x_1 + 2x_2 = 4$).

But the assumption of $\boldsymbol{Q} \succ 0$ is strong. Actually, we only need the KKT matrix to be invertible.

Nonsingularity of KKT matrices

Let $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x} + \boldsymbol{w}^\top \boldsymbol{x}$ where $\boldsymbol{Q} \succeq \mathbf{0}$. If there is no constraints, when does f attains it minimum value? The gradient is $\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{w}$. So there exists an optimal solution if and only if the equation $\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{w} = \mathbf{0}$ has solutions, namely, $\boldsymbol{w}$ is in the row space of $\boldsymbol{Q}$. Recall the lemma $\operatorname{im}(\boldsymbol{A}^\top) = \ker(\boldsymbol{A})^\perp$. It is further equivalent to $\boldsymbol{w} \perp \ker(\boldsymbol{Q})$, i.e.,

$$\forall \boldsymbol{v} \in \mathbb{R}^n, \qquad \boldsymbol{Q}\boldsymbol{v} = \mathbf{0} \implies \langle \boldsymbol{w}, \boldsymbol{v} \rangle = 0 .$$

There are two cases:

1. if $\boldsymbol{Q}$ is invertible / nonsingular, there exists a unique solution $\boldsymbol{x}$;
2. if $\boldsymbol{Q}$ is not invertible, there are infinite many solutions.

Now let $\boldsymbol{K}$ be the KKT matrix. Clearly the KKT system is solvable if and only if $\begin{pmatrix} -\boldsymbol{w} \\ \boldsymbol{b} \end{pmatrix} \perp \ker(\boldsymbol{K})$.

We first prove the following useful proposition.

Lemma

Suppose $Q \succeq 0$, then for any vector x, $Qx = \mathbf{0}$ if and only if $x^\top Q x = 0$.

Proof

- Necessity: Trivial.
- Sufficiency: Since $Q \succeq 0$, we can do eigendecomposition to Q.

$$Q = U\Lambda U^\top = [\xi_1 u_1, \dots, \xi_n u_n] \cdot \begin{pmatrix} u_1^\top \\ u_2^\top \\ \cdots \\ u_n^\top \end{pmatrix} = \sum_{i=1}^n \xi_i u_i u_i^\top,$$

  where $\xi_i \geq 0$ is the i-th eigenvalue and u_i is the i-th eigenvector of Q.

Then we have

$$x^\top Q x = \sum_{i=1}^{n} \xi_i (u_i^\top x)^2 = 0,$$

which means either $\xi_i = 0$ or $u_i^\top x = 0$. Hence,

$$Qx = \sum_{i=1}^{n} u_i (\xi_i (u_i^\top x)) = \mathbf{0}\,.$$

Now we can show that $\ker(\boldsymbol{K}) = \{(\boldsymbol{v}, \mathbf{0})^\top \mid \boldsymbol{v} \in \ker(\boldsymbol{Q}) \cap \ker(\boldsymbol{A})\}$. Note that if $\boldsymbol{v} \in \ker(\boldsymbol{Q}) \cap \ker(\boldsymbol{A})$, then

$$\begin{pmatrix} \boldsymbol{Q} & \boldsymbol{A}^\top \\ \boldsymbol{A} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{v} \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} -\boldsymbol{Q}\boldsymbol{v} \\ \boldsymbol{A}\boldsymbol{v} \end{pmatrix} = \mathbf{0}\,.$$

Conversely, we claim that if

$$\begin{pmatrix} Q & A^\top \\ A & \mathbf{0} \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \mathbf{0}\,,$$

then $u \in \ker(Q) \cap \ker(A)$ and $v = \mathbf{0}$. This linear equation system is equivalent to $Qu + A^\top v = \mathbf{0}$ and $Au = \mathbf{0}$. Note that by left multiplying $u^\top$, we obtain that

$$u^\top Q u = u^\top(-A^\top v) = -(Au)^\top v = 0$$

So $Qu = \mathbf{0}$ by the above lemma. Moreover, $A^\top v = -Qu = \mathbf{0}$, contradicts $\operatorname{rank}(A) = m$ if $v \neq \mathbf{0}$.

Then we focus on the case where $\boldsymbol{K}$ is invertible / nonsingular.

Theorem

KKT matrix is invertible (nonsingular) is equivalent to each one of the followings:

1. $\ker(A) \cap \ker(Q) = \{\mathbf{0}\}$, or
2. if $Ax = 0$ and $x \neq 0$, then $x^\top Q x > 0$, or
3. $\forall F \in \mathbb{R}^{n\times(n-m)}$, if $\operatorname{im}(F) \triangleq \{Fv : v \in \mathbb{R}^{n-m}\} = \ker(A)$, then $F^\top Q F \succ 0$.

Proof

- "invertible $\implies$ 1". If there exists $x \neq \mathbf{0}$ such that $x \in \ker(Q) \cap \ker(A)$, we have

$$\begin{pmatrix} Q & A^\top \\ A & \mathbf{0} \end{pmatrix} \cdot \begin{pmatrix} x \\ \mathbf{0} \end{pmatrix} = \mathbf{0},$$

  which means the KKT matrix is not invertible.
- "1 $\implies$ 2". If $Ax = 0$ and $x \neq 0$, then $x \in \ker(A) \setminus \{\mathbf{0}\}$. So $Qx \neq 0$. Thus $x^\top Qx \neq 0$ by the above lemma.
- "2 $\implies$ invertible". By the above lemma, item 2 is equivalent to if $Ax = 0$ and $x \neq 0$ then $Qx \neq \mathbf{0}$. So $\ker(A) \cap \ker(Q) = \{\mathbf{0}\}$. Thus $\ker(\boldsymbol{K}) = \{\mathbf{0}\}$, which implies that $\boldsymbol{K}$ is invertible.
- "2 $\implies$ 3". If $\mathrm{im}(F) = \ker(A)$, for all $x \neq \mathbf{0}$, $AFx = 0$. Note that $\dim \mathrm{im}(F) = \dim \mathrm{rank}(A) = n - m$. So $\mathrm{rank}(F) = n - m$. Thus $Fx \neq \mathbf{0}$ by linear independence. If $Fx \neq \mathbf{0}$ then $(Fx)^\top Q(Fx) = x^\top (F^\top QF)x > 0$. So $F^\top QF \succ 0$.
- "3 $\implies$ 2". If $\mathrm{im}(F) = \ker(A)$, for all $x \neq \mathbf{0}$ that $Ax = \mathbf{0}$, there exists $y \neq \mathbf{0}$ such that $x = Fy$. Since $F^\top QF \succ 0$, $x^\top Qx = y^\top F^\top QFy > 0$.

15.3 Newton's method

Now we consider how to solve general convex optimization with equality constraints.

Recall Newton's method. Given x_k, we do Taylor Expansion of $f(x)$ at x_k:

$$f(x) \approx \tilde{f}(x) \triangleq f(x_k) + \nabla f(x_k)^\top (x - x_k) + \frac{1}{2}(x - x_k)^\top \nabla^2 f(x_k)(x - x_k)$$

- If there is no constraints, we use $x_{k+1} = \arg\min \tilde{f}(x)$ to approximate the minimum point of f.
- If there are constraints $Ax - b = 0$, we may use $x_{k+1} = \underset{Ax-b=\mathbf{0}}{\arg\min} \tilde{f}(x)$ to approximate the minimum point of f under constraints $Ax - b = \mathbf{0}$.

Let $d = x - x_k$. We have

$$\tilde{f}(x) = f(x_k) + \nabla f(x_k)^\top d + \frac{1}{2} d^\top \nabla^2 f(x_k) d$$

and the constraints become $Ad = \mathbf{0}$ since

$$0 = A(x_k + d) - b = Ax_k + Ad - b\,.$$

If x_k is a feasible solution, we have $Ax_k - b = \mathbf{0}$, hence $Ad = \mathbf{0}$.
Now for d, the problem becomes

$$\underset{Ad=\mathbf{0}}{\arg\min} \tilde{f}(x) = f(x_k) + \nabla f(x_k)^\mathsf{T} d + \frac{1}{2} d^\mathsf{T} \nabla^2 f(x_k) d$$

and finally

$$x_{k+1} = \underset{Ax-b=\mathbf{0}}{\arg\min} \tilde{f}(x) = x_k + \underset{Ad=\mathbf{0}}{\arg\min} f(x_k) + \nabla f(x_k)^\mathsf{T} d + \frac{1}{2} d^\mathsf{T} \nabla^2 f(x_k) d\,.$$

The KKT system for optimizing d is

$$\begin{pmatrix} \nabla^2 f(x_k) & A^\mathsf{T} \\ A & \mathbf{0} \end{pmatrix} \begin{pmatrix} d \\ \lambda \end{pmatrix} = \begin{pmatrix} -\nabla f(x_k) \\ 0 \end{pmatrix}$$

ALGORITHM 1: Formal Definition of Newton's Method

```
Let x_0 be an arbitrary feasible solution ;
while d^T ∇^2 f(x_k) d ≤ δ do
    Compute d by solving KKT system;
    x_{k+1} ← x_k + d;
return {x_k}
```

Note that we use $d^\mathsf{T} \nabla^2 f(x_k) d \le \delta$ as stopping criteria instead of $\|\nabla f(x_k)\| < \delta$. This is because, when the algorithm should terminate, its gradient is not zero (it needs to follow the Lagrange condition).

Next, let us see some examples of using Newton's method.

Example 1

$$\begin{aligned} \min \quad & f(x_1, x_2) = x_1^2 + x_2^2 \\ s.t. \quad & x_1 + x_2 = 1 \end{aligned}$$

Start from $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$. The KKT system at this point is

$$\begin{pmatrix} 2 & 0 & 1 \\ 0 & 2 & 1 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \\ \lambda \end{pmatrix} = \begin{pmatrix} -2 \\ 0 \\ 0 \end{pmatrix}.$$

Its solution is $(d_1, d_2, \lambda) = \left(-\frac{1}{2}, \frac{1}{2}, -1\right)$. The next point x_1 is exactly the optimal solution $\left(\frac{1}{2}, \frac{1}{2}\right)$.

Example 2

$$\begin{array}{ll} \min & f(x_1, x_2) = x_1^2 \\ s.t. & x_1 + 2x_2 = b \end{array}$$

Start from $\begin{pmatrix} b \\ 0 \end{pmatrix}$. The KKT system at this point is

$$\begin{pmatrix} 2 & 0 & 1 \\ 0 & 0 & 2 \\ 1 & 2 & 0 \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \\ \lambda \end{pmatrix} = \begin{pmatrix} -2b \\ 0 \\ 0 \end{pmatrix}.$$

Its solution is $(d_1, d_2, \lambda) = \left(-b, \frac{b}{2}, 0\right)$. The next point x_1 is also exactly the optimal solution $\left(0, \frac{b}{2}\right)$.
In this example, although $\nabla^2 f$ is not invertible, the KKT matrix is still invertible.
We may check by the criterion introduced above, since $\ker(\nabla^2 f) = \left\{ s \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix} : s \in \mathbb{R} \right\}$ and $\ker(A) = \left\{ t \cdot \begin{pmatrix} -2 \\ 1 \end{pmatrix} : t \in \mathbb{R} \right\}$, which satisfy the previous condition of $\ker(A) \cap \ker(\nabla^2 f) = \{0\}$.

Example 3

$$\begin{array}{ll} \min & f(x_1, x_2) = e^{x_1^2 + x_2^2} \\ s.t. & x_1 + x_2 = 1 \end{array}$$

Start from $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$. The KKT system at this point is

$$\begin{pmatrix} 6e & 0 & 1 \\ 0 & 2e & 1 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \\ \lambda \end{pmatrix} = \begin{pmatrix} -2e \\ 0 \\ 0 \end{pmatrix}.$$

Its solution is $(d_1, d_2, \lambda) = \left(-\frac{1}{4}, \frac{1}{4}, -\frac{e}{2}\right)$. The next point x_1 is $\left(\frac{3}{4}, \frac{1}{4}\right)$. It is not the optimal solution, but it is also in the right direction.

15.4 Correctness and convergence

We will demonstrate the correctness of Newton's method via the following two propositions:

1. Each time we choose a descending direction: $\nabla f(x_k)^\top d \leq 0$.
2. The stopping criteria is correct: if $d^\top \nabla^2 f(x_k) d = 0$ then x_k is optimal.

Proposition 1

$\nabla f(x_k)^\top d \leq 0$.

Proof

The KKT system can be unfolded as follows:

$$\begin{cases} \nabla^2 f(x_k) d + A^\top \lambda = -\nabla f(x_k) \\ Ad = 0 \end{cases}$$

Use $d^\top$ to left multiple the first line, we can get

$$d^\top \nabla^2 f(x_k) d + \underbrace{d^\top A^\top \lambda}_{=0} = -d^\top \nabla f(x_k)\,,$$

which yields

$$d^\top \nabla f(x_k) = -d^\top \nabla^2 f(x_k) d \leq 0\,.$$

Proposition 2

If $d^\top \nabla^2 f(x_k) d = 0$ then x_k is optimal.

Proof

First, $d^\top \nabla^2 f(x_k) d = 0 \Leftrightarrow \nabla^2 f(x_k) d = \mathbf{0}$ (due to the lemma proved in Section 15.2). Then by the first equality in KKT system,

$$\underbrace{\nabla^2 f(x_k) d}_{=0} + A^\top \lambda = -\nabla f(x_k) \implies \nabla f(x_k) + A^\top \lambda = 0 \implies x_k \text{ is optimal.}$$

Now we analyzed the convergence of Newton's method with equality constraints. In fact, we can convert equality constrained problems to problems without constraint. For example, the following problem

$$\begin{aligned} \min \quad & f(x_1, x_2) \\ s.t. \quad & x_1 + x_2 = 1 \end{aligned}$$

is equivalent to $\min\ f(x_1, 1 - x_1)$.

In general, for an equality-constraintd problem, assume its feasible set is $\Omega = \{x \mid Ax = b\}$. It can be rewritten as $\Omega = \tilde{x} + \{Fz \mid z \in \mathbb{R}^{n-m}\}$ for some $\tilde{x} \in \Omega$ and $F \in \mathbb{R}^{n \times (n-m)}$, since Ω is an affine set.

Then the original problem is equivalent to the following one:

$$\min_{z \in \mathbb{R}^{n-m}} \quad g(z) \triangleq f\left(\tilde{x} + Fz\right)$$

Applying Newton's method to $g(z)$, we will get

$$\begin{aligned} z_{k+1} &= z_k - \left(\nabla^2 g(z_k)\right)^{-1} \nabla g(z_k) \\ &= z_k - \left(\nabla\left(F^\top \nabla f(\tilde{x} + Fz_k)\right)\right)^{-1} \nabla g(z_k) \\ &= z_k - \left(F^\top \nabla^2 f(\tilde{x} + Fz_k) F\right)^{-1} \nabla (F^\top \nabla f(\tilde{x} + Fz_k)) \end{aligned}$$

Assuming $x_0 = \tilde{x} + Fz_0$, we have the following proposition, which shows that the Newton’s method is *affinely invariant*.

Lemma

For all $k \geq 1$, $x_k = \tilde{x} + Fz_k$.

Proof

By induction, assume $x_k = \tilde{x} + Fz_k$. Let $d_{x_k} = x_{k+1} - x_k$ and $d_{z_k} = z_{k+1} - z_k$.

For all $v \in \mathbb{R}^n$, if $Av = \mathbf{0}$ then $\tilde{x} + v \in \Omega$, which implies $v \in \mathrm{im}(F)$, and vice versa. So $\mathrm{im}(F)$ is exactly $\ker(A)$.

Note that $Ad_{x_k} = 0$ by Newton's method. Thus there exists $u \in \mathbb{R}^{n-m}$ such that $d_{x_k} = Fu$.

Moreover, by the first equalty of KKT system, we know that

$$\nabla^2 f(x_k) d_{x_k} + A^\top \lambda = -\nabla f(x_k) \Rightarrow \nabla^2 f(x_k) F u + A^\top \lambda = -\nabla f(x_k)\,.$$

Left multiply $F^\top$ on both sides,

$$F^\top \nabla^2 f(x_k) F u + F^\top A^\top \lambda = -F^\top \nabla f(x_k)$$

Since $\mathrm{im}(F)$ is $\ker(A)$, we have $AFz = 0$ for all z, which implies that $AF = \mathbf{0}$.

Then, we have

$$F^\top \nabla^2 f(x_k) F u = -F^\top \nabla f(x_k)\,,$$

which implies that

$$u = -\left(F^\top \nabla^2 f(x_k) F\right)^{-1} \left(F^\top \nabla f(x_k)\right) = -\left(\nabla^2 g(z_k)\right)^{-1} \nabla g(z_k) = d_{z_k}\,.$$

Hence, $d_{x_k} = Fu = Fd_{z_k}$ and

$$x_{k+1} = x_k + d_{x_k} = x_k + Fd_{z_k} = \tilde{x} + Fz_{k+1}.$$

Therefore, the convergence of $\{z_k\}$ can lead to the convergence of $\{x_k\}$.