Chapter 12 – Theory of Constrained Optimization

12.1 Introduction

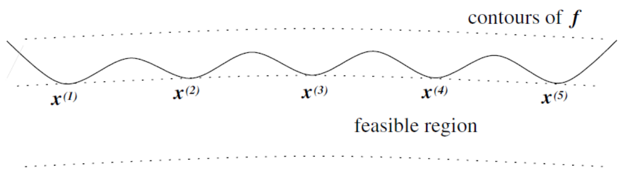
A general formulation for constrained optimization problem is.

$$\min_{x\in\mathbb{R}^n} f(x) \text{ s.t. } \begin{cases} c_i(x) = 0, & i \in E \\ c_i(x) \geq 0, & i \in I \end{cases} \tag{12.1}$$

where f and c_i are smooth, real-valued functions on a subset of $\mathbb{R}^n$, and E and I are two finite sets of indices.

A feasible set Ω is the sets of point x that satisfy the constraints, such that

$$\min_{x\in\Omega} f(x) \qquad \Omega = \{x \mid \; c_i(x) = 0, i \in E; \quad c_i(x) \geq 0, i \in I\} \tag{12.3}$$



A vector x^* is a

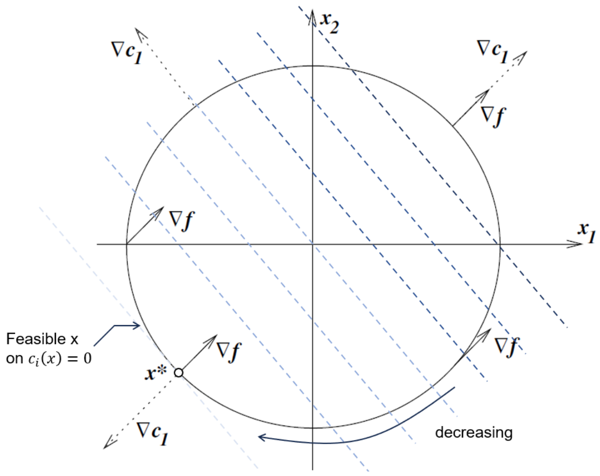
- Local solution if $x^* \in \Omega$ and there is a neighborhood $\mathcal{N}$ of x^* such that $f(x) \geq f(x^*)$ for all $x^* \in \mathcal{N} \cap \Omega$
- Strict local solution if $x^* \in \Omega$ and there is a neighborhood $\mathcal{N}$ of x^* such that $f(x) > f(x^*)$ for all $x^* \in \mathcal{N} \cap \Omega$ with $x \neq x^*$.
- Isolated local solution if $x^* \in \Omega$ and there is a neighborhood $\mathcal{N}$ of x^* such that x^* is the only solution in $\mathcal{N} \cap \Omega$.

Definition 12.1 – Active set

Active set $\mathcal{A}(x)$ at any feasible x consists of the equality constraint indices from E together with the indices of the inequality constraints i for which $c_i(x) = 0$; that is,

$$\mathcal{A}(x) = E \cup \{i \in I \mid c_i(x) = 0\}$$

At any feasible point x, the inequality constraint $i \in I$ is active if $c_i(x) = 0$ and inactive if strict inequality $c_i(x) > 0$ is satisfied.



At the solution x^* the constraint normal $\nabla c_1(x^*)$ is parallel to $\nabla f(x^*)$ such that

$$\nabla f(x^*) = \lambda_1^* \nabla c_1(x^*) \tag{12.10}$$

With first-order Taylor series approximation to the objectives and constraint functions, there is a direction d such that it would satisfy the equality constraint and create a decrease in the objective with

$$\nabla c_1(x)^T d = 0 \text{ and } \nabla f(x)^T d < 0 \tag{12.14}$$

If $\nabla f(x)$ and $\nabla c_1(x)$ are not parallel, we can set the following to create a decrease

$$\bar{d} = -\left(I - \frac{\nabla c_1(x)\nabla c_1(x)^T}{\|\nabla c_1(x)\|^2}\right)\nabla f(x); \quad d = \frac{\bar{d}}{\|\bar{d}\|} \tag{12.15}$$

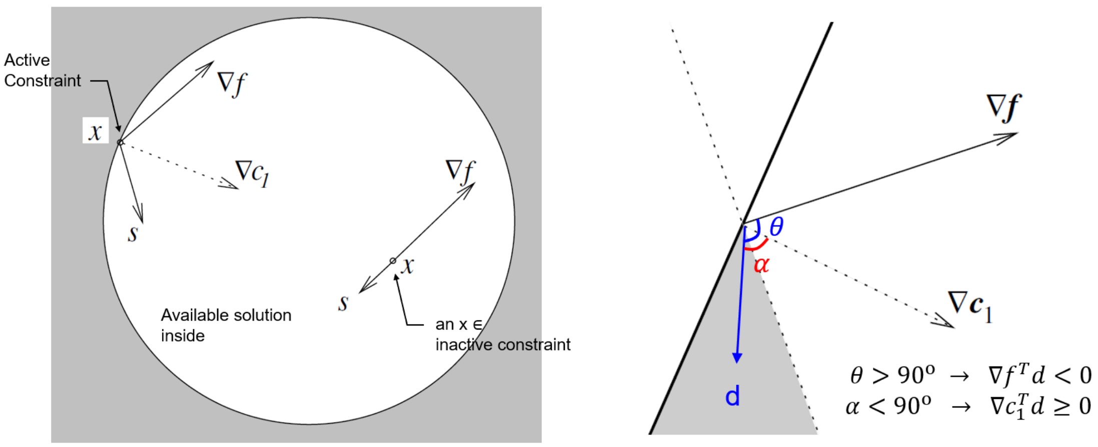
Introduce the Lagrangian Function

$$L(x, \lambda_1) = f(x) - \lambda_1 c_1(x) \quad \rightarrow \quad \nabla_x L(x, \lambda_1) = \nabla f(x) - \lambda_1 \nabla c_1(x) \tag{12.16}$$

the condition (12.10) is equivalent to

$$\nabla_x L(x^*, \lambda_1^*) = \nabla f(x^*) - \lambda_1^* \nabla c_1(x^*) = 0 \quad \rightarrow \quad \nabla f(x^*) = \lambda_1 \nabla c_1(x^*) \tag{12.17}$$

where $(.)^*$ denotes the optimal value.



Case 1 $(c_1 > 0,\ \lambda = 0)$: x lies strictly inside circle $\rightarrow \ s = -\alpha \nabla f \quad$ (s does not create a step if $\nabla f(x) = 0$)

Case 2 $(c_1 = 0,\ \lambda \geq 0)$: x lies on the boundary of the circle with $c_1 = 0. \ \rightarrow \ \nabla f(x)^T s < 0,\ \nabla c_1(x)^T s \geq 0$

It is clear from the figure that the available region is empty (in Case 2), i.e. no further descent direction, if $\nabla f(x)$ and $\nabla c_1(x)$ point in the same direction, i.e. $\nabla f(x) = \lambda_1 \nabla c_1(x)$ for some $\lambda_1 \geq 0$. It is noted that the sign of λ_1 is significant to indicate the feasible set.

When no first-order feasible descent direction exists at some points x^*, we have

$$\nabla_x L(x^*, \lambda_1^*) = 0 \ \ (\lambda_1^* > 0) \tag{12.22}$$

and

$$\lambda_1^* c_1(x^*) = 0 \tag{12.23}$$

The complementarity condition (12.23) implies that the Lagrange multiplier λ_1^* can be strictly positive only when the corresponding constraints $c_1(x^*)$ is active.

12.2 Tangent Cone and Constraint Qualification

Definition 12.2 – Tangent Cone

The vector d is said to be tangent (or tangent vector) to Ω at a point x if there are a feasible sequence $\{z_k\}$ approaching x and a sequence of positive scalars $\{t_k\}$ with $t_k \rightarrow 0$ such that

$$\lim_{k \to \infty} \frac{z_k - x}{t_k} = d \tag{12.29}$$

The set of all tangents to Ω at x^* is called the tangent cone and is denoted by $T_\Omega(x^*)$.

A cone is a set $\mathcal{F}$ with the property that for all $x \in \mathcal{F}$,

$$x \in \mathcal{F} \rightarrow \alpha x \in \mathcal{F} \qquad \forall\, \alpha > 0 \tag{A.36}$$

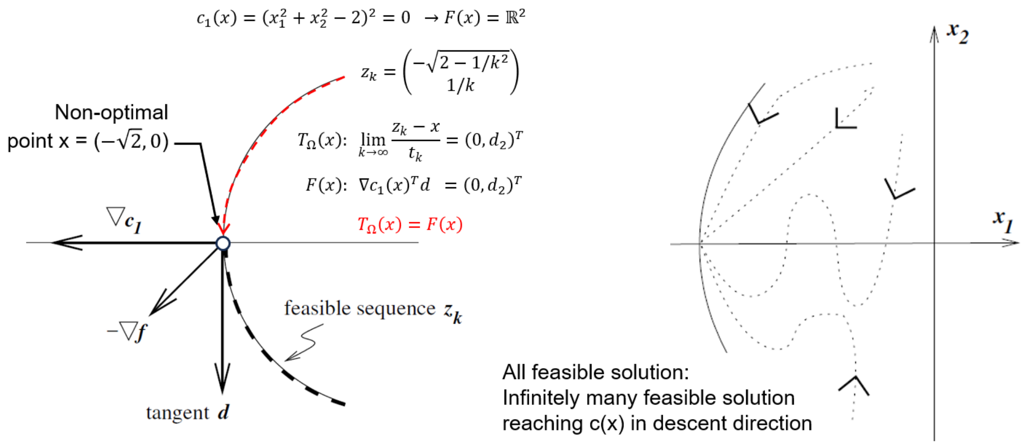
Definition 12.3 – Feasible Direction

Given a feasible point $x \in \Omega$ and the active constraint set $\mathcal{A}(x)$ of Definition 12.1, the set of linearized feasible direction $\mathcal{F}(x)$ is

$$\mathcal{F}(x) = \left\{ d \;\middle|\; \begin{matrix} d^T \nabla c_i(x) = 0 & \forall\, i\ \in E \\ d^T \nabla c_i(x) \geq 0 & \forall\, i\ \in I \end{matrix} \right\} \tag{12.30}$$

It is important to note that

- Definition of tangent cone relies only on its geometry, not on its algebraic specification of the set Ω.
- The linearized feasible set depends on the definition of constraint function $c_i,\ i \in E \cup I$



Constraint Qualifications are the conditions under which the linearized feasible set $\mathcal{F}(x)$, which is constructed by linearizing the algebraic description of the set Ω at any x, capture the essential geometric features of the set Ω in the vicinity of x, as represented by tangent cone $T_\Omega(x^*)$.

12.3 First Order Condition and its Proof

Definition 12.4 – Linear Independence Constraint Qualification (LICQ)

Given a point x and the active set $\mathcal{A}(x)$ defined in Definition 12.1, LICQ holds if the set of active constraint gradient

$$\{\nabla c_i(x), i \in \mathcal{A}(x)\} \text{ is linearly independent} \tag{12.32}$$

In general, if LICQ holds, none of the active constraint gradients $\{\nabla c_i(x), i \in \mathcal{A}(x)\}$ can be zero.

Define the Lagrangian Function for the general problem (12.1)

$$L(x, \lambda) = f(x) - \sum_{i \in E \cup I} \lambda_i c_i(x) \tag{12.33}$$

Theorem 12.1 – KKT Condition (First-order Necessary Condition)

Suppose 1) x^* is a local solution of (12.1)

2) $f(x)$ and $c_i(x)$ are continuously differentiable

3) LICQ holds at x^*, that is $\{\nabla c_i(x), i \in \mathcal{A}(x)\}$ is linearly independent

Then there is a Lagrange multiplier vector λ^*, with components $\lambda_i,\ i \in E \cup I$, such that the following conditions are satisfied with optimal (x^*, λ^*)

$$\nabla_x L(x^*, \lambda^*) = 0 \tag{12.34a}$$

$$c_i(x) = 0 \quad \forall i \in E \tag{12.34b}$$

$$c_i(x) \geq 0 \quad \forall i \in I \tag{12.34c}$$

$$\lambda_i^* \geq 0 \quad \forall i \in I \tag{12.34d}$$

$$\lambda_i^* c_i(x^*) = 0 \ \ \forall i \in E \cup I \tag{12.34e}$$

The condition (12.34e) are complementarity conditions, which implies either constraint i is active such that $c_i(x) = 0$ or $\lambda_i^* = 0$, which does not impact the optimization, or possibly both.

In fact, the inactive inequality constraint does not create an impact on the optimization problem. Hence the condition can be written as

$$\nabla_x L(x^*, \lambda^*) = \nabla f(x^*) - \sum_{i \in \mathcal{A}(x^*)} \lambda_i^* \nabla c_i(x^*) = 0 \quad \rightarrow \quad \nabla f(x^*) = \sum_{i \in \mathcal{A}(x^*)} \lambda_i^* \nabla c_i(x^*) \tag{12.35}$$

Note: For a given problem (12.1) and solution x^*, there may be many λ_i^* for which KKT condition is satisfied. When LICQ holds, however, the optimal λ_i^* is unique.

Definition 12.5 – Strict Complementarity

Given a local solution x^* of (12.1) and a vector λ^* satisfying KKT condition, the strict complementarity condition holds if exactly one of λ_i^* and $c_i(x^*)$ is zero for each index $i \in I$. In other words, $\lambda_i^* > 0$ for each $i \in I \cap \mathcal{A}(x^*)$.

Note: Satisfaction of strict complementarity conditions usually makes it easier for algorithms to determine the active sets $\mathcal{A}(x^*)$ and converge rapidly to the solution x^*.

Denote $A(x^*)$ with rows as the active constraint gradients at the optimal point, that is

$$A(x^*)^T = [\nabla c_i(x^*)]_{i \in \mathcal{A}(x^*)} \tag{12.37}$$

Lemma 12.2

Let x^* be a feasible point. The following two statement are true.

(i) $T_\Omega(x^*) \subset \mathcal{F}(x)$

(ii) If LICQ condition is satisfied at x^*, then $T_\Omega(x^*) = \mathcal{F}(x)$

Proof of (ii):

Since LICQ condition holds, the $m \times n$ matrix $A(x^*)$ of active constraint gradients has full row rank m. Let Z be a matrix whose columns are a basis for the nullspace of $A(x^*)$; that is

$$Z \in \mathbb{R}^{n \times (n-m)}, \quad Z \text{ has full column rank,} \quad A(x^*)_{m \times n} Z_{n \times (n-m)} = 0_{m \times (n-m)} \tag{12.39}$$

Choose $d \in \mathcal{F}(x^*)$ arbitrarily and suppose $\{t_k\}_{k=0}^{\infty}$ is any sequence of positive scalars such that $t_k \rightarrow 0$ when $k \rightarrow \infty$. Define the parametrized system of equations R: $\mathbb{R}^n \times \mathbb{R} \rightarrow \mathbb{R}^n$ by

$$R(z, t) = \begin{bmatrix} c(z) - tA(x^*)d \\ Z^T(z - x^* - td) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{12.40}$$

The solution $z = z_k$ of the systems for small $t = t_k > 0$ give a feasible sequence that approaches x^* and satisfies

$$\lim_{k \rightarrow \infty} \frac{z_k - x}{t_k} = d$$

At $t = 0, z = x^*$, and the Jacobian of R at this point is

$$\nabla_z R(x^*, 0) = \begin{bmatrix} A(x^*) \\ Z^T \end{bmatrix} \tag{12.41}$$

which is non-singular by construction of Z. Hence, according to the implicit function theorem, the system (12.40) has a unique solution z_k for all values of t_k sufficiently small.

Moreover, from (12.40) and Definition 12.3 that

$$i \in E \rightarrow \ c_i(z_k) = t_k \nabla c_i(x^*)^T d = 0 \tag{12.42a}$$

$$i \in \mathcal{A}(x^*) \cap I \rightarrow \ c_i(z_k) = t_k \nabla c_i(x^*)^T d \geq 0 \tag{12.42b}$$

which is the linearized feasible direction requirement, so that z_k is indeed feasible.

It remains to verify that $\lim_{k\to\infty} \frac{z_k - x}{t_k} = d$ holds for the choice of $\{z_k\}$. Using the fact that $R(z_k, t_k) = 0$ for all k together with Taylor's theorem, it is found that

$$R(z_k, t_k) = \begin{bmatrix} c(z_k) - t_k A(x^*) d \\ Z^T(z_k - x^* - t_k d) \end{bmatrix} = \begin{bmatrix} A(x^*)(z_k - x^*) + o(||z_k - x^*||) - t_k A(x^*) d \\ Z^T(z_k - x^* - t_k d) \end{bmatrix}$$

$$= \begin{bmatrix} A(x^*) \\ Z^T \end{bmatrix} (z_k - x^* - t_k d) + o(||z_k - x^*||) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Dividing the expression by t_k and using non-singularity of coefficient matrix in the first term,

$$\frac{z_k - x^*}{t_k} = d + o\left(\frac{||z_k - x^*||}{t_k}\right)$$

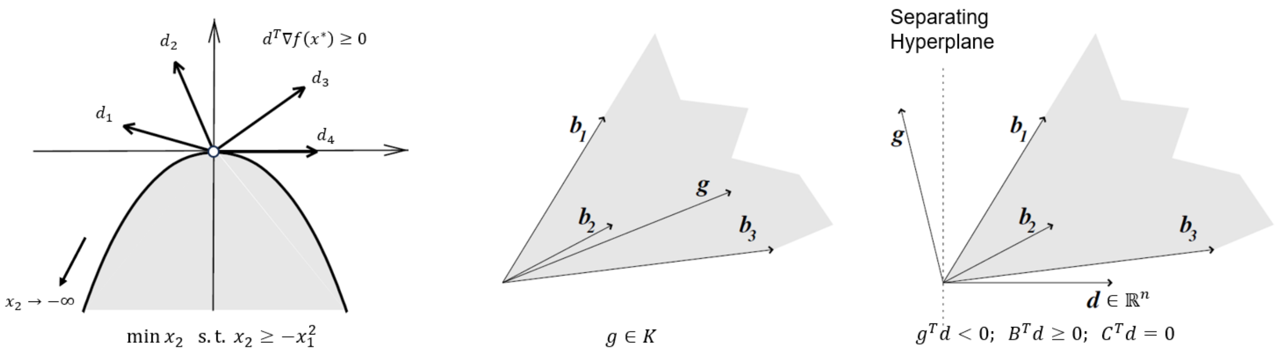
for which the tangent cone requirement is satisfied.

Hence, $d \in T_\Omega(x^*)$ for an arbitrary $d \in \mathcal{F}(x)$ and $T_\Omega(x^*) = \mathcal{F}(x^*)$.

Theorem 12.3

$$x^* \text{ is local solution of (12.1)} \rightarrow \ \nabla f(x^*)^T d \geq 0 \quad \forall d \in T_\Omega(x^*) \tag{12.43}$$

Note: The converse may not be true, i.e. $\nabla f(x^*)^T d \geq 0 \ \forall d \in T_\Omega(x^*)$, but x^* is not a local minimizer.



Lemma 12.4 – Farkas' Lemma

Consider a cone K defined where $B \in \mathbb{R}^{n\times m}$, $C \in \mathbb{R}^{n\times p}$ as

$$K = \{By + Cw \mid y \geq 0\} \tag{12.45}$$

Given a vector $g \in \mathbb{R}^n$. Only one among the two conditions are true.

1. $g \in K \ \rightarrow \ g^T d \geq 0$
2. There exists a $d \in \mathbb{R}^n$ such that $g^T d < 0, \ \ B^T d \geq 0, \ \ C^T d = 0,$
   where d defines a separating hyperplane that separates g from the cone K.

By applying Lemma 12.4 to the cone N defined by

$$N = \left\{ \sum_{i \in \mathcal{A}(x^*)} \lambda_i \nabla c_i(x^*), \quad \lambda_i \geq 0 \text{ for } i \in \mathcal{A}(x^*) \cap I \right\} \tag{12.50}$$

and setting $g = \nabla f(x^*)$, we have any one of the following conditions is true.

1. $\nabla f(x^*) = \sum_{i \in \mathcal{A}(x^*)} \lambda_i \nabla c_i(x^*) = A(x^*)^T \lambda_i, \quad \lambda_i \geq 0 \text{ for } i \in \mathcal{A}(x^*) \cap I$ (12.51)
2. There is a direction d such that $d^T \nabla f(x^*) < 0$ and $d \in \mathcal{F}(x^*)$

12.5 Second Order Conditions

The first order condition (KKT conditions) describes how $\nabla f(x)$ and active constraints $c_i(x)$ are related to each other at a solution x^*. When KKT conditions are satisfied, a move along any vector w from $\mathcal{F}(x^*)$ either increases the first order approximation to the objective function (i.e. $w^T \nabla f(x^*) > 0$) or keeps this value the same (i.e. $w^T \nabla f(x^*) = 0$).

When $w^T \nabla f(x^*) = 0$, it is not possible to tell whether the objective function f will increase or decrease along the direction w. Hence the second order condition is needed.

Given 1. $\mathcal{F}(x^*)$ from Definition 12.3,

2. some Lagrange multiplier λ_i satisfy KKT condition (12.34), and

3. f and $c_i, i \in E \cup I$ are twice differentiable.

Critical Cone $C(x^*, \lambda_i^*)$ is defined as

$$C(x^*, \lambda_i^*) = \{ w \in \mathcal{F}(x^*) \mid \nabla c_i(x^*)^T w = 0 \quad \forall i \in \mathcal{A}(x) \cap I, \lambda_i \geq 0 \}$$

or equivalently,

$$w \in C(x^*, \lambda_i^*) \leftrightarrow \begin{cases} \nabla c_i(x^*)^T w = 0 & \forall i \in E, \\ \nabla c_i(x^*)^T w = 0, & \forall i \in \mathcal{A}(x) \cap I,\ \lambda_i > 0 \\ \nabla c_i(x^*)^T w \geq 0, & \forall i \in \mathcal{A}(x) \cap I,\ \lambda_i = 0 \end{cases} \tag{12.53}$$

The critical cone contains those directions w that would tend to "adhere" to the active inequality constraints even when small changes to the objective and equality constraints are made.

From the definition (12.53) and the fact that $\lambda_i = 0$ for all inactive components $i \in I \backslash \mathcal{A}(x)$, it follows that

$$w \in C(x^*, \lambda_i^*) \rightarrow \quad \lambda_i^* \nabla c_i(x^*)^T w = 0 \quad \forall i \in E \cup I \tag{12.54}$$

From the first KKT condition ($\nabla_x L(x^*, \lambda^*) = 0$) and definition of Lagrange function (12.33),

$$w \in C(x^*, \lambda_i^*) \rightarrow \quad w^T \nabla f(x^*) = \sum_{i \in E \cup I} \lambda_i w^T \nabla c_i(x^*) \tag{12.55}$$

Theorem 12.5 – Second-order Necessary Condition

Suppose 1. x^* is a local solution of (12.1),

2. LICQ condition is satisfied, and

3. λ^* is the Lagrange multiplier vector for which KKT conditions (12.34) are satisfied.

Then

$$w^T \nabla_{xx}^2 L(x^*, \lambda^*) w \geq 0 \quad \forall w \in C(x^*, \lambda^*) \tag{12.57}$$

Theorem 12.6 – Second-order Sufficient Condition

Suppose 1. for some $x^* \in \mathbb{R}^n$ there is a λ^* such that the KKT condition (12.34) are satisfied,

2. $$w^T \nabla^2_{xx} L(x^*, \lambda^*) w > 0 \quad \forall w \in C(x^*, \lambda^*),\ w \neq 0 \tag{12.65}$$

Then x^* is a strict local solution for (12.1).

Note: If $\nabla^2_{xx} L(x^*, \lambda^*)$ is positive definite, the second-order condition is satisfied.

The second-order conditions are sometimes stated in a form that is slightly weaker but easier to verify than (12.57) and (12.65). This form uses a two-sided projection of the Lagrangian Hessian $\nabla^2_{xx} L(x^*, \lambda^*)$ onto a subspaces that are related to $C(x^*, \lambda^*)$.

The simplest case is obtained when the multiplier λ^* that satisfies KKT conditions (12.34) is unique (e.g. when LICQ condition holds), and strict complementarity holds. In this case, the definition of Critical cone (12.53) of $C(x^*, \lambda_i^*)$ reduces to

$$C(x^*, \lambda^*) = N(\nabla c_i(x^*)^T)_{i \in \mathcal{A}(x^*)} = N(A(x^*))$$

where $A(x^*)$ is the active constraint gradient defined as in (12.37).

As in (12.39), a matrix Z was defined with full column rank whose columns span the space $C(x^*, \lambda^*)$; that is

$$C(x^*, \lambda^*) = \{Zu \mid u \in \mathbb{R}^{|\mathcal{A}(x^*)|}\}$$

Hence, the condition (12.57) in **Theorem 12.5** can be restated as

$$w^T \nabla^2_{xx} L(x^*, \lambda^*) w \geq 0 \quad \rightarrow \quad u^T Z^T \nabla^2_{xx} L(x^*, \lambda^*) Z u \geq 0 \quad \forall u$$

or $Z^T \nabla^2_{xx} L(x^*, \lambda^*) Z$ is positive semi-definite.

Z can be computed numerically, so that the positive (semi)definiteness conditions can actually be checked by forming these matrices and finding their eigenvalues. One way to compute Z is to apply QR factorization to the matrix of active constraint gradients whose nullspace we seek.

$$A(x^*)^T = Q \begin{pmatrix} R \\ 0 \end{pmatrix} = (Q_1 \quad Q_2) \begin{pmatrix} R \\ 0 \end{pmatrix} = Q_1 R \tag{12.74}$$

where R is a square upper triangular matrix and Q is $n \times n$ orthogonal.

Note:

- If R is non-singular, $Z = Q_2$.
- If R is singular (indicating the active constraint gradients are linearly dependent), a slight enhancement of this procedure that make use of column pivoting during the QR procedure can be used to identify Z.

12.6 Other Constraint Qualifications

Constraint qualifications are the conditions to ensure that the linearized approximation to the feasible set Ω capture the essential shape of Ω in a neighborhood of x^*.

One situation in which the linearized feasible direction set $\mathcal{F}(x)$ is obviously an adequate representation of the actual feasible set occurs when all the active constraints are already linear; that is

$$c_i(x) = a_i^T x + b_i \tag{12.75}$$

for some $a_i \in \mathbb{R}^n$ and $b_i \in \mathbb{R}$.

Lemma 12.7

Suppose at some $x^* \in \Omega$, all active constraints $c_i(x^*),\ i \in \mathcal{A}(x)$, are linear functions. Then $T_\Omega(x^*) = \mathcal{F}(x)$.

Another useful generalization of the LICQ is the Mangasarian-Fromovitz constraint qualification (MFCQ).

Definition 12.6 – MFCQ

MFCQ holds if there exists a vector $w \in \mathbb{R}^n$

$$\left\{ w \in \mathbb{R}^n \;\middle|\; \begin{matrix} \nabla c_i(x^*)^T w > 0 & \forall i \in \mathcal{A}(x^*) \cap I \\ \nabla c_i(x^*)^T w = 0 & \forall i \in E \end{matrix} \right\} \to \text{MFCQ}$$

and the sets of equality constraints gradient $\{\nabla c_i(x^*), \forall i \in E\}$ are linearly independent.

Note: Strict inequality involves the active inequality constraints.

The MFCQ is a weaker condition than LICQ. If LICQ is satisfied, the system of equalities defined by

$$\left\{ w \in \mathbb{R}^n \;\middle|\; \begin{matrix} \nabla c_i(x^*)^T w = 1 & \forall i \in \mathcal{A}(x^*) \cap I \\ \nabla c_i(x^*)^T w = 0 & \forall i \in E \end{matrix} \right\}$$

has a solution w, by full rank of the active constraint gradient. Hence, w of **Definition 12.6** can be chosen precisely this vector.

For MFCQ, it is equivalent to the boundedness (or uniqueness as LICQ) of the set of Lagrange Multiplier vector λ^* for which KKT conditions (12.34) are satisfied.

12.7 A Geometric Viewpoint

An alternative first-order optimality condition that depends only on the geometry of the feasible set Ω and not on its algebraic description in terms of the constraint function $c_i,\ i \in E \cup I$.

$$\min f(x) \quad \text{s.t. } x \in \Omega \tag{12.76}$$

where Ω is the feasible set.

Definition 12.7 – Normal Cone

The normal cone to the set Ω at the point $x \in \Omega$ is defined as

$$N_\Omega(x) = \{v \mid v^T w \leq 0 \quad \forall\, w \in T_\Omega(x)\} \tag{12.77}$$

where $T_\Omega(x)$ is the tangent cone of Definition 12.2. Each vector $v \in N_\Omega(x)$ is said to be a normal vector.

Geometrically, each normal vector v makes an angle of at least $\pi/2$ with every tangent vector.

Recall cone N described as a sum of active constraint gradient

$$N = \left\{ \sum_{i \in \mathcal{A}(x^*)} \lambda_i \nabla c_i(x^*), \quad \lambda_i \geq 0 \text{ for } i \in \mathcal{A}(x^*) \cap I \right\} \tag{12.50}$$

Theorem 12.8

$$x^* \text{ is local minimizer of } f \text{ in } \Omega \to\ -\nabla f(x^*) \in N_\Omega(x) \tag{12.78}$$

It suggests a close relationship between $N_\Omega(x)$ and the conic combination N of active constraint gradients given by (12.50). When LICQ holds, the two are identical (to within change of sign).

Lemma 12.9

Suppose that the LICQ assumption holds at x^*. Then the normal cone $N_\Omega(x)$ is simply $-N$.

Proof: (from Farkas' Lemma and Definition of $N_\Omega(x)$)

$$g \in N \;\to\; g^T d \geq 0 \quad \forall\, d \in \mathcal{F}(x^*)$$

From Lemma 12.2, $\mathcal{F}(x^*) = T_\Omega(x^*)$, $\quad g \in -N \;\to\; g^T d \leq 0 \quad \forall\, d \in T_\Omega(x^*) \to\ N_\Omega(x) = -N.$

12.8 Lagrange Multiplier and Sensitivity

- Lagrange multiplier λ_i^* indicates the sensitivity of optimal objective value $f(x^*)$ to the presence of the constraint c_i. It shows how hard f is "pushing" or "pulling" the solution x^* against the particular c_i.
- If $\lambda_i^*||\nabla c_i(x^*)||$ is large, the optimal value is sensitive to the placement of the i-th constraint, while if the quantity is small, the dependence is not strong.
- If an inactive constraint $i \notin \mathcal{A}(x)$ such that $c_i(x^*) > 0$ is chosen, and $c_i(x^*)$ is perturbed by a tiny step, it will still be inactive and x^* will still be a local solution of the optimization problem. Since $\lambda_i^* = 0$ from (12.34e), the Lagrange multiplier indicates accurately that constraint i is not significant.

Definition 12.8 – Strongly Active

Let x^* be a solution of the problem (12.1), and suppose KKT condition (12.34) are satisfied.

An inequality constraint $c_i(x^*)$ is strongly active or binding if $i \in \mathcal{A}(x^*)$ and $\lambda_i^* > 0$ for some Lagrange multiplier λ_i^* satisfying KKT condition.

An inequality constraint $c_i(x^*)$ is weakly active if $i \in \mathcal{A}(x^*)$ and $\lambda_i^* = 0$.

Note: Sensitivity analysis is independent of scaling of individual constraints, i.e. $c_i \to \alpha c_i$.

12.9 Duality

Duality theory shows how an alternative problem (called dual) can be constructed from the function an data that define the original optimization problem (called primal), for which the dual is easier to solve computationally than the primal.

Consider a constrained optimization in form of

$$\min_{x\in\mathbb{R}^n} f(x) \quad \text{s.t.} \;\; c(x) \geq 0 \tag{12.81}$$

for which the Lagrange function with Lagrange multiplier $\lambda \in \mathbb{R}^m$ is $L(x,\lambda) = f(x) - \lambda^T c(x)$.

Define the dual objective function $q: \mathbb{R}^n \to \mathbb{R}$ as

$$q(\lambda) \triangleq \inf_{\mathrm{x}} L(x,\lambda)\,, \quad D \triangleq \{\lambda \mid q(\lambda) > -\infty\} \tag{12.82}$$

i.e. the domain of $q(\lambda)$ is the set of λ for which q is finite.

Note that the calculation of infimum in (12.82) requires finding the global minimizer of the function $L(x,\lambda)$. However, given that f and $-c_i$ are convex function with $\lambda \geq 0$ such that $L(x,\lambda)$ is also convex, <u>all local minimizers found are global minimizers</u>.

The dual problem to (12.81) is defined as

$$\max_{\lambda\in\mathbb{R}^n} q(\lambda) \;\; \text{s.t.} \;\; \lambda \geq 0 \tag{12.84}$$

Theorem 12.10

The function q defined by (12.82) is concave, and its domain D is convex.

Proof: For any λ^0 and λ^1 in $\mathbb{R}^m$, any $x \in \mathbb{R}^n$, and any $\alpha \in [0,1]$,

$$L(x,(1-\alpha)\lambda^0 + \alpha\lambda^1) = (1-\alpha)L(x,\lambda^0) + \alpha L(x,\lambda^1\,)$$

Taking infimum of both sides in this expression, using the definition (12.82), and using the results that the infimum of a sum is greater than sum of the infimums, $q\big((1-\alpha)\lambda^0 + \alpha\lambda^1\big) \geq (1-\alpha)q(\lambda^0) + \alpha q(\lambda^1\,)$

It confirms the concavity of q. If both λ^0 and λ^1 belong to D, it also implies $q\big((1-\alpha)\lambda^0+\alpha\lambda^1\big) \geq -\infty$ and hence $(1-\alpha)\lambda^0+\alpha\lambda^1 \in D$, verifying convexity of D.

Theorem 12.11 – Weak Duality

For any $\bar{x}$ feasible for (12.81) and any $\bar{\lambda} \geq 0$, $q(\bar{\lambda}) \geq f(\bar{x})$

Proof: $q(\bar{\lambda}) = \inf_x f(x) - \bar{\lambda}^T c(x) \leq f(\bar{x}) - \bar{\lambda}^T c(x) \leq f(\bar{x})$

Note: The optimal value of the dual problem (12.84) gives a lower bound on the optimal objective value for the primal problem (12.81).

Theorem 12.12 – KKT Condition for the Dual, Solution of Primal is Solution of Dual.

Suppose that $\bar{x}$ is a solution of (12.81) and that f and $-c_i, i=1,2,\dots,m$ are convex functions on $\mathbb{R}^n$ that are differentiable at $\bar{x}$. Then any $\bar{\lambda}$ for which $(\bar{x},\bar{\lambda})$ satisfy the KKT condition (12.87) is the solution of dual (12.84)

$$\nabla f(\bar{x}) - \nabla c(\bar{x})\bar{\lambda} = 0 \tag{12.87a}$$

$$c(\bar{x}) \geq 0 \tag{12.87b}$$

$$\bar{\lambda} \geq 0 \tag{12.87c}$$

$$\bar{\lambda}_i c_i(\bar{x}) = 0, i=1,2,\dots,m \tag{12.87d}$$

where $\nabla c(\bar{x})$ is the $n \times m$ matrix defined by $\nabla c(\bar{x}) = [\nabla c_1(\bar{x})\ \ \nabla c_2(\bar{x}) \dots \nabla c_m(\bar{x})]$.

Note that if the functions are continuously differentiable and a constraint qualification such as LICQ holds at $\bar{x}$, then an optimal Lagrange multiplier is guaranteed to exist, by **Theorem 12.1**.

Theorem 12.13 – Converse of Theorem 12.12, Solution of Dual is Solution of Primal.

Suppose f and $-c_i, i=1,2,\dots,m$ are convex functions and continuously differentiable on $\mathbb{R}^n$. With $\bar{x}$ as a solution of (12.81) at which LICQ holds, $\hat{\lambda}$ as the solution of the dual (12.84) with infimum $\inf_x L(x,\hat{\lambda})$ attained at $\hat{x}$ and $L(x,\hat{\lambda})$ as a convex function, $\bar{x}=\hat{x}$ ($\hat{x}$ is unique solution of (12.81)) and $f(\bar{x}) = L(\hat{x},\hat{\lambda})$

Another form of duality that is convenient for computation is known as Wolfe dual.

$$\max_{x,\lambda} L(x,\lambda) \quad \text{s.t.} \quad \nabla_x L(x,\lambda) = 0, \lambda \geq 0 \tag{12.88}$$

Theorem 12.14 – Solution of the Primal is Solution of Wolfe Dual.

Suppose that f and $-c_i, i=1,2,\dots,m$ are convex functions and continuously differentiable on $\mathbb{R}^n$, and $(\bar{x},\bar{\lambda})$ is a solution pair of (12.81) at which LICQ holds. Then $(\bar{x},\bar{\lambda})$ solves the problem (12.88).

Example 12.11 – Linear Programming

Primal: $$\min c^T x \quad \text{s.t.}\ Ax - b \geq 0 \tag{12.89}$$

Dual: $$q(\lambda) = \inf_x\ [c^T x - \lambda^T(Ax-b)] \rightarrow \max_\lambda b^T\lambda \quad \text{s.t.}\ A^T\lambda = c,\ \ \lambda \geq 0 \tag{12.90}$$

Wolfe Dual: $$\max_\lambda\ c^T x - \lambda^T(Ax-b) \quad \text{s.t.}\ A^T\lambda = c,\ \ \lambda \geq 0$$

Example 12.12 – Convex Quadratic Programming

Primal: $$\min \tfrac{1}{2}x^T G x + c^T x \quad \text{s.t.}\ Ax - b \geq 0 \tag{12.91}$$

Dual: $$q(\lambda) = \inf_x\ [\tfrac{1}{2}x^T G x + c^T x - \lambda^T(Ax-b)] \rightarrow \max_\lambda q(\lambda) = -\tfrac{1}{2}(A^T\lambda - c)^T G^{-1}(A^T\lambda - c) + b^T\lambda \tag{12.93}$$

Wolfe Dual: $$\max_{\lambda,x} -\tfrac{1}{2}x^T G x + \lambda^T b \quad \text{s.t.}\ Gx + c - A^T\lambda,\ \ \lambda \geq 0$$

THE END