7

CONSTRAINED NONLINEAR OPTIMIZATION

7.1 Quadratic optimization I: equality constraints

7.1.1 Problem formulation

In **equality constrained quadratic optimization**, the optimization problem is given by

$$\min_{x \in \mathbb{R}^n} f(x) = c^T x + \frac{1}{2} x^T H x, \ \text{subject to } Ax = b,$$

where $c \in \mathbb{R}^n$, $H \in \mathbb{R}^{n \times n}$ is symmetric, and $A \in \mathbb{R}^{m \times n}, m \leq n, rank(A) = m, b \in \mathbb{R}^M$. We allow H to be either positive semi-definite or negative definite.

Example 7.1.1.

$$\begin{cases} \max_{x_1, x_2 \in \mathbb{R}} & x_1 x_2 \\ \text{subject to} & x_1 + x_2 = 100 \end{cases}$$

Example 7.1.2.

$$\begin{cases} \max_{x_1, x_2, x_3 \in \mathbb{R}} & x_1^2 + x_1 x_2 + x_2^2 + x_3 \\ \text{subject to} & x_1 + x_2 + x_3 = 100 \\ & 2x_2 - x_3 = 10 \end{cases}$$

7.1.2 Optimality condition

7.1.2.1 *General case*

Similar to unconstrained optimization, the first optimality condition can be derived by requiring objective function is non-decreasing on the local tangent space defined by the constraints $Ax = b$.

Theorem 7.1.1 (first order necessary optimality condition, KKT condition). *Consider $x^* \in \mathbb{R}^n$ satisfying feasibility condition $Ax^* = b$. If x^* is a local minimizer of the equality constrained quadratic optimization, then there exists a vector $y^* \in \mathbb{R}^m$ such that*

$$\nabla f(x^*) = A^T y^*.$$

Since $\nabla f(x^*) = Hx^* + c$, we can also equivalently write the two conditions into the so-called ***KKT matrix****:*

$$\begin{pmatrix} H & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} x^* \\ -y^* \end{pmatrix} = \begin{pmatrix} -c \\ b \end{pmatrix}.$$

Proof. We use contradiction to prove. Suppose we cannot find any vector $y^* \in \mathbb{R}^m$ such that $g \triangleq \nabla f(x^*) = A^T y^*$, then we can conclude $\nabla f(x^*) \in \mathcal{R}(A^T)$. Note that we can decompose $g = g_R + g_N$ into a component g_R in $\mathcal{R}(A^T)$ and a **non-zero** component g_N in $\mathcal{N}(A)$. (This decomposition is guaranteed by rank-nullity decomposition theorem [Corollary 4.4.4.1]). Then we can find a nonzero vector $p \in \mathcal{N}(A)$ that $p^T g < 0$,

$$\nabla f(x^* + \alpha p) = f(x^*) + \alpha p^T g + O(\alpha^2) < f(x^*)$$

for sufficiently small α. This contradicts the fact that x^* is a local minimizer. □

We can view A^T as a matrix of column vectors that span the subspace perpendicular to the null space of A, which is the feasible movement for any iterate x. By specifying $\nabla f(x^*) = A^T y^*$, we are requiring the objective function has no tendency of increase or decrease when x move in allowed local region as defined by $A^T x = b$. As an illustration [Figure 7.1.1], we examine the optimality condition for $f(x_1, x_2) = x_1^2 + x_2^2$ under constraint $x_1 + x_2 = 1$. The KKT condition requires the gradient of f should align with the normal of the plane $x_1 + x_2 = 1$ at a local minimal.

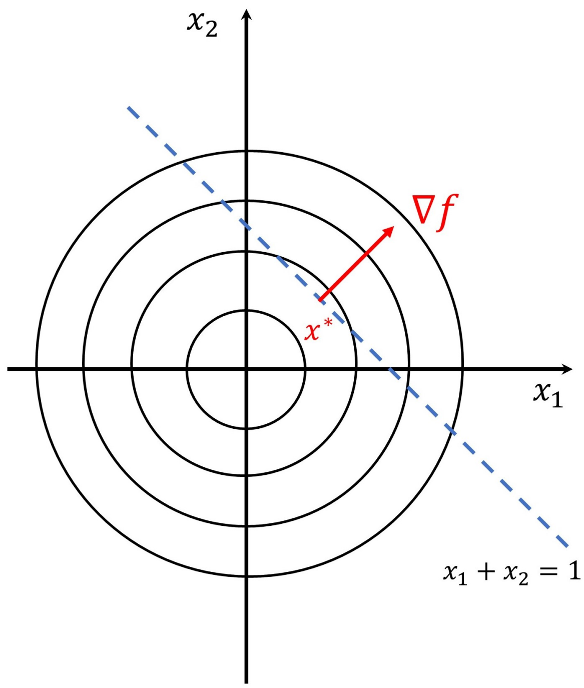

Figure 7.1.1: Demonstration of KKT condition at a local minimal for $f(x_1, x_2) = x_1^2 + x_2^2$ under constraint $x_1 + x_2 = 1$.

Lemma 7.1.1 (condition for nonsingular KKT matrix). *[1, lec 6] If A has full row rank and $Z^T HZ > 0$, where the columns of Z form a basis for the null space of A, then the KKT matrix*

$$K = \begin{pmatrix} H & A^T \\ A & 0 \end{pmatrix}$$

is nonsingular.

Proof. (a) Note that A full row rank is critical. Otherwise, KKT matrix will directly be rank deficient since

$$rank(KKT) = rank(H) + rank(A)$$

. (b) Suppose that

$$K \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

We want to show we must have $u = v = 0$. From above, we have

$$Au = 0, Hu + A^T v = 0$$

. Since Z columns span $\mathcal{N}(A)$, we have $u = Zw$. Multiply $Hu + A^T v = 0$ by u^T, we have

$$u^T Hu + u^T A^T v = w^T Z^T HZw = 0.$$

Since $Z^T HZ > 0$, we have $w = 0, u = Zw = 0$. Then,

$$Hu + A^T v = A^T v = 0 \implies v = 0,$$

since A^T has full column rank.

□

Based on the study of the reduced Hessian, we now have stronger sufficient conditions.

Theorem 7.1.2 (sufficiency of KKT condition in equality constrained quadratic optimization). *For the equality constraint quadratic optimization [Definition 7.3.1], we have the following possibilities:*

- $Z^T HZ > 0$. *There exists an unique minimizer from the unique solution of the KKT solution. Moreover, this unique minimizer must be an isolated/strict/global minimizer. Particularly, if H is positive definite, then $Z^T HZ > 0$.*
- $Z^T HZ \geq 0$, $Z^T HZ$ *is singular, KKT system is consistent but non-unique. Then, the quadratic programming has a unique minimum value but the minimizer is not unique.*
- $Z^T HZ$ *is not positive semidefinite. Then there exists a feasible ray upon which the objective function is unbounded blow.*

Proof. We can turn the constrained optimization problem into a unconstrained quadratic optimization, given by

$$\min_{y \in \mathbb{R}^{n-m}} c^T(x_0 + Zy) + \frac{1}{2}(x_0 + Zy)^T H(x_0 + Zy)$$

where $x_0 + Zy$,Z columns are basis of the $\mathcal{N}(A)$, is the solution to $Ax = b$. Then, we can easily get to the conclusions. □

7.1.2.2 *Positive semi-definitive quadratic programming*

If the objective function $f(x) = c^T x + \frac{1}{2} x^T Hx$ has H being positive semi-definitive, then we have more specific optimality conditions.

Lemma 7.1.2 (local minimum and global minimum in equality constrained quadratic programming). *Let x^* be a local minimizer of the equality constrained quadratic optimization problem.*

- *If H is positive semi-definite, then x^* is also a global minimizer.*
- *If H is positive definite, then the optimization problem only has at most one global minimizer and x^* is the only global minimizer.*

Proof. (1)Let x^* be a local minimum, suppose there is a point $x' \neq x^*$ being a global minimizer such that $f(x') < f(x^*)$. Then

$$f(x') = f(x^* + (x' - x^*)) = f(x^*) + \frac{1}{2}(x' - x^*)^T H(x' - x^*) \geq f(x*),$$

which contradicts that x' is global minimizer. (2) Suppose there are two global minimizer at x_1, x_2 such that $f(x_1) = f(x_2)$.

$$f(x_2) = f(x_1 + (x_2 - x_1)) = f(x_1) + \frac{1}{2}(x_2 - x_1)^T H(x_2 - x_1) > f(x_1),$$

which contradicts $f(x_1) = f(x_2)$. □

Theorem 7.1.3 (first order necessary and sufficient optimal condition for positive semi-definite quadratic programming, KKT condition). *Consider $x^* \in \mathbb{R}^n$ satisfying feasibility condition $Ax^* = b$. Let H be positive semi-definite. Then x^* is a global minimizer of the equality constrained quadratic optimization, if and only if there exists a vector $y^* \in \mathbb{R}^m$ such that*

$$\nabla f(x^*) = A^T y^*.$$

Proof. (1) If x^*, that fact that we have $\nabla f(x^*) = A^T y^*$ is addressed in Theorem 7.1.2. (2) The other direction is addressed in Theorem 7.1.2. □

7.1.3 Solving KKT systems

7.1.3.1 *Factorization approach*

The first way we introduce is to solve the KKT system by empolying symmetric factorization of the KKT matrix and then solving (x, y) sequentially. We have factorization given by

$$P^T K P = L D L^T$$

where P is an appropriately chosen permutation matrix, L is lower triangular with $diag(L) = I$, and D is block diagonal.

Since $K = PLDL^T P^T$, we can use the following procedure to sove the KKT system.

$$\begin{aligned}
&\text{solve} \quad Lz = P^T \begin{pmatrix} -c \\ b \end{pmatrix} \\
&\text{solve} \quad D\hat{z} = z \\
&\text{solve} \quad L^T \tilde{z} = \hat{z} \\
&\text{set} \quad \begin{pmatrix} x^* \\ y^* \end{pmatrix} = P\tilde{z}
\end{aligned}$$

where we use z to substitute $DL^T P^T \begin{pmatrix} x^* \\ y^* \end{pmatrix}$, $\hat{z}$ to substitute $L^T P^T \begin{pmatrix} x^* \\ y^* \end{pmatrix}$, etc.

7.1.3.2 *Range space approach*

Solving the KKT system involves solving

$$\begin{aligned} Hx - A^T y &= -c \\ Ax &= b \end{aligned}$$

Since A is full row rank, we can get one solution of x given by $x = (A^T A)^{-1} A^T b$. Multiply AH^{-1} on both sides of the first equation, we have

$$AH^{-1}Hx - AH^{-1}A^T y = -AH^{-1}c,$$

which simplifies to

$$b - AH^{-1}A^T y = -AH^{-1}c,$$

The range approach is quite effective under following conditions:

- B can be easily inverted or B^{-1} is known analytically.
- Small dimensionality problems.

7.1.4 Linear least square with linear constraints

7.1.4.1 *Least norm problem*

Lemma 7.1.3 (least norm with linear constraint). *The constrained minimizing problem*

$$\min \|x\|^2$$
$$\text{subject to } Cx = d$$

where

- $C \in \mathbb{R}^{p\times n}, d \in \mathbb{R}^p$
- *C has full row rank.*

has solution given by

$$x^* = C^T(CC^T)^{-1}d.$$

Note that $C^T(CC^T)^{-1}$ is the pseudo-inverse of C [Definition 4.14.1).

Proof. First $Cx^* = CC^T(CC^T)^{-1}d = d$ implies the constraint is satisfied. Second, for any $x \neq x^*, Cx = d$, we have

$$\begin{aligned} \|x\|^2 &= \|x - x^* + x^*\|^2 \\ &= \|x^*\|^2 + \|x - x^*\|^2 + 2(x - x^*)^T x^* \\ &= \|x^*\|^2 + \|x - x^*\|^2 \\ &\geq \|x^*\|^2 \end{aligned}$$

where we use the fact that

$$(x - x^*)^T x^* = (x - x^*)^T C^T (CC^T)^{-1} d = (Cx - Cx^*)^T (CC^T)^{-1} d = 0.$$

□

Lemma 7.1.4 (least norm with linear constraint). *The constrained minimizing problem*

$$\min \|Ax - b\|^2$$
$$\text{subject to } Cx = d$$

where

- $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m$
- $C \in \mathbb{R}^{p\times n}, d \in \mathbb{R}^p$
- *A, C has full row rank.*

has

- *the solution pair* (x^*, y^*) *satisfying*

$$Cx^* = d, A^T A x^* + A^T b = C^T y^*$$

or equivalently,

$$\begin{pmatrix} A^T A & C^T \\ C & 0 \end{pmatrix} \begin{pmatrix} x^* \\ -y^* \end{pmatrix} = \begin{pmatrix} -A^T b \\ d \end{pmatrix}.$$

- x^* *is given by*

$$x^* = x_0 + (A^T A)^{-1} C^T (C(A^T A)^{-1} C^T)^{-1} (d - Cx_0), x_0 = (A^T A)^{-1} A^T b.$$

Proof. (1) Theorem 7.1.1 (2) Use block matrix inversion Lemma A.8.6 to get solve the KKT equation.

□

Remark 7.1.1 (interpretation).

- Without the linear constraint, we get the normal equation solution
$$x_0 = (A^T A)^{-1} A^T b.$$
- The additional linear constraint acts as adjusting the original solution.

7.1.5 Application: Markovitz Portfolio Optimization Model

Definition 7.1.1. *A* ***portfolio vector*** *is a vector* $w \in \mathbb{R}^d$, *with the constraint* $\sum_i w_i = 1$. *We require* $\sum_{i=1}^n w_i = 1$ *can be thought as we split a unit money into investment of different assets.*

Suppose in our universe, there are n stocks. We are further given n estimated return $E(r_i)$, and the covariance matrix $Cov(r_i, r_j) = \Sigma$. For a portfolio characterized by $w^T r$, where w is the portfolio vector $w \in \mathbb{R}^n, \sum_{i=1}^n w_i = 1$ and r is the random variable vector $r \in \mathbb{R}^n$ with each component r_i characterized the return rate of asset i, the expected total return and the variance are given as

$$E(w^T r) = w^T E(r) = \sum_{i=1}^n w_i E(r_i)$$

$$Var(w^T r) = w^T Cov(i,j) w = \sum_{i,j}^n w_i w_j \sigma_{ij}$$

Definition 7.1.2 (mean-variance optimization formulation, minimum variance at fixed return). *Consider n assets with one period return given by $r_1, r_2, ..., r_n$. Denote the return mean by $\mu_1, \mu_2, ..., \mu_n$ and the covariance matrix by $\Sigma, \sigma_{ij} \triangleq \Sigma_{ij}$. Given an arbitrary value μ_0, we want to construct a portfolio,characterized by weight vector $w \in \mathbb{R}^n$ with such return μ_0 and minimum variance. The mean-variance optimization problem is given as*

$$\begin{aligned} &\min_w \frac{1}{2} \sum_{i,j}^n w_i w_j \sigma_{ij} \\ &s.t. \sum_{i=1}^n w_i \mu_i = \mu_0 \\ &\sum_i^n w_i = 1. \end{aligned}$$

Theorem 7.1.4 (sufficient and necessary conditions for efficient portfolio). *Consider n assets with one period return given by $r_1, r_2, ..., r_n$. Denote the return mean by $\mu_1, \mu_2, ..., \mu_n > 0$ and the **positive-definite** covariance matrix by $\Sigma, \sigma_{ij} \triangleq \Sigma_{ij}$. Consider the mean-variance optimization problem Given an arbitrary value μ_0, we want to construct with such return μ_0 and minimum variance. The mean-variance optimization problem is given as*

$$\begin{aligned} &\min_w \frac{1}{2} \sum_{i,j}^n w_i w_j \sigma_{ij} \\ &s.t. \sum_{i=1}^n w_i \mu_i = \mu_0 \\ &\sum_i^n w_i = 1. \end{aligned}$$

where the portfolio is characterized by weight vector $w \in \mathbb{R}^n$ and μ_0 is preselected number.

It follows that

- *There exists weights* $w_i, i = 1, ..., n$ *and the two Lagrange multipliers* λ_1 *and* λ_2 *satisfying*

$$\sum_{j=1}^{n} \sigma_{ij} w_j - \lambda_1 \mu_i - \lambda_2 = 0, i = 1, ..., n$$

$$\sum_{i=1}^{n} w_i \mu_i = \mu_0$$

$$\sum_{i=1}^{n} w_i = 1$$

- *Because* Σ *is positive definite, there exists unique* $(w, \lambda_1, \lambda_2)$ *to the above linear system; Moreover,* w *is the unique strict global minimizer.*

Proof. See the KKT condition for quadratic optimization Theorem 7.1.2. □

7.2 Quadratic optimization II: inequality constraints

7.2.1 Problem formulation

In inequality constraint quadratic programming, the

$$\min_{x\in\mathbb{R}^n} q(x) = c^T x + \frac{1}{2}x^T H x$$

subject to:

$$A_{\mathcal{E}}x = b_{\mathcal{E}}, A_{\mathcal{I}}x \geq b_{\mathcal{I}},$$

where

- $c \in \mathbb{R}^n, H \in \mathbb{R}^{n\times n}$ and H is symmetric and positive definite, $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m$.
- $\mathcal{E} = \{1, 2, ..., m_{\mathcal{E}}\}, \mathcal{I} = \{m_{\mathcal{E}} + 1, ..., m\}$, and $rank(A_{\mathcal{E}}) = m_{\mathcal{E}} \leq n$.

Example 7.2.1.

$$\begin{cases} \max_{x_1,x_2,x_3\in\mathbb{R}} & x_1^2 + x_1x_2 + 2x_2^2 + 3x_3 \\ \text{subject to} & x_1 + x_2 + x_3 = 100 \\ & x_2 + x_3 \geq 5 \end{cases}$$

7.2.2 Optimality conditions

7.2.2.1 *Pure inequality case*

Suppose in our inquality constrained optimization we only have inequality constraints $Ax \geq b$. The idea of first order optimality condition at x^* is that as we move around x^* while satisfying the constraint $Ax \geq b$, the objective function is non-decreasing.

There are two situations. First suppose x^* lies inside the interior of the feasible region defined by $Ax \geq b$, then the optimality condition is the same as a unconstrained optimization problem, i.e., $\nabla f(x^*) = 0$. This is because the constraint is not limiting the movement of x around x^*; that is, the constraint is inactive.

Second, suppose x^* lies on the boundary of the feasible region, then the optimality condition will have the form of $\nabla f(x^*) = A^T y^*, y^* > 0$, which encodes the idea that objective function will not decrease when move around x^* within the feasible region.

In the following, we first studied the Farka's Lemma, which enables us to transform the geometry insight into algebraic conditions. Then we present the KKT condition that accommodates the above two situations.

Lemma 7.2.1 (Farkas's lemma, alternative). *Let $g \in \mathbb{R}^n$ and $A \in \mathbb{R}^{r\times n}$. It follows that*

$$g^T p \geq 0, \forall p \in \{p : Ap \geq 0\}$$

if and only if there exists $\lambda \in \mathbb{R}^r, \lambda \geq 0$ such that

$$g = A^T \lambda.$$

Proof. (1) forward. If $g = A^T\lambda$, then $g^T p = \lambda A p \geq 0$; (2) converse. If $g \notin \mathcal{X} = \{A^T\lambda, \lambda \geq 0\}$, then g is the point lying outside the cone $\mathcal{X}$. Based on Theorem 9.2.3, there exists p such that $Ap \geq 0$(all the basis vectors of the cone lying on one halfspace of a hyperplane passing origin and having norm vector p) and $g^T p < 0$(the element g lying on the other halfspace). This contradicts that for all p such that $Ap \geq 0$, $g^T p \geq 0$. □

Theorem 7.2.1 (KKT optimality condition for pure inequality constrained quadratic programming). *If x^* is a minimizer of quadratic programming [subsection 7.2.1), then there exists $y \in \mathbb{R}^m$ such that*

$$\begin{aligned} a_i^T x^* &\geq b_i, \forall i \in \mathcal{I} \\ y_i^* &= 0, \forall i \in \mathcal{I}/\mathcal{A}(x^*) \\ y_i^* &\geq 0, \forall i \in \mathcal{A}(x^*) \cap \mathcal{I} \\ Hx^* + c = A^T y^* &= \sum_{i \leq i \leq m} y_i^* a_i \end{aligned}$$

where we use a_i^T to denote the ith row of A, $\mathcal{A}(x^*)$ denotes the active/binding index set at x^*, defined as

$$\mathcal{A}(x^*) = \{i : a_i^T x^* = b_i, \forall i \in \mathcal{I}\}$$

Proof. We consider a small movement step δx from x^*. To ensure $x^* + \delta x$ to remain in the feasible region, we require $A\delta x \geq 0$. Because x^* is a local minimizer, the gradient $g = Hx^* + c$ must satisfy $g^T \delta \geq 0$ (we can prove this via contradiction argument, similar to Theorem 7.1.2). Using Farkas' Lemma [Lemma 7.2.1], we must have $g = A^T y^*, y^* \geq 0$. For constraints that are inactive, we can set their associated $y*$ to zero. □

As an illustration [Figure 7.2.1], we examine the optimality condition for $f(x_1, x_2) = x_1^2 + x_2^2$ under constraint $x_1 + x_2 \geq 1$. The KKT condition for a local minimal requires the decreasing of f cannot be achieved when x moves inside the feasible region.

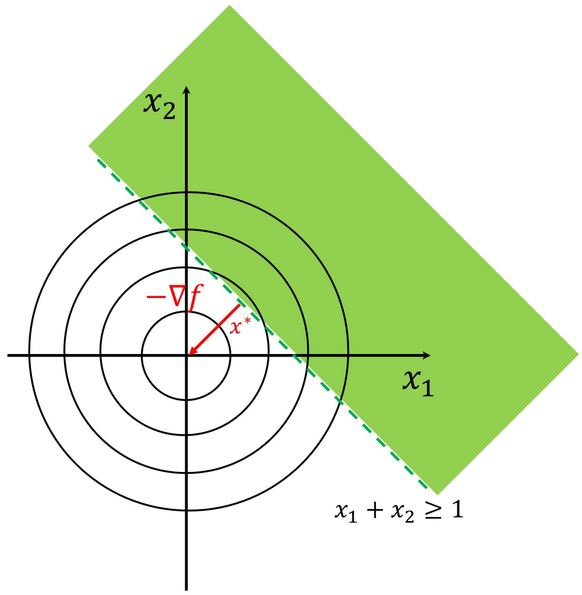

Figure 7.2.1: Demonstration of KKT condition at a local minimal for $f(x_1, x_2) = x_1^2 + x_2^2$ under constraint $x_1 + x_2 \geq 1$.

7.2.2.2 *General constrained optimization*

We can combine KKT conditions of equality constraint case and pure inequality constraint case and arrive at the following KKT conditions for general constrained quadratic optimization.

Theorem 7.2.2 (KKT optimality condition general constrained quadratic programming). *If x^* is a minimizer of quadratic programming [subsection 7.2.1], then there exists $y \in \mathbb{R}^m$ [a] such that*

$$\begin{aligned} a_i^T x^* &= b_i, \forall i \in \mathcal{E} \\ a_i^T x^* &\geq b_i, \forall i \in \mathcal{I} \\ y_i^* &= 0, \forall i \in \mathcal{I}/\mathcal{A}(x^*) \\ y_i^* &\geq 0, \forall i \in \mathcal{A}(x^*) \cap \mathcal{I} \\ Hx^* + c = A^T y^* &= \sum_{i \leq i \leq m} y_i^* a_i \end{aligned}$$

where $\mathcal{A}(x^*)$ denotes the active set at x^*, defined as

$$\mathcal{A}(x^*) = \mathcal{E} \cup \{i : a_i^T x^* = b_i, \forall i \in \mathcal{I}\}$$

[a] for y_i associated with equality constraints, y_i will not subject to non-negative constraint.

7.2.2.3 *Positive semi-definitive quadratic programming*

Lemma 7.2.2 (local minimum and global minimum in inequality constrained quadratic programming). *Let x^* be a local minimizer of the equality constrained quadratic optimization problem.*

- *If H is positive semi-definite, then x^* is also a global minimizer.*
- *If H is positive definite, then the optimization problem only has at most one global minimizer and x^* is the only global minimizer.*

Proof. Because the feasible regions are a convex set, we can use same proof technique in Lemma 7.1.2. □

Theorem 7.2.3 (KKT necessary and sufficient optimality condition for positive semi-definite quadratic programming). *If x^* is a minimizer of quadratic programming [subsection 7.2.1), then there exists $y \in \mathbb{R}^m$ such that*

$$\begin{aligned} a_i^T x^* &= b_i, \forall i \in \mathcal{E} \\ a_i^T x^* &\geq b_i, \forall i \in \mathcal{I} \\ y_i^* &= 0, \forall i \in \mathcal{I}/\mathcal{A}(x^*) \\ y_i^* &\geq 0, \forall i \in \mathcal{A}(x^*) \cap \mathcal{I} \\ Hx^* + c = A^T y^* &= \sum_{i \leq i \leq m} y_i^* a_i \end{aligned}$$

where $\mathcal{A}(x^*)$ denotes the active set at x^*, defined as

$$\mathcal{A}(x^*) = \mathcal{E} \cup \{i : a_i^T x^* = b_i, \forall i \in \mathcal{I}\}$$

Moreover, if H is semi-positive definite, then above condition for x^* is sufficient for being global minimizer.

Proof. (1) Forward direction is from KKT condition [Theorem 7.2.2]. (2) Consider another feasible point $x' = x^* + p$ where $p \neq 0, p^T a_i = 0, \forall i \in \mathcal{E}, p^T a_i \geq 0, \forall i \in \mathcal{A} \cap \mathcal{I}$. Then

$$\begin{aligned}&q(x^* + p) - q(x^*)\\ =&\frac{1}{2}p^T H p + p^T c + p^T H x^*\\ =&\frac{1}{2}p^T H p + p^T (A^T y^*)\\ =&\frac{1}{2}p^T H p + \sum_{i \in \mathcal{I}} p^T a_i y_i^* \geq 0\end{aligned}$$

where we use the fact that $y_i^* = 0, \forall i \in \mathcal{I}/\mathcal{A}$ (inactive constraint), and $p^T a_i \geq 0, \forall i \in \mathcal{A} \cap \mathcal{I}$. Because p is arbitrarily chosen, therefore x^* is a local minimum. □

7.2.3 Primal active-set method

Definition 7.2.1 (working set). *A working set $\mathcal{W}_k$ is a set of index such that*

$$\mathcal{E} \subseteq \mathcal{W}_k \subseteq \mathcal{A}(x_k)$$

Definition 7.2.2 (subspace minimizer). *Given the working set $\mathcal{W}_k$ associated with x_k, we say that the solution x_k^* to*

$$\min_{x \in \mathbb{R}^n} q(x) = c^T x + \frac{1}{2} x^T H x, \text{ subject to } A_{\mathcal{W}_k} x = b_{\mathcal{W}_k}$$

is a subspace minimizer associated with $\mathcal{W}_k$, where $A_{\mathcal{W}_k}$ denotes the rows from A correspond to $\mathcal{W}_k$.

Lemma 7.2.3 (step to subspace minimizer). *Given current iterate x_k, the step p_k to the subspace minimizer such that $x_k^* = x_k + p_k$ is also the minimizer of the following minimizer problem*

$$\min_{p \in \mathbb{R}^n} \frac{1}{2} p^T H p + g_k^T p, \text{ subject to } A_{\mathcal{W}_k} p = 0,$$

where $g_k^T = \nabla q(x_k)$.

Proof.

$$
\begin{aligned}
q(p+x_k) &= c^T(p+x_k) + \frac{1}{2}(p+x_k)^T H(p+x_k) \\
&= c^T p + c^T x + \frac{1}{2}p^T Hp + p^T Hx_k = p^T g_k + \frac{1}{2}p^T Hp
\end{aligned}
$$

where $g_k = Hx_k + c$. □

Lemma 7.2.4 (subspace minimizer step as descend step). *Let p_k be the subspace minimizer associated with constraint matrix $A_{\mathcal{W}_k}$. Assume that $p_k \neq 0$ and $H > 0$, then*

- $g_k^T p_k = -p_k^T Hp_k < 0.$
- $q(x_k + \alpha p_k) < q(x_k)$ *for all* $0 < \alpha < 2.$

Proof. (1) The KKT condition for p_k to be a minimizer of

$$\min_{p\in\mathbb{R}^n} \frac{1}{2}p^T Hp + g_k^T p, \text{ subject to } A_{\mathcal{W}_k} p = 0,$$

is given by

$$Hp_k - A_{\mathcal{W}_k}^T y = -g_k, A_{\mathcal{W}_k} p_k = 0.$$

Then,

$$p_k^T g_k = -p_k^T Hp_k + p_k^T A_{\mathcal{W}_k}^T y = -p_k^T Hp_k < 0.$$

(2)

$$q(x_k + \alpha p_k) - q(x_k) = \frac{1}{2}\alpha^2 p_k^T Hp_k + \alpha p^T g_k = \frac{1}{2}\alpha^2 p_k^T Hp_k - \alpha p_k^T Hp_k = \frac{1}{2}\alpha(\alpha-2)p_k^T Hp_k,$$

when $2 > \alpha > 0$, we have $q(x_k + \alpha p_k) < q(x_k)$. □

Remark 7.2.1 (overshooting issue). Note that when we take $1 < \alpha < 2$, we are moving more than p_k, but we can still achieve decrease in the q due to the quadratic nature of q.

Lemma 7.2.5 (decreasing and blocking constraints in the search and descent direction). *Let p_k be the subspace minimizer associated with constraint matrix $A_{\mathcal{W}_k}$. Assume that $p_k \neq 0$ and $H > 0$.*

The set of ***decreasing and blocking constraints*** *$\mathcal{D}_k$ at the point x_k is given by*

$$\mathcal{D}_k \triangleq \{j \in [1:m] : j \notin \mathcal{W}_k, a_j^T p_k < 0\},$$

where a_j^T is the jth row of the constraint matrix.

The maximum step size α before hitting a blocking constraint i is given by:

$$\alpha = \frac{b_i - a_i^T p_k}{a_i^T p_k}.$$

Note that constraint $i \in \mathcal{W}_k$ will not be a blocking constraint since $a_i^T p_i = 0$.

Proof. Note that

$$a_i^T(x_k + \alpha p_k) = b_i \implies \alpha = \frac{b_i - a_i^T p_k}{a_i^T p_k}.$$

□

Remark 7.2.2 (geometry of decreasing constraints)**.** When we move along direction p_k, we might encounter hyperplanes(i.e. other constraints). The set of hyperplanes encountered are $\mathcal{D}_k$.

Moreover, if we do not encounter any hyperplanes, then we can take the minimizer step p_k.

Algorithm 12: Primal active-set method for strictly convex quadratic programming

Input: Initial feasible x_0 with associated working set $\mathcal{W}_0$ such that $\mathcal{E} \subseteq \mathcal{W}_0 \subseteq \mathcal{A}(x_0)$

Set $k = 0$
repeat
  Compute gradient $g_k = \nabla q(x_k) = Hx_k + c$.
  Compute y_k as $[y_k]_{i \notin \mathcal{W}_k} = 0$ and $[y_k]_{\mathcal{W}_k} = \hat{y}_k$, where $(p_k, \hat{y}_k)$ as the solution to

$$\min_p \frac{1}{2} p^T H p + g_k^T p, \text{ subject to } A_{\mathcal{W}_k} p = 0.$$

  if $p_k = 0$ **then**
    if $[y_k]_i \geq 0$ *for all* $i \in \mathcal{W}_k \cap \mathcal{I}$ **then**
      return x_k as the minimizer since KKT condition is satisfied
    end
    else
      Set $s = \arg\min_{i \in \mathcal{W}_k \cap I} [y_k]_i, x_{k+1} = x_k, \mathcal{W}_{k+1} = \mathcal{W}_k / \{s\}$,
    end
  end
  else
    Set $x_{k+1} = x_k + \alpha_k p_k$ and α_k is the minimum step length such that

$$\alpha_k = \min(1, \min_{i \in \mathcal{D}_k} \frac{b_i - a_i^T x_k}{a_i^T p_k}),$$

    where $\mathcal{D}_k = \{i \in [1:m] : i \notin \mathcal{W}_k, a_i^T p_k < 0\}$.
    if $\alpha_k < 1$ **then**
      set $\mathcal{W}_{k+1} = \mathcal{W}_k \cup \{t\}$ for some $t \in \mathcal{D}_k$ satisfying $\frac{b_t - a_t^T x_k}{a_t^T p_k} = \alpha_k$
    end
    else
      set $\mathcal{W}_{k+1} = \mathcal{W}_k$.
    end
  end
  Set $k = k + 1$.
until *termination condition satisfied*;

Output: approximate minimizer x_k

Remark 7.2.3 (interpretation).

- We set $[y_k]_{i \notin \mathcal{W}_k} = 0$ means that we assume inequality constraints not in the working set $\mathcal{W}_k$ will be set as inactive.
- If $p_k = 0$ and $[y_k]_i \geq 0$ for all $i \in \mathcal{W}_k \cap \mathcal{I}$, then the KKT condition is satisfied; If $p_k = 0$, but not $[y_k]_i \geq 0$ for all $i \in \mathcal{W}_k \cap \mathcal{I}$, then that mean by stepping off some active constraint, we can get objective function to decrease(see the following lemma).
- If $p_k > 0$, then we can show that p_k is a descent direction [Lemma 7.2.4), so we move along p_k until a constraint is hit.

Lemma 7.2.6 (feasible direction when stepping off constraints). *[2, p. 470] Let p_k be the subspace minimizer associated with constraint matrix $A_{\mathcal{W}_k}$. Assume*

- $p_k = 0$
- $\{a_i\}_{i \in \mathcal{W}_k}$ *is a linearly independent set*
- *there exists some $j \in \mathcal{W}_k \cap \mathcal{I}$ such that $[y_k]_j < 0$.*

Then the direction p_{k+1} computed as the solution to

$$\min_{p \in \mathbb{R}^n} \frac{1}{2} p^T H p + g_k^T p, \text{ subject to } a_i^T p = 0, \forall i \in \mathcal{W}_k / \{j\}$$

satisfies $a_j^T p_k \geq 0$, i.e., it is a feasible direction for constraint j.

Moreover, if $H > 0$, then

$$a_j^T p_k > 0,$$

satisfies s

Proof. Note that $p_k = 0$, we have

$$Hp_k = 0 = -g + \sum_{i \in \mathcal{W}_k} y_i a_i.$$

For p_{k+1}, we have

$$Hp_{k+1} = 0 = -g + \sum_{i \in \mathcal{W}_k - \{j\}} y_i' a_i.$$

Subtract, we get

$$-Hp_{k+1} = \sum_{i \in \mathcal{W}_k - \{j\}} (y_i - y_i') a_i + a_j^T y_j.$$

Multiply both sides by p_{k+1}, we get

$$0 > -p_{k+1} H p_{k+1} = p_{k+1}^T a_j y_j \implies p_{k+1}^T a_j > 0.$$

□

Remark 7.2.4. This lemma says if we encounter the situation $p_k = 0$ but there exists negative multiplier in the active constraint. Then we should remove the constraint and solve the subspace minimizer problem again. The new subspace minimizer will be descent directionLemma 7.2.4 and step off the constraint to a feasible region.

7.2.4 Gradient projection method

Definition 7.2.3 (bound-constrained quadratic optimization). *The bound-constrained quadratic optimization problem is given by*

$$\min_{x\in\mathbb{R}^n} q(x) = c^T x + \frac{1}{2}x^T Hx, \textit{ subject to } l \leq x \leq u$$

where

- $c \in \mathbb{R}^n$ *and* $H \in \mathbb{R}^{n\times n}$ *is symmetric.*
- H *may be indefinite, i.e.,* q *may be nonconvex.*
- $l \in \mathbb{R}^n$*(may contain* $-\infty$*) and* $u \in \mathbb{R}^n$*(may contain* ∞*). Assume* $l < u$.

Definition 7.2.4 (projection operator). *The projection operator* $P : \mathbb{R}^n \to \mathbb{R}^n$ *associated with the constraints* $l \leq x \leq u$ *is defined component-wise as*

$$[P(x;l,u)]_i = \begin{cases} l_i, & if\ x_i < l_i \\ x_i, & if\ l_i \leq x_i \leq u_i \\ u_i, & if\ x_i > u_i \end{cases}$$

Remark 7.2.5. the range of the projection operator is the feasible set

Definition 7.2.5 (projected gradient path). *The projected gradient path is defined by the piece-wise linear path* $\{x(t) : t \geq 0\}$ *such that*

$$x(t) = P(x - t\nabla q(x); l, u), \forall t \geq 0.$$

Definition 7.2.6 (Cauchy step). *The Cauchy step* x^c *is defined as*

$$x^c = x(t^c),$$

where t^c is the solution of [a]

$$\min_{t\geq 0} q(x(t)).$$

[a] The detailed calculation procedure for the Cauchy step can be found at [2, p. 486].

Algorithm 13: First order gradient projection algorithm

Input: Initial $x_{in} \in \mathbb{R}^n$
1. Set $x_0 = P(x_{in}; l, u)$ to make x_0 feasible. Set $k = 0$
2. **repeat**
3. **if** *x_k is first order KKT point* **then**
4. **return** x_k as the minimizer
5. **end**
6. With $x = x_k$, compute the Cauchy point x^c.
7. Set $x_{k+1} = x^c$.
8. Set $k = k + 1$.
9. **until** *termination condition satisfied*;

Output: approximate minimizer x_k

7.2.5 Dual convex quadratic programming

Lemma 7.2.7 (dual form inequality constraint convex quadratic optimization). *[3, p. 437] The primal problem of*

$$\min_{x\in\mathbb{R}^n} \frac{1}{2}x^T Hx + c^T x, \text{ subject to } Ax \geq b$$

with $H > 0$, $A \in \mathbb{R}^{m\times n}$, and $rank(A) = n > m$ has the dual form given by

$$\max_{y\in\mathbb{R}^m} -\frac{1}{2}(A^T y - c)^T H^{-1}(A^T y - c) + b^T y, \text{ subject to } y \geq 0$$

Proof. The Lagrange function is

$$L(x, y) = \frac{1}{2}x^T Hx + c^T x - y^T(Ax - b), y \geq 0.$$

To minimize over x, we set

$$\nabla_x L(x, y) = 0 \implies Hx + c - A^T y = 0.$$

Then, the dual function

$$d(y) = \min_x L(x, y) = -(A^T y - c)^T H^{-1}(A^T y - c) + b^T y, y \geq 0.$$

□

Remark 7.2.6 (convert to easier bound-constrained optimization). Note that using the dual formulation we can convert the convex quadratic optimization to a simplier bound-constrained convex quadratic optimization [Definition 7.2.3].

Lemma 7.2.8 (dual form equality constraint convex quadratic optimization). *The primal problem of*

$$\min_{x\in\mathbb{R}^n} \frac{1}{2}x^T Hx + c^T x, \text{ subject to } Ax = b$$

with $H > 0$, $A \in \mathbb{R}^{m\times n}$, *and* $\mathrm{rank}(A) = m \leq n$ *has the dual form given by*

$$\max_{y\in\mathbb{R}^m} -\frac{1}{2}(A^T y - c)^T H^{-1}(A^T y - c) + b^T y$$

Proof. Similar to inequality constraint case. □

Remark 7.2.7 (convert to easier unconstrained problem). Note that using the dual formulation we can convert the convex quadratic optimization to a simplier unconstrained optimization.

7.3 General equality constrained optimization

Notations

- Jacobian $J(x) = \nabla c \in \mathbb{R}^{m \times n}$, each row is ∇c_i^T

7.3.1 Feasible path and optimality

Definition 7.3.1 (Equality-constraint optimization). *A equality constrained linear programming is given as:*

$$\min_{x \in \mathbb{R}^n} f(x)$$
$$\textit{subject to } c(x) = 0$$

where $f : \mathbb{R}^n \to \mathbb{R}, c : \mathbb{R}^m \to \mathbb{R}^n$ *where* $c = [c_1, ..., c_m]^T, c_i : \mathbb{R}^n \to R$.

Definition 7.3.2 (feasible path, tangent vector). *A feasible path is a curve for constraints* $c(x) = 0$, *represented by a twice continuously differentiable function* $x(t)$, *that emanates from a feasible point* x_0 *such that*

$$x(0) = x_0, c(x(t)) = 0$$

for all $0 \leq t < \sigma, \sigma > 0$, *and such that* $dx/dt|_{t=0} \neq 0$ *The tangent vector of the feasible path is given as*

$$p = dx(t)/dt|_{t=0}.$$

Definition 7.3.3 (tangent cone). *Given constraints* $c(x) = 0$, *the set*

$$\mathcal{T}(x) = \{p : p \textit{ is a nonzero vector tangent to a feasible path emanating from } x\} \cup \{0\}\}$$

is called the tangent cone of c *at the point* x.

Remark 7.3.1 (tangent cone is a cone). The tangent cone of constraints $c(x) = 0$ at x is a cone; that is, given $p \in \mathcal{T}(x), \alpha p \in \mathcal{T}(x), \alpha > 0$. Note that p is vector tangent to some feasible path emanating from x, then αp will still be a vector tangent to the original path.

Theorem 7.3.1 (first order necessary condition, geometric form). *[1, lec 4]If x^* is a local minimizer, then $c(x^*) = 0$ and $\nabla f(x^*)^T p \geq 0, \forall p \in \mathcal{T}(x^*)$.*

Proof. Suppose for some p, we have $\nabla f(x^*)^T p < 0$, then in direction p, let $\gamma(t)$ be the feasible curve emanating from x^* with tangent vector p, we have

$$f(\gamma(\alpha)) = f(x^*) + \alpha f(x^*)^T p + O(\alpha^2) < f(x^*)$$

as $\alpha \to 0, \alpha > 0$, which contradicts that fact that $f(x^*)$ is local minimum. □

7.3.2 Constraint qualification and Lagrange theory

Lemma 7.3.1 (algebraic characterization of tangent cone).

$$\mathcal{T}(x) \subseteq \mathcal{N}(J(x)),$$

that is, if $p \in \mathcal{T}(x)$, then $(\nabla c_i)^T p = 0, \forall i = 1, ..., m$, or equivalently, $p \in \mathcal{N}(J(x))$.

Proof. Since $c(x(\alpha)) = 0$, for $\alpha \in [0, \sigma)$ along a feasible path $x(\alpha)$ emanating from x, we know that

$$0 = \frac{d}{d\alpha} c_i(x(\alpha))|_{\alpha=0} = \nabla c_i(x)^T p$$

where

$$p = \frac{d}{d\alpha} x(\alpha)|_{\alpha=0} = \lim_{\alpha \to 0^+} \frac{x(\alpha) - x(0)}{\alpha}.$$

Therefore, we know that p satisfy $(\nabla c_i)^T p = 0, i = 1, ..., m$, which implies

$$\sum_{i=1}^{m} (\nabla c_i)^T p = 0 \Leftrightarrow Jp = 0$$

Note that $Jp = \sum_{i=1}^{m} (\nabla c_i)^T p = 0$ can not generally implies $(\nabla c_i)^T p = 0, \forall i = 1, ..., m$ unless J has linearly independent rows. □

Example 7.3.1. [4] Consider constraints $c_1(x) = x_1^3 - x_2 = 0, c_2(x) = x_2 = 0$. At a feasible point $x^* = 0$, we have

$$J = \begin{bmatrix} 0 & -1 \\ 0 & 1 \end{bmatrix},$$

and

$$\mathcal{N}(J(x^*)) = \{(\gamma, 0)^T, \gamma \in \mathbb{R}\}.$$

However, x^* is the only feasible point so that no feasible paths exist.

Therefore, we have

$$\{0\} = \mathcal{T}(x^*) \subset \mathcal{N}(J(x^*)) = \{(\gamma, 0)^T, \gamma \in \mathbb{R}\}.$$

Definition 7.3.4 (constraint qualification). *We say that the constraint qualification of equality constraint $c(x) = 0$ holds at a feasible point x if every nonzero vector p satisfying $J(x)p = 0$ implies $p \in \mathcal{T}(x)$.*

Remark 7.3.2 (purpose of constraint qualifications)**.**

- constraint qualifications is simple the assumptions on constraint such that later KKT condition relies on
- There are different types of constraint qualifications, such as linear independence constraint qualification, Managasarian-Fromovitz constraint qualification.

Lemma 7.3.2. *If constraint qualifications of equality constraint $c(x) = 0$ holds, then*

$$\mathcal{T}(x) = \mathcal{N}(J(x))$$

Proof. directly form above lemma and the definition and constraint qualification. □

Lemma 7.3.3 (linear constraints always satisfy constraint qualification). *The constraint qualification of equality constraint $c(x) = 0$ holds at x if $c(x) = Ax - b = 0($ no matter A has full row rank or not $)$.*

Remark 7.3.3. Assume $Ax = b$ has infinitely many solution, which form a subspace. Consider a feasible point x_0, then from the definition, $\mathcal{T}(x^*)$ is just $\mathcal{N}(A)$ since $A(x + \alpha p) = b, \forall \alpha \geq 0, p \in \mathcal{N}(A)$.

Lemma 7.3.4 (sufficient condition for constraint qualifications, nonlinear constraint case). *[2, p. 324] The constraint qualification of equality constraint $c(x) = 0$ holds at x if $J(x)$ has full row rank (i.e. each row, ∇c_i, are linearly independent of each other).*

Lemma 7.3.5 (existence of Lagrange multipliers for equality constraints). *[1, lec 4] Assume that f and c are differentiable at a feasible point x^*. Then*

$$\nabla f(x^*)^T p \geq 0, \forall p \in \mathcal{N}(J(x^*))$$

if and only if $\nabla f(x^*) \in \mathcal{R}(J(x^*))$; that is there exist some λ such that

$$\nabla f(x^*) = J(x^*)^T \lambda$$

Proof. (1) forward. If $\nabla f(x^*) \in \mathcal{R}(J(x^*))$, then $\nabla f(x^*)^T p = 0$, since $\mathcal{R}(J(x^*)^T) \perp \mathcal{N}(J(x^*))$; (2) converse. Let $\nabla f(x^*) = g = g_N + g_R$ where $g_R \in \mathcal{R}(J(x^*)), g_N \in \mathcal{N}(J^T)$. (This decomposition is guaranteed by rank-nullity decomposition theorem). Then suppose $g \notin \mathcal{R}(J)$, which implies $g_N \neq 0$. Then $g^T p = -g_N^T p < 0$ if $p = g_N \in \mathcal{N}(J^T)$. That is, there exist some p such that $\nabla f(x^*)^T p < 0$. □

Remark 7.3.4. This lemma enables us to formulate our first-order optimality condition more concisely.

Theorem 7.3.2 (first order necessary condition, KKT condition, equality constraints). *Assume that the constraint qualification holds. It follows that if x^* is a local minimizer, then*

$$c(x^*) = 0 \text{ and } g(x^*) = J(x^*)^T \lambda,$$

for some vector λ^*, known as Lagrange multiplier. The pair (x^*, λ^*) is also known as KKT points.

Proof. When constraint qualification holds and x^* is a local minimizer, we know that

$$c(x^*) = 0, \nabla f(x^*)^T p \geq 0, \forall p \in \mathcal{N}(J(x^*)).$$

Then follow the above lemma, we have

$$c(x^*) = 0 \text{ and } g(x^*) = J(x^*)^T \lambda.$$

□

Example 7.3.2. Consider an equality constrained quadratic optimization is given by

$$\min_{x \in \mathbb{R}^n} f(x) = c^T x + \frac{1}{2} x^T H x, \text{ subject to } Ax = b,$$

where $c \in \mathbb{R}^n$, $H \in \mathbb{R}^{n \times n}$ is symmetric, and $A \in \mathbb{R}^{m \times n}, m \leq n, rank(A) = m, b \in \mathbb{R}^M$.

The first-order optimality conditions for a minimizer x^* of the equality constrained quadratic optimization problem is that there exists a vector $y^* \in \mathbb{R}^m$ such that

$$Ax^* = b(feasible) \text{ and } Hx^* + c = A^T y^*.$$

Note 7.3.1 (interpretation of Lagrange multipliers, equality constraint). Assume linear independence qualification holds. Suppose that (x^*, λ^*) is a KKT point such that

$$g(x^*) = J(x^*)^T \lambda^* \Leftrightarrow g \perp \mathcal{N}(J).$$

We have the following observations:

- If $\lambda_i^* = 0$, then the constraint i is redundant; that is solution will not change if we ignore this constraint.
- Consider an arbitrary direction $p \in \mathbb{R}^n$:
  - If $J(x)p = 0$, then $p \in \mathcal{N}(J(x^*))$, $g(x^*)^T p = \lambda^T Jp = 0$; that is, p is in the feasible direction, however, since x^* is the local minimizer, moving in p will not change the value of objective function.
  - If p is such that $Jp = e_i$(e_i is the ith standard basis), then moving in direction p will only step off the constraint i but maintain within tangent space of other constraints. To see this, we have $c(x + \alpha p) \approx c(x) + e_i$. Moreover, if $y_i > 0$, then $g^T p = y^T Jp = y_i$, then moving in direction p will increase the value of objective function; vice versa.
  - If p is such that $Jp \geq 0$, then

    $$g^T p = [y^*]^T Jp,$$

    whose value depends on specific signs of components in y^*.
- Note that when we want to examine the effect of a specific constraint on the optimal objective function value, we need to find a direction only step off this specific constraint.
- If $J(x^*)$ has full row rank, then y^* is **unique** since $g = J^T y$.

Remark 7.3.5 (how to calculate (x^*, λ^*)). Let us assume $x \in \mathbb{R}^n$ and $\lambda^* \in \mathbb{R}^m$(i.e. there are m constraints). The feasibility condition $c(x^*)$ provides m equations, the condition $g(x^*) = J(x^*)^T \lambda$ provides n equations. In principle, we can solve (x^*, λ^*) by solving roots to nonlinear equations.

7.3.3 Second order condition

Theorem 7.3.3 (second order necessary condition, geometric form). *[1, lec 4] Assume that the first-order constraint qualification holds. Suppose x^* is a constrained minimizer and consider a feasible path $x(\alpha)$ such that*

$$x(0) = x^*, 0 \neq p \triangleq \frac{d}{d\alpha} x(\alpha)|_{\alpha=0}, v \triangleq \frac{d^2}{d\alpha^2} x(\alpha)|_{\alpha=0}.$$

Then

$$\frac{d^2}{d\alpha^2} f(x(\alpha)) \geq 0,$$

for all such feasible paths.

Proof. Use Taylor expansion along the path $x(\alpha)$, we have

$$\begin{aligned} f(x(\alpha)) &= f(x(0)) + \alpha \frac{d}{d\alpha} f(x(\alpha))|_{\alpha=0} + \frac{\alpha^2}{2} \frac{d^2}{d\alpha^2} f(x(\alpha))|_{\alpha=0} + O(\alpha^3) \\ &= f(x^*) + \alpha g(x^*)^T p + \frac{\alpha^2}{2} \frac{d^2}{d\alpha^2} f(x(\alpha))|_{\alpha=0} + O(\alpha^3) \end{aligned}$$

When the constraint qualification holds, we have

$$g(x^*) = J^T y*, Jp = 0 \implies \alpha g(x^*)^T p = 0.$$

Therefore,

$$f(x(\alpha)) = f(x^*) + \frac{\alpha^2}{2} \frac{d^2}{d\alpha^2} f(x(\alpha))|_{\alpha=0} + O(\alpha^3).$$

Use contradiction argument, we can show that

$$\frac{d^2}{d\alpha^2} f(x(\alpha))|_{\alpha=0} \geq 0.$$

□

Theorem 7.3.4 (second order necessary conditions, algebraic form). *Assume that the constraint qualification holds at x^*. It follows that if x^* is a local minimizer, then*

- *(feasible) $c(x^*) = 0$.*
- *There exists a Lagrange multiplier $y^* \in \mathbb{R}^m$ such that $\nabla f = J(x^*)^T y^*$.*
- *For y^* in (2), we have*

$$p^T [\sum_{i=1}^m -y_i^* \nabla^2 c_i(x^*) + H(x^*)] p \geq 0$$

holds for all $p \in \mathcal{N}(J(x^*))$.

Proof. Use that fact that

$$\frac{d^2}{d\alpha^2}f(x(\alpha)) = \frac{d}{d\alpha}[g(x(\alpha)^T x'(\alpha)] = g(x(\alpha))^T x''(\alpha) + x'(\alpha)^T H(x(\alpha))x'(\alpha)$$

When the constraint qualification holds, we have

$$g(x^*) = J^T y * .$$

Therefore, use Theorem 7.3.3, we have

$$0 \leq \frac{d^2}{d\alpha^2}f(x(\alpha))|_{\alpha=0} = [y^*]^T J(x^*)v + p^T H(x^*)p = \sum_{i=1}^{m} y_i^*(\nabla c_i(x^*))^T v + p^T H(x^*)p.$$

For a feasible path satisfying $c_i(x(\alpha)) = 0, \forall i$, we have

$$0 = \frac{d}{d\alpha}c_i(x(\alpha)) = \frac{d}{d\alpha}[\nabla c_i(x(\alpha))^T x'(\alpha)] = \nabla c_i(x(\alpha))^T x''(\alpha) + x'(\alpha)^T \nabla^2 c_i(x(\alpha))x'(\alpha).$$

Simplify, we have

$$0 = \nabla c_i(x^*)^T v + p^T \nabla^2 c_i(x^*)p.$$

Then,

$$0 \leq \frac{d^2}{d\alpha^2}f(x(\alpha))|_{\alpha=0} = \sum_{i=1}^{m} y_i^*(\nabla c_i(x^*))^T v + p^T H(x^*)p$$

will reduce to

$$p^T[\sum_{i=1}^{m} -y_i^* \nabla^2 c_i(x^*) + H(x^*)]p \geq 0.$$

□

Remark 7.3.6 (interpretation, analog to unconstrained optimization using reduced Hessian).

- The Hessian condition in (3) says that the Hessian of the Lagrange is positive definite for vectors p of linear feasible direction set(i.e. $p \in \mathcal{N}(J(x^*))$).
- If we define $Z(x^*)$ to be the matrix whose columns form a basis of $cN(J(x^*))$, then condition (3) is equivalent to
$$Z^T HZ > 0.$$
- The matrix $Z^T HZ$ is called **reduced Hessian**, and plays the role analogous to $H(x^*)$ in unconstrained optimization.

Theorem 7.3.5 (second order sufficient optimality condition, strict local minimizer). *[1, lec 4][3, p. 364] The vector $x^* \in \mathbb{R}^n$ is a strict local minimizer of the optimization problem [Definition 7.3.1] if*

- *(feasible) $c(x^*) = 0$.*
- *There exists a Lagrange multiplier $y^* \in \mathbb{R}^m$ such that $\nabla f = J(x^*)\lambda^*$.*
- *For y^* in (2), we have*

$$p^T[\sum_{i=1}^m -y^*\nabla^2 c_i(x^*) + H(x^*)]p > 0$$

holds for all $p \neq 0$ and $J(x^*)p = 0$.

In fact, if the above conditions holds, it can be showed that there exists scalars $\gamma > 0$ and $\epsilon > 0$ such that

$$f(x) \geq f(x^*) + \frac{\gamma}{2}\|x - x^*\|, \forall x \text{ such that } h(x) = 0, \|x - x^*\| < \epsilon.$$

Proof. **Note that we do not require constraint qualification.** all feasible directions $\mathcal{T} \subseteq \mathcal{N}(J(x^*))$. We can use similar techniques used in proving necessary conditions. □

7.4 General inequality constrained optimization

Notations

- Jacobian $J(x) = \nabla c \in \mathbb{R}^{n \times m}$, each column is ∇c_i
- Active set $\mathcal{A}(x) = \{i : c_i(x) = 0\}$

7.4.1 Feasible path and optimality

Definition 7.4.1 (inequality-constraint optimization). *A equality constrained linear programming is given as:*

$$\min_{x \in \mathbb{R}^n} f(x)$$
$$\text{subject to } c(x) \geq 0$$

where $f : \mathbb{R}^n \to \mathbb{R}, c : \mathbb{R}^m \to \mathbb{R}^n$ *where* $c = [c_1, ..., c_m]^T, c_i : \mathbb{R}^n \to R$.

Definition 7.4.2. *Consider a x satisifying a constraint* $c_i(x) \geq 0$. *We said*

- *the constraint is* ***active or binding*** *at* x_0 *if* $c(x_0) = 0$;
- *the constraint is* ***inactive or not binding*** *at* x_0 *if* $c(x_0) > 0$.

Definition 7.4.3 (feasible path, tangent vector). *[1, lec 4]A feasible path is a curve for constraints* $c(x) = 0$, *represented by a twice continuously differentiable function* $x(t)$, *that emanates from a feasible point* x_0 *such that*

$$x(0) = x_0, c(x(t)) = 0$$

for all $0 \leq t < \sigma, \sigma > 0$, *and such that* $dx/dt|_{t=0} \neq 0$ *The tangent vector of the feasible path is given as*

$$p = dx(t)/dt|_{t=0}.$$

Definition 7.4.4 (tangent cone). *[1, lec 4]Given constraints* $c(x) \geq 0$, *the set*

$$\mathcal{T}(x) = \{p : p \text{ is a nonzero vector tangent to a feasible path emanating from } x\} \cup \{0\}\}$$

is called the tangent cone of c at the point x.

Theorem 7.4.1 (first order necessary condition, geometric form). *[1, lec 4]If x^* is a local minimizer for inequality constraint optimization problem, then*

$$c(x^*) \geq 0 \text{ and } \nabla f(x^*)^T p \geq 0, \forall p \in \mathcal{T}(x^*)$$

.

Proof. Suppose for some p, we have $\nabla f(x^*)^T p < 0$, then in direction p, let $\gamma(t)$ be the feasible curve emanating from x^* with tangent vector p, we have

$$f(\gamma(\alpha)) = f(x^*) + \alpha f(x^*)^T p + O(\alpha^2) < f(x^*)$$

as $\alpha \to 0, \alpha > 0$, which contradicts that fact that $f(x^*)$ is local minimum. □

Remark 7.4.1 (the need for algebraic condition)**.** Note that this theorem gives a necessary condition in geometric form, which is not easy to use in practice.(for example, the set $\mathcal{T}(x^*)$ is not explicit.) We need to convert these geometric condition to algebraic conditions for the ease of use. These algebraic condition is known as **KKT** condition.

7.4.2 Constraint qualifications and KKT conditions

Definition 7.4.5 (linearly feasible direction). *Given a feasible point x for the constraint $c(x) \geq 0$ and $\mathcal{A} = \mathcal{A}(x)$, we define the set of **linearly feasible directions** as*

$$\mathcal{T}_L(x) = \{p : J_{\mathcal{A}}(x)p \geq 0\}$$

Remark 7.4.2 (interpretation on linearly feasible)**.** Let x_0 be a feasible point, and let $\mathcal{A}$ denote the active constraint set. Then $c_i(x_0) = 0, \forall i \in \mathcal{A}$. Let $p \in \mathcal{T}_L$, then for sufficiently small $\alpha > 0$, we have

$$c_i(x_0 + \alpha p) \approx c_i(x_0) + \alpha \nabla c_i^T p \geq 0.$$

That is, $x_0 + \alpha p$ is still in feasible region, and p is a feasible direction.

Lemma 7.4.1.

$$\mathcal{T}(x) \subseteq \mathcal{T}_L(x)$$

Proof. Let $r(t)$ be a feasible path for c_i, then $c_i(r(t)) \geq 0$. Take the derivative repsect to r at $t = 0$, we have

$$\nabla c_i \cdot \frac{dr}{dt}|_{t=0} = \nabla c_i \cdot p \geq 0$$

Therefore, if $p \in \mathcal{T}(x)$, we have $\nabla c_i \cdot p \geq 0, i \in \mathcal{A}$ and implies $J_A^T p \geq 0$. □

Definition 7.4.6 (constraint qualification). *We say that the constraint qualification of inequality constraint $c(x) \geq 0$ holds at a feasible point x if every nonzero vector p satisfying $J(x)_A^T p = 0$ implies $p \in \mathcal{T}_L(x)$.*

Remark 7.4.3 (purpose of constraint qualifications).

- constraint qualifications is simple the assumptions on constraint such that later KKT condition relies on
- There are different types of constraint qualifications, such as linear independence constraint qualification, Managasarian-Fromovitz constraint qualification.

Lemma 7.4.2. *If constraint qualifications of inequality constraint $c(x) \geq 0$ holds, then*

$$\mathcal{T}(x) = \mathcal{T}_L(x)$$

Proof. Directly from above lemma $\mathcal{T}(x) \subseteq \mathcal{T}_L(x)$ and the definition and constraint qualification. □

Lemma 7.4.3 (linear constraints always satisfy constraint qualifications). *The constraint qualification of equality constraint $c(x) \geq 0$ holds at x if $c(x) = Ax - b \geq 0$(no matter A is full row rank or not.)*

Proof. Let A_a denote the constraint matrix consists of active constraints. Denote x_0 as the feasible point. Then $A_a x_0 = b_a$. $A_a(x_0 + \alpha p) = b_a$ for sufficiently small $\alpha > 0$ and $p \in \mathcal{N}(A_a)$. Based on the definition, it satisfies the constraint qualification. □

Lemma 7.4.4 (sufficient condition: linear independence constraint qualifications, nonlinear constraint case). *[2, p. 324] The constraint qualification of inequality constraint $c(x) \geq 0$ holds at x if $J_A(x)$ has full row rank(i.e. $\nabla c_i, i \in \mathcal{A}$ are linearly independent of each other). And this is called* ***linear independence constraint qualification****.*

Lemma 7.4.5 (Farkas's lemma, alternative). *Let $g \in \mathbb{R}^n$ and $A \in \mathbb{R}^{r\times n}$. It follows that*

$$g^T p \geq 0, \forall p \in \{p : Ap \geq 0\}$$

if and only if there exists $\lambda \in \mathbb{R}^r, \lambda \geq 0$ such that

$$g = A^T \lambda.$$

Proof. (1) forward. If $g = A^T\lambda$, then $g^T p = \lambda Ap \geq 0$; (2) converse. If $g \notin \mathcal{X} = \{A^T\lambda, \lambda \geq 0\}$, then g is the point lying outside the cone $\mathcal{X}$. Based on Theorem 9.2.3, there exists p such that $Ap \geq 0$(all the basis vectors of the cone lying on one halfspace of a hyperplane passing origin and having norm vector p) and $g^T p < 0$(the element g lying on the other halfspace). This contradicts that for all p such that $Ap \geq 0$, $g^T p \geq 0$. □

Theorem 7.4.2 (first order condition for inequality constraints, KKT condition). *Assume that the first-order constraint qualification holds. If x^* is a local solution to nonlinear optimization with inequality constraint [Definition 7.4.1), then there exists a Lagrange multiplier vector y^* such that*

$$\begin{aligned} c(x^*) &\geq 0 \ (feasiblity) \\ g(x^*) &= \sum_{i\in\mathcal{A}} [y^*]_i \nabla c_i(x^*), [y^*]_i \geq 0, \forall i \in \mathcal{A} \\ [y^*]_i &= 0, \forall i \in \mathcal{I} \end{aligned}$$

Or equivalently,

$$\begin{aligned} c(x^*) &\geq 0 \\ g(x^*) &= J(x^*)^T y^*, y^* \geq 0 \\ c_i(x^*) y_i^* &= 0, \forall i \end{aligned}$$

Proof. When the constraint qualification holds, we have $\mathcal{T}(x^*) = \mathcal{T}_L(x^*) = \{p : J_{\mathcal{A}}(x^*)p \geq 0\}$. The optimality condition [Theorem 7.4.1] says if x^* is a local solution, then $c(x^*) \geq 0$ and $g^T(x^*)p \geq 0, \forall p \in \mathcal{T}_L(x^*) = \{p : J_{\mathcal{A}}(x^*)p \geq 0\}$. The Farkas' lemma [Lemma 7.4.5] says that the gradient at x^* must satisfy

$$g(x^*) = J_{\mathcal{A}}^T y^*, y_{\mathcal{A}}^* \geq 0.$$

Further considering inactive constraints, we can view them as non-existence. So we directly set $y_{\mathcal{I}}^* = 0$. □

Remark 7.4.4 (comparison with equality-constraint optimization).

- In equality constraint problem, there is no constraints on the multipliers' value; in inequality constraint problem, we require the the multipliers' value to be nonnegative.
- The equality constraint problem can be constructed from inequality constraint problems by pairing constraints, i.e. $c(x) = 0 \Leftrightarrow c(x) \geq 0, -c(x) \geq 0$. Then, Theorem 7.4.2 will reduce to

Note 7.4.1 (interpretation of Lagrange multipliers in inequality constraints). [1, lec 4] Assume (x^*, y^*) is a first-order KKT point and $J_{\mathcal{A}}(x^*)$ has full row rank. Let p_i be a vector(such p_i will always exists since $J_{\mathcal{A}}$ is full row rank) such that

$$J_{\mathcal{A}} p_i = e_i, \forall i = 1, ..., |\mathcal{A}|,$$

where e_i is the ith coordinate basis vector. Then moving in the direction p_i will step off the ith active constraint, but still tangent to the other active constraints.

Observe that

$$g(x^*)^T p_i = (y^*_{\mathcal{A}})^T J_{\mathcal{A}}(x^*) p_i = (y^*_{\mathcal{A}})^T e_i = [y^*_{\mathcal{A}}]_i.$$

Therefore, if $[y^*_{\mathcal{A}}]_i > 0$, then moving in the direction p_i will result in initial increasing of the objective function; vice versa. If $[y^*_{\mathcal{A}}]_i = 0$, then moving in the direction p_i will not result in initial change.

In summary,

- A positive Lagrange multiplier associated with an active constraint implies that the objective function initially increases when stepping off of that constraint.
- A zero Lagrange multiplier associated with an active constraint implies that the objective function f initially is 'flat' when stepping off of that constraint. This further implies the minimizer is not strict.
- A negative Lagrange multiplier associated with an active constraint implies that the objective function f initially is decreasing when stepping off of that constraint. This further implies this is not a minimizer.
- If a constraint is inactive, then its Lagrange multipliers must be zero.
- We require $y^* \geq 0$ indicates any deviation of current point will result initial non-decreasing.

Note 7.4.2 (geometry of moving with inequality constraints). Consider a direction vector $p \in \mathbb{R}^n$. Let i denote an active inequality constraint at a feasible x.

- If $\nabla c_i(x)^T p = 0$, then $c_i(x + \alpha\alpha p) \approx c_i(x) + \alpha\nabla c_i(x)^T p = 0$.
- If $\nabla c_i(x)^T p < 0$, then $c_i(x + \alpha\alpha p) \approx c_i(x) + \alpha\nabla c_i(x)^T p < 0$.
- If $\nabla c_i(x)^T p > 0$, then $c_i(x + \alpha\alpha p) \approx c_i(x) + \alpha\nabla c_i(x)^T p > 0$.

7.4.3 Second order conditions

Notations:

$$\mathcal{A}_+(x^*) = \mathcal{A}(x^*) \cap \{i : y_i^* > 0\}$$
$$\mathcal{A}_0(x^*) = \mathcal{A}(x^*) \cap \{i : y_i^* = 0\}$$

Remark 7.4.5 (general remarks on second order condition).

- Second order condition will examine curvature effect along paths for which f is initially 'flat'.
- Simple examples include $f(x) = x^3$ and $f(x) = x^4$. Only second order condition can help distinguish optimality at $x = 0$.
- If $y_i^* > 0$, then from 7.4.1 we know that stepping off i constraint will increase objective function; however, move along the constraint surface will not.
- If $y_i^* = 0$ for an active constraint, stepping off i constraint or move along i will be flat.

Definition 7.4.7 (initial flat linearized direction set).

$$\mathcal{S} = \{p : \nabla c_i(x^*)^T p = 0, \forall i \in \mathcal{A}_+(x^*) \cup \nabla c_i(x^*)^T p \geq 0, \forall i \in \mathcal{A}_0(x^*)\}$$

Remark 7.4.6. If $\nabla c_i(x^*)^T p = 0$, then move along p will maintain the acitve constraint i active; If $\nabla c_i(x^*)^T p > 0$, then move along p will step off the acitve constraint i into feasible region; If $\nabla c_i(x^*)^T p < 0$, then move along p will step off the acitve constraint i into infeasible region.

Lemma 7.4.6. *For any $p \in \mathcal{S}$,*

$$g^T p = 0.$$

Proof. Note that

$$g(x^*)^T p = \sum_{i \in \mathcal{A}} y_i^* (\nabla c_i(x^*))^T p.$$

If $y_i^* > 0$, we require $\nabla c_i(x^*)^T p = 0$. If $y_i^* = 0$, we relax to $\nabla c_i(x^*)^T p \geq 0$. $\square$

Definition 7.4.8 (second order constraint qualification for inequality problems). *The second order constraint qualification for inequality constraint $c(x) \geq 0$ holds at a KKT poyint (x^*, y^*) if for all y in*

$$\mathcal{M}(x^*) \triangleq \{y : g(x^*) = J(x^*)^T y, y \geq 0 \text{ and } c_i(x^*) * y_i = 0\}$$

It follows that every nonzero direction p in the set

$$\mathcal{S} = \{p : \nabla c_i(x^*)^T p = 0, \forall i \in \mathcal{A}_+(x^*) \cup \nabla c_i(x^*)^T p = 0, \forall i \in \mathcal{A}_0(x^*)\}$$

is tangent to a twice-differentiable path $x(\alpha)$ such that $c_{\mathcal{A}_+}(x(\alpha)) = 0$ and $c_{\mathcal{A}_0}(x(\alpha)) \geq 0$ for all $0 < \alpha \leq \sigma$ and some $\sigma > 0$.

Remark 7.4.7 (interpretation). The set of twice-differentiable path $x(\alpha)$ such that $c_{\mathcal{A}_+}(x(\alpha)) = 0$ and $c_{\mathcal{A}_0}(x(\alpha)) \geq 0$ for all $0 < \alpha \leq \sigma$ and some $\sigma > 0$ **is a set of path that initially flat and feasible**. Usually, the tangent vector set of this set of paths might be different from $\mathcal{S}$. The constraint qualification says that they are equivalent.

Remark 7.4.8 (comparison with first order constraint qualification).

- The first order constraint qualification does not imply the second order constraint qualification.
- The second order constraint qualification does not imply the first order constraint qualification.
- If active constraints are linear, then both first and second order constraint qualification hold.
- Linear independence qualification is sufficient for both first and second order constraint qualification.

Theorem 7.4.3 (second-order necessary conditions). *[1, lec 4] Suppose that $f, c \in C^2$. If x^* is local constrained minimizer of the inequality constrained problem at which the first and the second order constraint qualifications are satisfied, then there exists a vector of Lagrange multiplier y^* such that*

- $c(x^*) \geq 0$
- $g(x^*) = J(x^*)^T y^*, y^* \geq 0$
- $c(x^*) \cdot y^* = 0$*(complementary slackness)*
- $$p^T [\sum_{i=1}^{m} -\lambda_i^* \nabla^2 c_i(x^*) + H(x^*)] p \geq 0$$

  holds for all $p \in \mathcal{S}$.

Theorem 7.4.4 (second-order sufficient condition). *[1, lec 4] Suppose that $f, c \in C^2$. If there exist (x^*, y^*) such that*

- $c(x^*) \geq 0$
- $g(x^*) = J(x^*)^T y^*, y^* \geq 0$
- $c(x^*) \cdot y^* = 0$ *(complementary slackness)*
- $$p^T[\sum_{i=1}^{m} -\lambda_i^* \nabla^2 c_i(x^*) + H(x^*)]p \geq \omega \|p\|$$
  for some $\omega > 0$ and for all $p \in \mathcal{S}$.

Then, x^* is local constrained minimizer.

Example 7.4.1. Consider a trust region subproblem [Theorem 6.3.1]

$$\min_{s \in \mathbb{R}^n} m(s) = f + s^T g + \frac{1}{2} s^T B s, \text{ subject to } \|s\|_2 \leq \delta.$$

Note that we can write the trust region constraint as $-\frac{1}{2}s^T s \geq -\frac{1}{2}\delta^2$, whose $J(s) = -s$

A vector s^* is a local minimizer if and only if $\|s^*\| \leq \delta$ and there exists a scalar $\lambda^* \geq 0$ such that

- $\|s^*\| \leq \delta$.
- $(g + Bs) = -\lambda^* s^*$ gives $(B + \lambda^* I)s^* = -g$.
- (complementary slackness)$\lambda^*(\|s\|_2 - \delta) = 0$
- $B + \lambda^* I$ is positive definite.

Definition 7.4.9 (strict complementarity). *We say that strict complementarity holds at the KKT point x^* if there exists a Lagrange multiplier vector y^* such that $y^* > 0$ for all $i \in \mathcal{A}(x^*)$.*

Remark 7.4.9 (why require strict complementarity). If a Lagrange multiplier y_i associated with an active constraint is zero, then it implies the minimizer is not strict. See 7.4.1.

Corollary 7.4.4.1 (second-order sufficient condition). *[1, lec 4] Suppose that $f, c \in C^2$. If there exist (x^*, y^*) such that*

- $c(x^*) \geq 0$
- $g(x^*) = J(x^*)^T y^*, y^* \geq 0$

- *$c(x^*) \cdot y^* = 0$(complementary slackness)*
- ***strict complementarity is satisfied**.*
- $$p^T[\sum_{i=1}^{m} -\lambda_i^* \nabla^2 c_i(x^*) + H(x^*)]p \geq \omega \|p\|_2^2$$

  for some $\omega > 0$ and for all $p \in \mathcal{N}(J(x^*))$.

Then, x^* is local constrained minimizer.

Proof. Based on the definition, strict complementarity means that $\mathcal{A}_0 = \emptyset$ and $\mathcal{S} = \mathcal{N}(J_{\mathcal{A}}(x^*))$. □

7.5 Envelope theorem and sensitive analysis

Lemma 7.5.1 (sensitivity of unconstrained optimization). *[5, p. 369] Consider a maximization problem given by*

$$M(a) = \max_{x \in \mathbb{R}^n} f(x,a)$$

where a is an external parameter. Denote the maximizers of f as $x(a)$.Assume $M(a)$ and $x(a)$ are differentiable with respect to a. Then

$$\frac{dM(a)}{da} = \frac{\partial f(x,a)}{\partial a}|_{x=x(a)}.$$

Proof.

$$\frac{dM(a)}{da} = \sum_{i=1}^{n} \frac{\partial f(x,a)}{\partial x_i}\frac{dx_i(a)}{da} + \frac{\partial f(x,a)}{\partial a}.$$

Note that the first term of the right-hand-side is zero due to first order necessary condition of unconstrained optimization

$$\frac{\partial f}{\partial x_i} = 0, i = 1,2,...,n.$$

□

Lemma 7.5.2 (sensitivity of equality constrained optimization,envelope theorem). *[5, p. 369][6, p. 605] Consider a maximization problem given by*

$$M(a) = \max_{x \in \mathbb{R}^n} f(x,a),\ s.t.\ g_i(x,a) = 0, j = 1,2,...,m$$

where a is an external parameter. Denote the maximizers of f as $(x(a), \lambda(a))$. Denote the Lagrange function as

$$L(x,\lambda,a) = f(x,a) - \sum_{j=1}^{m} \lambda_j g_j(x,a).$$

Assume $M(a), x(a), \lambda(a)$ are differentiable with respect to a. Then

$$\frac{dM(a)}{da} = \frac{\partial L(x,\lambda,a)}{\partial a}|_{x=x(a),\lambda=\lambda(a)}.$$

Proof.

$$\frac{dM(a)}{da} = \nabla_x f(x,a) \cdot \frac{dx}{da} + \frac{\partial f(x,a)}{\partial a}.$$

Use the first order condition on the gradient, we have

$$\nabla_x f(x,a) = \sum_{j=1}^{m} \lambda_j \nabla_x g_j(x(a),a).$$

Use the feasibility condition $g_j(x(a),a) = 0, j = 1,2,...,m$ and take derivative with respect to a, we have

$$\nabla_x g_j \cdot \frac{dx}{da} + \frac{\partial g_j}{\partial a} = 0.$$

Therefore, we have

$$\frac{dM(a)}{da} = -\sum_{j=1}^{m} \lambda_j \frac{\partial g_j(x,a)}{\partial a} + \frac{\partial f(x,a)}{\partial a} = \frac{\partial L(x,\lambda,a)}{\partial a}.$$

□

Remark 7.5.1 (same conclusion for minimization problem). For the minimization problem, the same conclusion will hold. Note in the proof we do not distinguish whether we are minimizing or maximizing.

Corollary 7.5.0.1 (sensitivity of equality constrained optimization). *[5, p. 369][6, p. 605] Consider a maximization problem given by*

$$M(a) = \max_{x\in\mathbb{R}^n} f(x),\ s.t.\ g_i(x) = a_i, j = 1,2,...,m$$

where a is an external parameter. Denote the maximizers of f as $(x(a),\lambda(a))$. Denote the Lagrange function as

$$L(x,\lambda,a) = f(x,a) - \sum_{j=1}^{m} \lambda_j g_j(x,a).$$

Assume $M(a), x(a), \lambda(a)$ are differentiable with respect to a. Then

$$\frac{\partial M(a)}{\partial a} = \lambda_i.$$

Proof. Note that

$$L(x,\lambda,a) = f(x) - \sum_{i=1}^{m} \lambda_i (g_i(x) - a_i) \implies \partial L/\ Paa_i = \lambda_i.$$

□

Lemma 7.5.3 (sensitivity of inequality constrained optimization). *[5, p. 369][6, p. 605] Consider a maximization problem given by*

$$M(a) = \max_{x \in \mathbb{R}^n} f(x,a), \; s.t. \; g_i(x,a) \geq 0, j = 1, 2, ..., m$$

where a is an external parameter. Denote the maximizers of f as $(x(a), \lambda(a))$. Denote the Lagrange function as

$$L(x, \lambda, a) = f(x,a) - \sum_{j=1}^{m} \lambda_j g_j(x,a).$$

Assume $M(a), x(a), \lambda(a)$ are differentiable with respect to a. Then

$$\frac{dM(a)}{da} = \frac{\partial L(x, \lambda, a)}{\partial a}|_{x=x(a), \lambda=\lambda(a)}.$$

Proof. Note that at optimality, λ_i is 0 for inactive constraints. Therefore, we should only consider the problem as an equality constrained problem [Lemma 7.5.2]. □

7.6 Notes on bibliography

For introductory level of optimization theory, see [7].

For intermediate treatment, see [1][2].

BIBLIOGRAPHY

1. Robinson, D. *Nonlinear optimization II lecture notes* (Johns Hopkins University, 2015).
2. Nocedal, J. & Wright, S. *Numerical optimization* (Springer Science & Business Media, 2006).
3. Bertsekas, D. *Nonlinear programming* ISBN: 9781886529007 (Athena Scientific, 2016).
4. Robinson, D. *Nonlinear optimization I lecture notes* (Johns Hopkins University, 2015).
5. Jiang, D. *advanced microeconomics (chineses)* (Peking University Publisher, 2000).
6. Jehle, G. & Reny, P. *Advanced Microeconomic Theory* ISBN: 9780273731917 (Financial Times/Prentice Hall, 2011).
7. Sundaram, R. K. *A first course in optimization theory* (Cambridge university press, 1996).