Part II

Convex Optimization Models

Chapter 3

Convexity

3.1 Convex Sets

3.1.1 Open and Closed Sets, Interior and Boundary

- **Open**. A set $\mathcal{X} \subseteq \mathbb{R}^n$ is *open* if for any point $x \in \mathcal{X}$, there exist a Euclidean ball centered in x, $B_\epsilon(x)$, which is contained in $\mathcal{X}$.

  $$B_\epsilon(x) \doteq \{z : \|z - x\|_2 < \epsilon\}$$

  Then $\mathcal{X} \subseteq \mathbb{R}^n$ is open if

  $$\forall x \in \mathcal{X}, \exists \epsilon > 0 : B_\epsilon(x) \subset \mathcal{X}$$

  That is, if a set is open, then we can always find a Euclidean ball with radius ϵ centered at $x \in \mathcal{X}$ that is still in the set $\mathcal{X}$.

- **Closed**. A set $\mathcal{X} \subseteq \mathbb{R}^n$ is said to be *closed* if its complement $\mathbb{R}^n \backslash \mathcal{X}$ is *open*. A closed set can be defined as a set which ***contains all its limit points***. In a complete metric space, a closed set is a set which is closed under the limit operation.

- The whole space $\mathbb{R}^n$ and the empty set $\emptyset$ are declared *open* by definition and they are also *closed* by definition.

- **Interior**. The *interior* of a set $\mathcal{X} \subseteq \mathbb{R}^n$ is defined as

$$\text{int}\,\mathcal{X} = \{z \in \mathcal{X} : B_\epsilon(z) \subseteq \mathcal{X},\ \text{for some } \epsilon > 0\}$$

  The interior of a subset S consists of all points of S that ***do not belong to the boundary*** of S. That is, any point in the set that is able to construct a Euclidean ball in the set.

- **Closure**. The *closure* of a set $\mathcal{X} \subseteq \mathbb{R}^n$ is defined as

$$\bar{\mathcal{X}} = \{z \in \mathbb{R}^n : z = \lim_{k\to\infty} x_k,\ x_k \in \mathcal{X}, \forall k\}$$

  The closure of $\mathcal{X}$ is the set of limits of sequences in $\mathcal{X}$. The sequence converges to an element in the closed set.

- In other word, the closure of $\mathcal{X}$ consists of ***all points in*** S ***together with all limit points of*** S. It may equivalently be defined as the union of S and its *boundary*, and also as the intersection of all closed sets containing S. Intuitively, the closure can be thought of as all the points that are either in S or "near" S. The notion of closure is in many ways *dual* to the notion of interior.

- **Boundary**. The *boundary* of $\mathcal{X}$ is defined as

$$\partial\mathcal{X} = \bar{\mathcal{X}} \backslash \text{int}\,\mathcal{X}$$

  where int denotes *interior* (i.e. closure = interior + boundary).

- **Bounded**. A set is said to be *bounded* if it is contained in a ball of finite radius, meaning that every element is at most a finite distance away from each other.

- **Compact**. If $\mathcal{X} \subseteq \mathbb{R}^n$ is *closed* and *bounded*, then it is said to be *compact*.

3.1.2 Combinations and Hulls

- **Linear hull (subspace)**. Given a set of points (vectors), $\mathcal{P} = \{x_1, \cdots, x_m\}$, the *linear hull (subspace)* generated by these points is the set of all possible *linear combinations* of the points.

$$x = \lambda_1 x_1 + \cdots + \lambda_m x_m,\ \lambda_i \in \mathbb{R},\ i = 1, \cdots, m$$

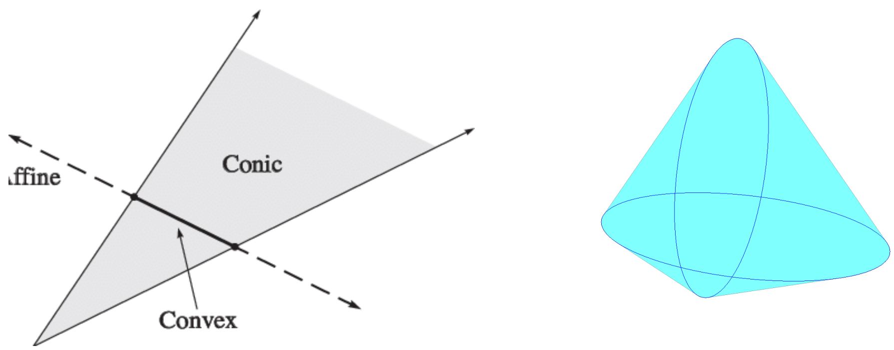

Figure 3.1: Left: A conic hull in $\mathbb{R}^2$, an affine hull in $\mathbb{R}^1$ and a convex hull in $\mathbb{R}^1$. Convex hull is a subset of affine hull. Right: The convex hull of the union of two ellipses.

- **Affine hull**. The *affine hull*, aff $\mathcal{P}$ is the set generated by taking all possible linear combinations of the points in $\mathcal{P}$, under the restriction that the coefficients λ_i ***sum up to 1***.

- **Convex combination**. This is a special type of linear combination $x = \lambda_1 x_1 + \cdots + \lambda_m x_m$ in which the coefficients λ_i are restricted to be ***non-negative and to sum up to 1***. Intuitively, this is a weighted average of the points.

- **Convex hull**. The set of all possible convex combination of the point set

$$\text{co}(x_1, \cdots, x_m) = \left\{ x = \sum_{i=1}^{m} \lambda_i x_i : \lambda_i \geq 0,\ i = 1, \cdots, m; \sum_{i=1}^{m} \lambda_i = 1 \right\}$$

- Convex hull is a subset of affine hull. We can think of convex hull as it is contained in the affine hull.

- **Conic combination**. Linear combination where the coefficients are restricted to be ***non-negative***.

- **Conic hull**. The set of all possible conic combination of the point set

$$\text{conic}(x_1, \cdots, x_m) = \left\{ x = \sum_{i=1}^{m} \lambda_i x_i : \lambda_i \geq 0,\ i = 1, \cdots, m \right\}$$

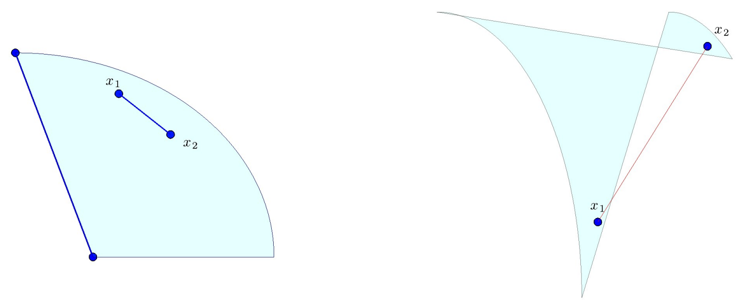

Figure 3.2: Convex and non-convex set

3.1.3 Convex Sets

- **Convex**. A subset $C \subseteq \mathbb{R}^n$ is *convex* if it contains the line segment between 2 points in it.

$$x_1, x_2 \in C,\ \lambda \in [0,1] \Rightarrow \lambda x_1 + (1-\lambda)x_2 \in C$$

- **Dimension of convex set**. The *dimension* d of a convex set $C \subseteq \mathbb{R}^n$ is defined as the dimension of its **affine hull**. Notices that it ccan happen that $d < n$.

- Subspace and affine sets, such as lines, planes, and higher-dimensional "flat" sets are convex, as they contain the entire line passing through any 2 points, not just the line segment. Half-spaces are also convex.

- **Cone**. A set C is a *cone* if it has the property that if $x \in C$, then $\alpha x \in C$, for every $\alpha \geq 0$.

- **Convex cone**. A set C is a *convex cone* if it is convex and it is a cone. The conic hull of a set is a convex cone.

3.1.4 Operations that Preserve Convexity

- **Intersection**. The intersection of a (possibly infinite) family of convex sets is convex. This property can be used to prove convexity for a wide variety of situations.

- **Affine transformation**. If a map $f : \mathbb{R}^n \to \mathbb{R}^m$ is affine, and $C \subset \mathbb{R}^n$ is convex, then the image set $f(c)$ is convex.

$$f(C) = \{f(x) :\in C\}$$

3.1.5 Supporting and Separating Hyperplanes

Theorem 3 (Supporting hyperplane theorem). *If $C \subseteq \mathbb{R}^n$ is convex and $z \in \partial C$, then there exists a supporting hyperplane for C at a boundary point z.*

Theorem 4 (Separating hyperplane theorem). *Let C_1, C_2 be convex subsets of $\mathbb{R}^n$ that do not intersect. Then there exists a separating hyperplane $\mathcal{H}$ for C_1, C_2. Furthermore, if C_1 is **closed** and **bounded** and C_2 is **closed**, then C_1, C_2 can be strictly separated.*

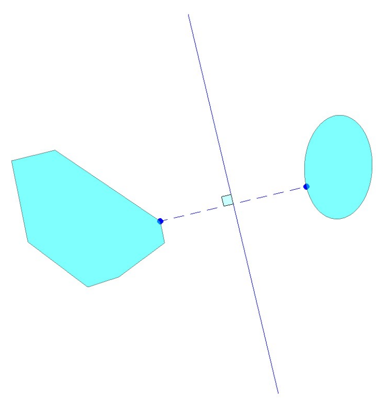

Figure 3.3: Seperating hyperplane

3.2 Convex Functions

3.2.1 Definitions

- **Effective domain**. The *effective domain (domain)* of a function $f : \mathbb{R}^n \to \mathbb{R}$ is the set over which the function is well defined.

$$\text{dom} f = \{x \in \mathbb{R}^n : -\infty < f(x) < \infty\}$$

- **Convexity**. A function $f : \mathbb{R}^n \to \mathbb{R}$ is *convex* if ***dom**f **is a convex set*** and for all $x, y \in \text{dom} f$ and all $\lambda \in [0, 1]$ it holds that

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$$

- **Concave**. A function is *concave* if $-f$ is *convex*.
- **Strictly convex**. A function f is *strictly convex* if inequality above holds strictly (i.e., with < instead of $\leq$).
- **Epigraph**. Given a function $f : \mathbb{R}^n \to (-\infty, +\infty]$, its epigraph (the set of points lying above the graph of the function) is the set

$$\text{epi}\, f = \{(x, t), x \in \text{dom} f, t \in \mathbb{R} : f(x) \leq t\}$$

- If f is a convex function if and only if epi f is a convex set.
- **Sublevel set**. The α-sublevel set of f is defined as

$$S_\alpha = \{x \in \mathbb{R}^n : f(x) \leq \alpha\}$$

- If f is a convex function, then S_α is a convex set.
- **Closed function**. A function $f \to (-\infty, \infty]$ is *closed* if its epigraph is a closed set. This is equivalent to that every sublevel set S_α of $f, \alpha \in \mathbb{R}$ is closed.
- **Sum of convex functions**. If $f_i : \mathbb{R}^n \to R, i = 1, \cdots, m$ are convex functions, then the function

$$f(x) = \sum_{i=1}^{m} \alpha_i f_i(x),\ \alpha_i \geq 0, i = 1, \cdots, m$$

is also convex over $\cap_i$ dom f_i.

- **Affine variable transformation**. Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex, and define

$$g(x) = f(Ax + b), \quad A \in \mathbb{R}^{n,m},\ b \in \mathbb{R}^n$$

  Then g is convex over dom $g = \{x : Ax + b \in \text{dom} f\}$.

- **Example**. $f(z) = -\log(x)$, is convex over dom $f = \mathbb{R}_{++}$, hence $f(x) = -\log(ax + b)$ is also convex over $ax + b > 0$.

3.2.2 Alternative Characterizations of Convexity

Besides the definition, there are other rules or conditions that can characterize convexity of a function. From now on, when mentioning convexity of a function, it is implicitly assumed that dom f is convex.

First-order Conditions

- If f is differentiable (i.e., dom f is open and the gradient exists everywhere on the domain), then f is convex if and only if

$$\forall x, y \in \text{dom } f,\ f(y) \geq f(x) + \nabla f(x)^T (y - x)$$

- The geometric interpretation is that the graph of f is bounded below everywhere by any one of its tangent hyperplanes, or that any tagent hyperplane is a supporting hyperplane for epi f.

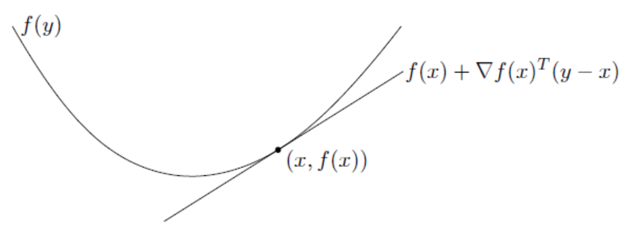

Figure 3.4

- The gradient of a convex function at a point $x \in \mathbb{R}^n$ divides the whole space in two half-spaces.

Second-order Conditions

- If f is twice differentiable, then f is convex if and only if its Hessian matrix $\nabla^2 f$ is ***positive semidefinite*** everywhere on the (open) domain of f, that is if and only if $\nabla^2 f \succeq 0$ for all $x \in \text{dom } f$.

Restriction to a Line

- A function f is convex if and only if its restriction to *any* line is convex. Restriction to a line means that for every $x_0 \in \mathbb{R}^n$ and $v \in \mathbb{R}^n$, the function of scalar variable t

$$g(t) = f(x_0 + tv)$$

is convex.

3.2.3 Operations that Preserve Convexity

- **Composition with an affine function**. The composition with an affine function preserves convexity. If $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m, f : \mathbb{R}^m \to \mathbb{R}$ is convex, then the function $g : \mathbb{R}^m \to \mathbb{R}, g(x) = f(Ax + b)$ is convex.

- **Pointwise supremum or maximum** The *pointwise maximum* of a family of convex functions is convex. If $(f_\alpha)_{\alpha \in \mathcal{A}}$ is a family of convex functions indexed by the parameter α, and $\mathcal{A}$ is an compact set for α, then the pointwise maximum function

$$f(x) = \max_{\alpha \in \mathcal{A}} f_\alpha(x)$$

is convex.

- **Example**. The *dual norm*

$$f(x) = \|x\|^* = \max_{y : \|y\| \leq 1} y^T x$$

is convex, since it is defined as the maximum of convex (in fact, linear) functions, indexed by the vector y.

- **Example**. The *largest sigular value* of a matrix

$$f(X) = \sigma_{\max}(X) = \max_{v:\|v\|_2=1} \|Xv\|_2$$

  is convex, since it is the pointwise maximum of convex functions which are the composition of the Euclidean norm (convex function) with the affine function $X \to Xv$.

- **Non-negative weighted sum**. The nonnegative weighted sum of convex functions is convex.

- **Partial minimum**. If f is a convex function in (x, z) (jointly convex in the variables x and z) then the function

$$g(x) = \min_z f(x, z)$$

  is convex.

- **Composition with monotone convex functions**. If $f = h \circ g$, with h, g convex and h *non-decreasing*, then f is convex. The condition $f(x) \leq z$ corresponds to $h(g(x)) \leq z$, which is equivalent to the existence of y such that

$$h(y) \leq z, g(x) \leq y$$

- The above rules have direct extensions ot functions of a vector argument. If the component functions $g_i : \mathbb{R}^n \to \mathbb{R}, i = 1, \cdot, k$, are convex and $h : \mathbb{R}^k \to R$ is convex and non-decreasing with dom $g_i =$ dom $h = \mathbb{R}$, then

$$x \to (h \circ g)(x) \doteq h(g_1(x), \cdots, g_k(x))$$

  is convex.

- **Example**. If g_i are convex, then $\log \sum_i e^{g_i(x)}$ is also convex.

3.2.4 Subgradients and Subdifferentials

The characterization of a convex differentiable function

$$\forall x, y \in \text{dom}\, f, f(y) \geq f(x) + \nabla f(x)^T (y - x)$$

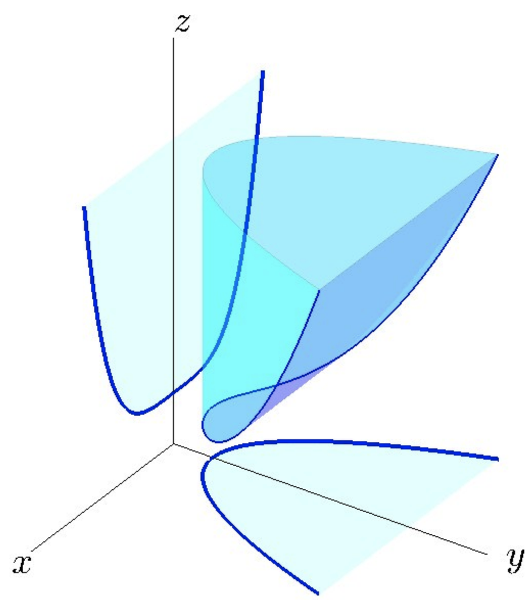

Figure 3.5: The convex set $\{(x, y, z) : h(y) \leq z, g(x) \leq y\}$ and its projection on the space of (x, z)-variables is the epigraph of f. The epigraph of g is the projection of (x, y)-variables.

states that at any point $x \in \text{dom} f$, the function $f(x)$ is lower bounded by an affine function of y, and that the bound is exact at x.

$$f(y) \geq f(x) + g_x^T(y - x), \forall y \in \text{dom} f \tag{3.1}$$

where $g_x = \nabla f(x)$. Even when f is non-differentiable (gradient may not exist at some points), the relation may still hold for suitable vectors g_x.

- **Subgraident**. If $x \in \text{dom} f$ and (3.1) holds for some vectors $g_x \in \mathbb{R}^n$, then g_x is called a *subgradient* of f at x.
- **Subdifferential**. The set of all subgradients of f at x is called the *subdifferential*, denoted by $\partial f(x)$.
- A subgradient is a ***surrogate*** of the gradient. It coincides with the gradient whenever a gradient exists, and it generalizes gradient at points where f is non-differentiable.

Theorem 5. *For a convex f, a subgradent always exists at all points in the relative interior of the domain. Moreover, f is **directionally** differentiable at all such points. For a convex function f,*

$$\begin{aligned} f(y) \geq f(x) + g_x^T(y-x) \\ \forall y \in \text{dom} f \\ \forall g_x \in \partial f(x) \\ \forall x \in \text{relint dom} f \end{aligned}$$

1. The subdifferential $\partial f(x)$ is a closed, convex, nonempty and bounded set.
2. If f is differentiable at x, the $\partial f(x)$ contains only one element: the gradient of f at x, that is, $\partial f(x) = \{\nabla f(x)\}$.
3. For any $v \in \mathbb{R}^n$

$$f'_v(x) \doteq \lim_{t \to 0^+} \frac{f(x+tv) - f(x)}{t} = \max_{g \in \partial f(x)} v^T g$$

   where $f'_v(x)$ is the ***directional derivative*** of f at x along the direction v.

- **Example**. The absolute value function $f(x) = |x|$, $x \in \mathbb{R}$ has subgradients $g \in [-1, 1]$ at 0.

$$\partial |x| = \begin{cases} \text{sgn}(x) & \text{if } x \neq 0, \\ [-1, 1] & \text{if } x = 0 \end{cases}$$

3.3 Convex Problems

3.3.1 Definition

- **Optimization problem**. An *optimization problem* of the form

$$p^* = \min_{x \in \mathbb{R}^n} \quad f_0(x) \tag{3.2}$$

$$\text{subject to:} \quad f_i(x) \leq 0, \quad i = 1, \cdots, m \tag{3.3}$$

$$h_i(x) = 0, \quad i = 1, \cdots, q \tag{3.4}$$

  is called a convex optimization problem, if

1. the object function f_0 is convex;
2. the functions defining the inequality constraints, f_i, are convex;
3. the functions defining the equality constraints, h_i, are affine.

- The problem can equivalently be defined as

$$p^* = \min_{x \in \mathcal{X}} f_0(x) \tag{3.5}$$

where the decision variable x must belong to a convex set $\mathcal{X}$. When $x \in \mathbb{R}^n$, then the problem is *unconstrained*.

- **Feasible set**. The *feasible set* of the problem is the set of points x that satisfy the constraints. If $\mathcal{X}$ is empty, then the problem is *infeasible*. In such a case, it is customary to set $p^* = +\infty$.

- **Optimal set**. The *optimal set* is defined as the set of feasible points for which the objective funtion attains the optimal value:

$$\begin{aligned}\mathcal{X}_{\text{opt}} &= \{x \in \mathcal{X} : f_0(x) = p^*\} \\ &= \arg\min_{x \in \mathcal{X}} f_0(x)\end{aligned}$$

- **Unbounded below**. If the problem is feasible and $p^* = -\infty$, we say that the problem is *unbounded below*.

- **Attained**. If the problem is feasible but still no optimal solution exists (e.g. the solution only exists in the limit), we say that the optimal value p^* is not *attained*. at any finite point.

- **Feasibility problem**. If we are only interested in verifying if the problem is feasible or not, then the problem is called a *feasibility problem*.

find $x \in \mathcal{X}$ or prove that $\mathcal{X}$ is empty.

3.3.2 Local and Global Optima

- **Local optima**. A point z is *local optimum* if there exists a ball B_r of radius $r > 0$ centered at z such that z minimizes f_0 locally in the ball B_r.

$$\min_{x \in \mathcal{X}} f_0(x) \text{ subject to:} \|x - z\|_2 < r$$

where $\forall x \in B_r \cap \mathcal{X}$, $f_0(x) \geq f_0(z)$.

- **Global optima**. If z is a global optimum point, then it holds instead that $f_0(x) \geq f_0(z), \forall x \in \mathcal{X}$.

Theorem 6. *Consider the optimization problem*

$$\min_{x \in \mathcal{X}} f_0(x)$$

If $f_0(x)$ is a convex function and $\mathcal{X}$ is a convex set, then any locally optimal solution is also globally optimal. Moreover, the set $\mathcal{X}_{opt}$ of optimal points is convex.

3.3.3 Uniqueness of the Optimal Solution

For any convex optimization problem, any locally optimal solution is also globally optimal, but this does not mean, in general, that the optimal solution is ***unique***. Intuitively, such a lack of uniqueness is in the case due to the **flatness** of the objective function around the optimal points.

Theorem 7. *If f_0 in the optimization problem*

$$p^* = \min_{x \in \mathcal{X}} f_0(x)$$

*is a **strictly convex function**, $\mathcal{X}$ is a convex set, and x^* is an optimal solution to the problem, then x^* is the unique optimal solution, that is $\mathcal{X}_{opt} = x^*$.*

Theorem 8. *Let f_0 is a **non-constant linear function** ($f_0 = c^T x, c \neq 0$), and $\mathcal{X}$ is closed, full-dimensional, and strictly convex. If the problem admits an optimal solution x^*, then this solution is unique.*

- **Alternative condition for uniqueness**. This states that another sufficient condition for uniqueness of the optimal solution is the class of convex programs with ***linear objective function*** and ***strictly convex feasible set***.

3.3.4 Problem Transformation

An optimization problem can be transformed, or reformulated, into an *equivalent* one by means of several useful "tricks".

Monotone Objective Transformation

- Consider an optimization problem of the form (3.5). Let $\varphi : \mathbb{R} \to \mathbb{R}$ be a continuous and strictly increasing function over $\mathcal{X}$, and consider the transoformed problem

$$
\begin{aligned}
g^* = \min_{x \in \mathbb{R}^n} \quad & \varphi(f_0(x)) && (3.6)\\
\text{subject to:} \quad & f_i(x) \leq 0, \quad i = 1, \cdots, m && (3.7)\\
& h_i(x) = 0, \quad i = 1, \cdots, q && (3.8)
\end{aligned}
$$

- Problems (3.5) and (3.8) have the same feasible set and the same set of optimal solutions.
- A common convexity-preserving objective transformation consists of **squaring** a (non-negative) objective.

Monotone Constraint Transformation

- If a constraint in a problem can be expressed as $l(x) \leq r(x)$, and φ is a continuous and ***strictly increasing*** function over $\mathcal{X}$, then this constraint is equivalent to

$$\varphi(l(x)) \leq \varphi(r(x))$$

- If φ is continuous and ***strictly decreasing*** over $\mathcal{X}$, then the constraints is equivalent to

$$\varphi(l(x)) \geq \varphi(r(x))$$

Change of Variables

If $F : X \to Y$ is an invertible mapping (i.e. $\forall y \in Y$ there exist a unique $x \in \mathcal{X}$ such that $F(x) = y, F^{-1}(y) = x$), describing a change of variables where the set X includes the intersection of the domain of f_0 with the feasible set $\mathcal{X}$ of the problem.

Then problem (3.4) can be reformulated as

$$\begin{aligned} p^* = \min_{y \in \mathbb{R}^n} \quad & g_0(y) \\ \text{subject to:} \quad & g_i(y) \leq 0, \quad i = 1, \cdots, m \\ & s_i(y) = 0, \quad i, = 1, \cdots, q \end{aligned}$$

where $g_i(y) = f_i(F^{-1}(y))$ and $s_i(y) = h_i(F^{-1}(y))$.

Addition of Slack Variables

Bringing in new *slack* variables into the problem is a equivalent to the original problem. Consider the problem with the objective involves the **sum** of terms,

$$\begin{aligned} p^* = \min_{x} \quad & \sum_{i=1}^{r} \varphi_i(x) \\ \text{s.t.:} \quad & x \in \mathcal{X} \end{aligned}$$

Introducing slack variables t_i, we reformulate the problem as

$$\begin{aligned} g^* = \min_{x,t} \quad & \sum_{i=1}^{r} t_i \\ \text{s.t.:} \quad & x \in \mathcal{X} \\ & \varphi_i(x) \leq t_i, \quad i = 1, \cdots, r \end{aligned}$$

where the new problem has the original variable x, plus the vector of slack variables $t = (t_1, \cdots, t_r)$.

Other Transformations

- Substituting equality constraints with inequality constraints.
- Elimination of inactive constraints.

3.3.5 Optimality Conditions

- Consider the optimization problem $\min_{x \in \mathcal{X}} f_0(x)$, where f_0 is convex and differentiable, and $\mathcal{X}$ is convex. Then,

$$x \in \mathcal{X} \text{ is optimal} \Leftrightarrow \nabla f_0(x)^T(y-x) \geq 0, \; \forall y \in \mathcal{X}$$

- If $\nabla f_0(x) \neq 0$, then $\nabla f_0(x)$ is a normal direction defining a *supporting hyperplane* $\{y : \nabla f_0(x)^T(y-x) = 0\}$.

- When the problem is **unconstrained** (i.e. $\mathcal{X}$ in $\mathbb{R}^n$), then the optimality condition becomes $\nabla f_0(x) = 0$.