Optimization and Machine Learning M - Theorems and Definitions

Dante Piotto

spring semester 2024

Contents

Chapter 1

Non Linear Programming

1.1 Unconstrained Optimization

The problem to be solved is defined as:

$$\min f(x) \quad x \in \mathbb{R}^n$$

1.1.1 Necessary conditions

Definition 1.1 (descendant direction)
a vector $d \in \mathbb{R}^n$ is a *descendant direction* for function f in x if $\exists \delta > 0 : f(x + \alpha d) < f(x) \quad \forall \alpha \in (0, \delta)$. We denote with $D(x)$ the set of all descendant directions for f in x

Definition 1.2 (stationary point)
A point $x \in \mathbb{R}^n$ is a *stationary point* for f if $\nabla f(x) = 0$

Theorem 1.1 (Firs-Order Necessary Condition)
Let $f \in C^1$. If $\bar{x} \in \mathbb{R}^n$ is a local minimum for problem (1.1), then $\nabla f(\bar{x}) = 0$

Proof. Let $\bar{x} \in \mathbb{R}^n$ be a local minimum for problem (1.1). The proof is by contradiction, thus assume that $\nabla f(\bar{x}) \neq 0$. Define a direction $d^* = -\dfrac{\nabla f(\bar{x})}{\|\nabla f(\bar{x})\|_2}$ and a point $y = \bar{x} + \alpha d^*$, for some $\alpha > 0$. It follows that $y \neq \bar{x}$ for any value of $\alpha > 0$.

For a sufficiently small value of α, one can approximate function f in y according to the Taylor series up to the first order as follows:

$$f(y) = f(\bar{x}) + \nabla f(\bar{x})^T (y - \bar{x}) + R_1(\bar{x}, \alpha) = f(\bar{x}) - \alpha \|\nabla f(\bar{x})\|_2 + R_1(\bar{x}, \alpha)$$

with $\lim_{\alpha \to 0} \dfrac{R_1(\bar{x}, \alpha)}{\alpha} \to 0$

Thus, for a sufficiently small value of α, the associated point y is such that $f(y) < f(\bar{x})$, giving a contradiction with the hypothesis that $\bar{x}$ is a local minimum □

Theorem 1.2 (Second-Order Necessary Condition)
Let $f \in C^2$ if $\bar{x} \in \mathbb{R}^n$ is a local minimum for problem (1.1), then

1. $\nabla f(\bar{x}) = 0$

2. $d^T \nabla^2 f(\bar{x}) d \geq 0$

Proof. The first condition has alread been proved in the previous theorem

We now prove condition 2 by contradiction, and assume this condition is not satisfied by a local minimum $\bar{x} \in \mathbb{R}^n$. Thus, assume that $\nabla^2 f(\bar{x})$ is not positive semidifinite.

Since 2 is not satisfied, it is possible to find a vector $d^* \in \mathbb{R}^n$ such that ${d^*}^T \nabla^2 f(\bar{x}) d^* < 0$. Note that $d^* \neq 0$. For the sake of simplicity, assume that d^* has been normalized so as to have $\|d^*\| = 1$. Define a new point $y = \bar{x} + \alpha d^*$ for some scalar α, and note that $y \neq \bar{x}$ for all $\alpha > 0$

For a sufficiently small value of α, one can approximate function f in y according to the Taylor series up to the second order as follows:

$$f(y) = f(\bar{x}) + \nabla f(\bar{x})(y - \bar{x}) + \frac{1}{2}(y - \bar{x})^T \nabla^2 f(\bar{x})(y - \bar{x}) + R_2(\bar{x}, \alpha)$$

with $\lim_{\alpha \to 0} \dfrac{R_2(\bar{x}, \alpha)}{\alpha} = 0$

As condition 1 states that $\nabla f(\bar{x}) = 0$, and

$$(y - \bar{x})^T \nabla^2 f(\bar{x})(y - \bar{x}) = (\alpha d^*)^T \nabla^2 f(\bar{x})(\alpha d^*) = \alpha^2 {d^*}^T \nabla^2 f(\bar{x}) d^* < 0$$

then we get

$$f(y) = f(\bar{x}) + \frac{1}{2}\alpha^2 {d^*}^T \nabla^2 f(\bar{x}) d^* < f(\bar{x})$$

This implies that for any sufficiently small value of α there exists a point y for which $f(y) < f(\bar{x})$, which contradicts the hypothesis that $\bar{x}$ is a local minimum. □

Theorem 1.3 (Second-Order Sufficient Condition)
Let $f \in C^2$. A solution $\bar{x} \in \mathbb{R}^n$ that satisfies the following conditions:

1. $\nabla f(\bar{x}) = 0$
2. $\nabla^2 f(\bar{x})$ is positive definite

is a (strict) local minimum for problem (1.1)

Proof. Let $\bar{x} \in \mathbb{R}^n$ be a solution that satisfies conditions 1 and 2. Let $\rho > 0$ and define a neighbourhood of $\bar{x}$ with radius ρ as follows:

$$N(\bar{x}, \rho) = \{y \in \mathbb{R}^n : \|y - \bar{x}\| \le \rho\}$$

Let $y \in N(\bar{x})$ be a point in this neighbourhood that is distinct from $\bar{x}$, i.e., defined by some $d \in \mathbb{R}^n$ with $\|d\| = 1$ and some $\alpha > 0$. The Taylor series for function f in y up to the second order is:

$$f(y) = f(\bar{x} + \alpha d) = f(\bar{x}) + \nabla f(\bar{x})^T \alpha d + \frac{1}{2}(\alpha d)^T \nabla^2 f(\bar{x})(\alpha d) + R_2(\bar{x}, \alpha) = f(\bar{x}) + \frac{1}{2}\alpha^2 d^T \nabla^2 f(\bar{x}) d + R_2(\bar{x}, \alpha)$$

where the last equality derives from condition 1.

For a sufficiently small value of α, the last term $R_2(\bar{x}, \alpha)$ is negligible. Thus, recalling the properties of positive definite matrices we have

$$f(y) \ge f(\bar{x}) + \frac{1}{2}\alpha^2 \lambda_{min}$$

where λ_{min} is the smallest eigenvalue of matrix $\nabla^2 f(\bar{x})$. As this is a positive definite matrix, we have $\lambda_{min} > 0$. This implies that $f(y) > f(\bar{x})$ for sufficiently small $\alpha > 0$ □

1.2 Algorithms for unconstrained optimization

Iterative schemes:

$$x^{k+1} = x^k + \alpha_k d^k$$

- $d^k \in \mathbb{R}^n, \|d^k\| = 1$ search direction
- $\alpha_k \in \mathbb{R}_+$ step size

1.2.1 Line Search Algorithms

1. if x^k is optimal stop
2. determine a descendent direction d^k for the objective function
3. determine the step size α_k along direction d^k starting from x^k
4. define the nuew solution $x^{k+1} = \alpha_k d^k$ and iterate

Determining the search direction

Typically

$$d^k = -D^k \nabla f(x^k)^T$$

where D^k is symmetric and nonsingular. Whenever D^k is positive definite, d^k is a descendant direction.

The gradient method

based on the approximation of the objective function f according to the Taylor series up to the first order

$$f(x^k + \alpha d) = f(x^k) + \alpha \nabla f(x^k)^T d$$

considering this expression as a function of d we get a minimum for

$$d^k = -\frac{\nabla f(x^k)}{\|\nabla f(x^k)\|}$$

Newton's method

second order Taylor approximation:

$$f(x^k + h) = f(x^k) + \nabla f(x^k)^T h + \frac{1}{2} h^T \nabla^2 f(x^k) h$$

setting to zero the gradient wrt h:

$$h = -\nabla^2 f(x^k)^{-1} \nabla f(x^k)$$

so the algorithm takes:

$$d^k = -\frac{\nabla^2 f(x^k)^{-1} \nabla f(x^k)}{\|\nabla^2 f(x^k)^{-1} \nabla f(x^k)\|} \quad \text{and} \quad \alpha^k = \|\nabla^2 f(x^k)^{-1} \nabla f(x^k)\|$$

Modified Newton's method

Performance of Newton's method can be improved by calculating step size according to a line search algorithm

Quasi-Newton's method

To reduce computational effort one can compute the search direction as

$$d^k = -\bar{B}^{-1} \nabla f(x^k)$$

where matrix $\bar{B}$ is some approximation of the current Hessian matrix. In particular, we can write

$$\nabla f(x^{k+1}) \simeq \nabla f(x^k) + \nabla^2 f(x)(x^{k+1} - x^k)$$

hence

$$\bar{B}(x^{k+1} - x^k) \simeq \nabla f(x^{k+1}) - \nabla f(x^k)$$

Step size selection

Let

$$\phi^k : \mathbb{R}_+ \to \mathbb{R}, \alpha \to \phi^k(\alpha) = f(x^k + \alpha d^k)$$

The "best step size" for iteration k is

$$\alpha^k = \arg\min_{\alpha \geq 0} \phi(\alpha)$$

but can be computationally expensive

$$\phi'(\alpha) = \nabla f(x^k + \alpha d^k)^T d^k = 0$$

id $d^k = -\nabla f(x^k)$ the best step size policy gives $\nabla f(x^{k+1})^T \nabla f(x^k) = 0$, i.e. for every pair of consecutive iterations the gradients of function f are orthogonal to each other. Can produce fluctuations in the resulting objective function value by producing a new point that is quite far from the previous one.

limited best step size

A common choice is to impose a maximum value for the distance between consecutive points:

$$\alpha^k = \arg \min_{0 \le \alpha \le \bar{\alpha}} \phi(\alpha)$$

Constant step size

faster

Wolfe Conditions

Allow to determine an approximate solution for the problem; typically good performance in convergence and computing time

Conditions require that α be such that:

$$f(x^k + \alpha d^k) \le f(x^k) + c_1 \alpha \nabla f(x^k)^T d^k$$
$$\nabla f(x^k + \alpha d^k)^T d^k \ge c_2 \nabla f(x^k)^T d^k$$

where $0 < c_1 < c_2 < 1$ are two parameters of the algorithm. The first condition is know as *Armijo condition* and can be rewritten as

$$f(x^k) - f(x^k + \alpha d^k) \ge -c_1 \alpha \nabla f(x^k)^T d^k$$

This condition ensures that α is improving wrt $\alpha = 0$ for function $\phi(\alpha)$, with a value reduction taht is proportional to α and to $\phi'(0) = \nabla f(x^k)^T d^k$. This condition does not ensure convergence, hence the second condition, known as *curvature condition*. We can rewrite it as

$$\phi'(\alpha) \ge c_2 \phi'(0)$$

When the condition is satisfied, it signifies that we cannot expect much decrease of the objective function by increasing α, and it is not satisfied for small values of α.

Algorithm:

1. set $i = 0$ and determine an initial value $\alpha(0)$
2. compute $f(x^k + \alpha(i) d^k)$
3. if $f(x^k + \alpha(i) d^k) > f(x^k) + c_1 \alpha(i) \nabla f(x^k)^T d^k$ set $\alpha(i+1) = \alpha(i)/2, i = i + 1$ and goto step 2
4. if $\nabla f(x^k + \alpha(i) d^k)^T d^k < c_2 \nabla f(x^k)^T d^k$ set $\alpha(i+1) = 2\alpha(i), i = i + 1$ and goto step 2
5. set $\alpha_k = \alpha_i$ and return

Typically, the value for c_1 is very small (e.g. $c_1 = 10^{-4}$), while c_2 is considerably larger (e.g. $c_2 = 0.9$). It can be proven that, beside pathological conditions, the Wolfe conditions define at least one interval $[\alpha_1, \alpha_2]$ that includes candidate values for the next step size.

1.2.2 Trust-region algorithms

A region T in which an approximation $\tilde{f}$ of the cost function is considered to be valid. The search direction is given by

$$p^k = \arg\min\{\tilde{f}(x^k + p) : x^k + p \in T\}$$

Typically the trust region is defined by all points within a distance of x^k and the approximating function $\tilde{f}$ is given by the Taylor series up to the second order. For this choice the determination of p^k requires optimizing a quadratic function over a convex set.

In practical algorithms the region size is chosen according to the performance of the algorithm during previous iterations: The size of the trust region is updated according to the ratio

$$r_k = \frac{f(x^k) - f(x^k + p)}{\tilde{f}(x^k) - \tilde{f}(x^k + p)}$$

Values close to 1 indicate that the model is consistently reliable and the trust region may be increased, whereas if r_k is small the model is an inadequate representation of the objective function over the current trast region which should be reduced in size.

It can be proved that, if the x^k points generated belong to a bounded set, then there exists a limit point of the sequence that satisfies the second order necessary conditions.

1.3 Constrained Optimization

Theorem 1.4 (Gordan's Theorem)
Let A be an $m \times n$ matrix. The system $Ax < 0$ has no solution iff there exists a $y \in \mathbb{R}^m, y \geq 0, y \neq 0$ such that $A^T y = 0$

Proof. Given the $m \times n$ matrix A, define the following problems:

$$P_1: \text{ is there an } x \in \mathbb{R}^n \text{ such that } Ax < 0?$$
$$P_1: \text{ is there a } y \in \mathbb{R}^m \text{ such that } y \geq 0, y \neq 0 \text{ and } A^T y = 0?$$

Observe that it cannot happen that both problems have answer "yes". Assume indeed that there exists both an $x \in \mathbb{R}^n$ such that $Ax < 0$ and a $y \in \mathbb{R}^m$ such that $y \geq 0, y \neq 0$ and $A^T y = 0$. We have $0 = 0^T x = (A^T y)^T x = (y^T A)x = y^T(Ax) = y^T z < 0$, where we introduced $z = Ax < 0$, and the last inequality derives from $z < 0, y \geq 0$ and $y \neq 0$

Now assume that problem P_1 has answer "no" and define the following sets:

$$S_1 = \{z \in \mathbb{R}^m : z < 0\} \quad \text{and} \quad S_2 = \{z \in \mathbb{R}^m : z = Ax \text{ for some } x \in \mathbb{R}^n\}$$

As $S_1 \cap S_2 = \emptyset$ there should exist an hyperplane, associated with a vector $y \in \mathbb{R}^m$, that separates S_1 and S_2, i.e. such that

$$y^T z < 0 \quad \forall z \in S_1 \quad \text{and} \quad y^T z \geq 0 \quad \forall z \in S_2$$

Vector y must satisfy $A^T y = 0$; indeed, if $A^T y \neq 0$, one could define $\bar{x} = -(y^T A)^T = -A^T y$, such that $\bar{x} \neq 0$. Imposing $y^T Ax \geq 0$ for $x = \bar{x}$ we should have $0 \leq (y^T A)\bar{x} = (-\bar{x}^T)\bar{x}$, while this is impossible as $\|\bar{x}\| > 0$

Furthermore, by definition y satisfies $y^T z < 0 \quad \forall z \in S_1$. In order to check possible y vectors that satisfy these conditions, let's impose this condition for different z vectors. In particular, we consider m distinct vectors $\tilde{z}_j \in \mathbb{R}^m$, one for each $j = 1, \ldots, m$, the j-th being defined as follows: $\tilde{z}_j = -\varepsilon 1^T - e_j$. For every $\varepsilon \in (0, 1)$, each vector z_j has all components that are negative, hence it belongs to S_1. Thus, for $j = 1, \ldots, m$ and $\forall \varepsilon > 0$ it should be $y^T \tilde{z}_j = -\varepsilon 1^T y - y_j < 0$, which implies that y cannot be the null vector and $y \geq 0$. Thus, y is a solution of problem P_2 that has anser "yes". Summarizing: if problem P_1 has answer "no", then P_2 has answer "yes". This concludes the proof. □

1.3.1 Fist-order necessary conditions

We consider optimization problems with explicit constraints

$$\begin{aligned} \min f(x) & \\ x \in \mathbb{R}^n & \\ g_i(x) \leq 0 & \qquad i \in I \\ h_j(x) = 0 & \qquad j \in E \end{aligned}$$

and we assume $f, g_i, h_j \in C^1$

Definition 1.3 (feasible direction)
A vector $d \in \mathbb{R}^n, d \neq 0$ is a feasible direction in $x \in F$ if $\exists \delta > 0 : x + \alpha d \in F \quad \forall \alpha \in (0, \delta)$

Definition 1.4 (descendant direction)
A vector $d \in \mathbb{R}^n, d \neq 0$ is a descendant direction for f in $x \in F$ if $\exists \delta > 0 :: f(x + \alpha d) < f(x) \quad \forall \alpha \in (0, \delta)$

Theorem 1.5
Let $f : F \to \mathbb{R}$ be a continuous function. if $\bar{x} \in F$ is a local minimum for problem (P), then $D(\bar{x}) \cap F(\bar{x}) = \emptyset$

Proof. The proof is by contradiction. Assume that the thesis is false: there exists a vector $d \in D(\bar{x}) \cap F(\bar{x})$ and two positive numbers δ_1, δ_2 such that $f(\bar{x} + \alpha d) < f(\bar{x}) \quad \forall \alpha \in (0, \delta_1)$ and $f(\bar{x} + \alpha d) \in F \quad \forall \alpha \in (0, \delta_2)$. It follows that, for every $\alpha \in (0, \min\{\delta_1, \delta_2\})$, the point $y = \bar{x} + \alpha d$ belongs to F and has $f(y) < f(\bar{x})$, i.e., $\bar{x}$ cannot be a local minimum. □

Special case: only inequalities

Definition 1.5 (set of active constraints)
we define with

$$I_a(\bar{x}) = \{i \in I : g_i(\bar{x}) = 0\}$$

the set of active constraints

Definition 1.6 (set of feasible directions)
We define with

$$F_s(\bar{x}) = \{d \in \mathbb{R}^n, d \neq 0 : \nabla g_i(\bar{x})^T d < 0 \forall i \in I_a(\bar{x})\}$$

Theorem 1.6
Let $f : F \to \mathbb{R}$ be a continuous function. If $\bar{x} \in F$ is a local minimum for (P), then $D(\bar{x}) \cap F_s(\bar{x}) = \emptyset$

Proof. Let $\bar{x} \in F$ be a local minimum. By contradiction, assume that there exists a vector $d \in \mathbb{R}^n$ such that $d \in D(\bar{x}) \cap F_s(\bar{x})$. As $F_s(\bar{x}) \subseteq F(\bar{x})$ it muyst be $d \in D(\bar{x}) \cap F(\bar{x})$, thus contradicting Theorem 1.5 □

Theorem 1.7 (Fritz-John conditions)
Let $f \in C^1$ and $g_i \in C^1 \forall i \in I$. If $\bar{x} \in F$ is a local minimum for f over F, then there exist scalar numbers λ_0 and $\lambda_i (i \in I)$ such that

1. $\lambda_0 \nabla f(\bar{x}) + \sum_{i \in I} \lambda_i \nabla g_i(\bar{x}) = 0$
2. $\lambda_i g_i(\bar{x}) = 0 \quad \forall i \in I$
3. $\lambda_0 \geq 0, \lambda_i \geq 0 \quad (\forall i \in I)$ and not all λ are zero

Proof. Let $\bar{x} \in F$ be a local minimum. Let $, = |I_a(\bar{x})|$, and define an $(m+1) \times n$ matrix A in which

- row 0 corresponds to $\nabla f(\bar{x})^T$
- each row $i (i = 1, \ldots, m)$ corresponds to $\nabla g_i(\bar{x})^T$

According to theorem 1.6 there exists no vector $d \in \mathbb{R}^n$ such that

$$\nabla f(\bar{x})^T d < 0 \quad \text{and} \quad \nabla g_i(\bar{x})^T d < 0 \quad \forall i \in I_a(\bar{x})$$

i.e., there exists no vector $d \in \mathbb{R}^n$ such that $A^T d < 0$

Using Gordan's Theorem 1.4, this implies the existence of $m + 1$ scalars $\lambda_i \geq 0 (i = 0, \ldots, m)$ that are not all equal to zero and such that

$$\lambda_0 \nabla f(\bar{x}) + \sum_{i \in I_a(\bar{x})} \lambda_i \nabla g_i(\bar{x}) = 0$$

Setting $\lambda_i = 0 \quad \forall i \notin I_a(\bar{x})$ we prove 1. By construction, thus, each constraint $i \notin I_a(\bar{x})$ has $\lambda_i = 0$; as each remaining constraint $i \in I_a(\bar{x})$ has $g_i(\bar{x}) = 0$, 2. follows. Finally, 3. is a direct consequence of Gordan's Theorem. □

Definition 1.7 (Fritz-John point)
A point $x \in F$ is a Fritz-John point if it satisfies the Fritz-John conditions.

1.3.2 Karush-Kuhn-Tucker Conditions

Consider Fritz-John points for which $\lambda_0 > 0$, Wlog, we can assume $\lambda_0 = 1$

Definition 1.8 (Karush-Kuhn-Tucker point)
A point $x \in \mathbb{R}^n$ is a KKT point it there exist scalars $\lambda_i \geq 0 (i \in I)$ such that:

$$\nabla f(x) + \sum_{i \in I} \lambda_i \nabla g_i(x) = 0 \tag{1.1}$$

$$\lambda_i g_i(x) = 0 \qquad i \in I \tag{1.2}$$

$$g_i(x) \leq 0 \qquad i \in I \tag{1.3}$$

$$\lambda_i \geq 0 \qquad i \in I \tag{1.4}$$

Under some assumptions, it can be proven that each local minimum is a KKT point.

Definition 1.9 (Constraint qualification conditions)
The feasible set $F = \{x \in \mathbb{R}^n : g_i(x) \leq 0 (i \in I)\}$ satisfies the constraint qualification if one of the following conditions is satisfied:

1. linear constraints:
   all funcions g_i are linear, i.e., $g_i(x) = a_i^T x + b_i \quad \forall i \in I$
2. Slater's condition:
   all functions g_i are convex and differentiable, and F contains (at least) one point $\bar{x}$ in its strict interior (i.e., $g_i(\bar{x}) < 0 \quad \forall i \in I$)
3. Mangasarian-Fromovitz constraint qualification (MFCQ):
   the gradients of the active constraints in x are linearly independent, i.e., it is impossible to find scalars $\alpha_i \geq 0, i \in I_a(x)$, not all zero, such that $\sum_{i \in I_a(x)} \alpha_i \nabla g_i(x) = 0$

A solution $\bar{x}$ for which constraint qualification conditions are satisfied is called a *regular point*

Theorem 1.8
Let $f \in C^1$ and $g_i \in C^1 \quad \forall i \in I$. IF $\bar{x} \in F$ is a local minimum for f over F and some constraint qualification conditions are satisfied, then $\bar{x}$ is a KKT point.

Proof. For the sake of simplicity we only prove the result in the case where MFCQ are satisfied. Let $\bar{x}$ be a local minimum, and assume that the gradients of the active constraints in $\bar{x}$ are linearly independent. As $\bar{x}$ is a local minimum, Theorem 1.7 ensures that there are scalars $\lambda_0, \lambda_i (i \in I)$ that are not equal to zero and such that

$$\lambda_0 \nabla f(\bar{x}) + \sum_{i \in I} \lambda_i \nabla g_i(\bar{x}) = 0, \qquad \lambda_i g_i(\bar{x}) = 0 \quad \forall i \in I, \qquad \lambda_0 \geq 0, \qquad \lambda_i \geq 0 \forall i \in I$$

We want to show that these multipliers exist also in the case $\lambda_0 \neq 0$.

Note that $\lambda_i = 0$ for all constraints $i \notin I_a(\bar{x})$, thus $\lambda_0 \nabla f(\bar{x}) + \sum_{i \in I_a(\bar{x})} \xi_i \nabla g_i(\bar{x}) = 0$. If it were $\lambda_0 = 0$ is should be $\sum_{i \in I_a(\bar{x})} \lambda_i \nabla g_i(\bar{x}) = 0$, i.e. there would be a linear combination of the gradients of the active constraints that is equal to zero. By hypothesis, not all these multipliers are zero, meaning that the gradients of the active constraints should be linearly dependent, hence contradicting the hypothesis. □

1.3.3 General constrained optimization

The results of the previous section can be extended to the most general case in which the feasible set is defined by both equations and inequalities.

$$\begin{aligned} \min f(x) & \\ x \in \mathbb{R}^n & \\ g_i(x) \leq 0 & \qquad i \in I \\ h_j(x) = 0 & \qquad j \in E \end{aligned}$$

The following is a generalization of Fritz-John points

Theorem 1.9
Let $f \in C^1$, $g_i \in C^1 \forall i \in I$ and $h_j \in C^1 \forall j \in E$. If $\bar{x} \in F$ is a local minimum for f over F, then there exist scalars $\lambda_0, \lambda_i (i \in I)$ and $\mu_j (j \in E)$ that are not all zero and such that

1. $\lambda_0 \nabla f(\bar{x}) + \sum_{i \in I} \lambda_i \nabla g_i(\bar{x}) + \sum_{j \in E} \mu_j \nabla h_j(\bar{x}) = 0$
2. $\lambda_i g_i(\bar{x}) = 0 \quad \forall i \in i$
3. $\lambda_0 \geq 0, \quad \lambda_i \geq 0 \quad \forall i \in I$

Simliarly, we can define KKT points

Theorem 1.10 (KKT points)
Let $f \in C^1$, $g_i \in C^1 \forall i \in I$ and $h_j \in C^1 \forall j \in E$. If $\bar{x} \in F$ is a local minimum for f over F, and some constraint qualification conditions are satisfied, then $\bar{x}$ is a KKT point.

1.4 Lagrangian approaches to constrained optimization

Definition 1.10 (Relaxation)
Consider an optimization problem $\mathcal{P}$ defined as follows

$$(\mathcal{P}) \qquad z = \min f(x), \quad x \in F(\mathcal{P})$$

and define the following auxiliary problem $\mathcal{R}$:

$$(\mathcal{R}) \qquad z_r = \min \Phi(x), \quad x \in F(\mathcal{R})$$

We will say that $\mathcal{R}$ is a *relaxed problem* of $\mathcal{P}$ if the following conditions are satisfied:

(a) $F(\mathcal{P}) \subseteq F(\mathcal{R})$

(b) $\Phi(x) \leq f(x) \quad \forall x \in F(\mathcal{P})$

Thus a relaxation is a technique that: (a) enlarges the feasible set and/or (b) defines a new objective function that is ”better” than the original one in all points of the original feasible set

Theorem 1.11
Let $(\mathcal{P})$ be an optimization probelm (in min form) with optimal value z. Let $(\mathcal{R})$ be a relaxation of $\mathcal{P}$ with optimal value z_r. Then, $z_r \leq z$.

Proof. Let $\bar{x}$ be the optimal solution of problem $(\mathcal{P})$. By condition (b) we have $F(\mathcal{P}) \subseteq F(\mathcal{R})$, hence $\bar{x} \in F(\mathcal{R})$. By condition (a) we have $\Phi(\bar{x}) \leq f(\bar{x}) = z$. As the optimal solution of problem $(\mathcal{R})$ cannot be worse than the solution value in $\bar{x}$ (which is a feasible point for this problem), we have $z_r \leq \Phi(\bar{x})$, which concludes the proof. □

1.4.1 Lagrangian Relaxation

Let problem $(\mathcal{P})$ be defined as follows

$$\begin{aligned} (\mathcal{P}) \qquad z = \min f(x) & \\ x \in X & \\ g_i(x) \leq 0 & \qquad i \in I \\ h_j(x) = 0 & \qquad j \in E \end{aligned}$$

where $x \subseteq \mathbb{R}^n$ is a subset of $\mathbb{R}^n$ defined by implicit constraints. In the following, we will denote $m = |I|$ and $p = |E|$.

Let us introduce some parameters called *Lagrangian multipliers*

$$u_i \geq 0 \quad \forall i \in I \quad \text{and} \quad v_j \lesseqgtr 0 \quad \forall j \in E$$

and define an auxiliary problem

$$(\mathcal{R}) \qquad \ell(u,v) = \min_{x\in X} \mathcal{L}(x;u,v)$$

where the *Lagrangian function* $\mathcal{L} : X \to \mathbb{R}$ is a function that depends on parameters u and v, defined as

$$\mathcal{L}(x;u,v) = f(x) + \sum_{i\in I} u_i g_i(x) + \sum_{j\in E} v_j h_j(x)$$

Observe that multipliers u_i are associated with inequalities $g_i(x) \leq= 0$ and must be non-negative, while multipliers v_j are associated with equalities $h_j(x) = 0$ and have no such requirement.

Theorem 1.12 (Weak duality)
For any choice of the multipliers $u \in \mathbb{R}^m, u \geq 0$ and $v \in \mathbb{R}^p$ we have $\ell(u,v) \leq z$.

Proof. We now prove that $(\mathcal{R})$ is a relaxation of $(\mathcal{P})$. First observe that $F(\mathcal{P}) \subseteq F(\mathcal{R})$ as the latter is obtained by the former by removing some constraints. As per the second condition, let x be a feasible solution for problem $(\mathcal{P})$. We have

$$\forall i \in I : g_i(x) \leq 0 \quad \text{and} \quad u_i \geq 0 \quad \to \sum_{i\in I} u_i g_i(x) \leq 0$$

and

$$\forall j \in E : h_j(x) = 0 \qquad \to \sum_{j\in E} v_j h_j(x) = 0$$

which gives $\mathcal{L}(x;u,v) \leq f(x)$.

Thus, the direct application of Theorem 1.11 gives the result. □

1.4.2 Dual Lagrangian problem

The value of the lower bound $\ell(u,v)$ given by the Lagrangian relaxation depends on the choice of the u,v multipliers, and is a lower bound on z for all valid choices of the multipliers. We are interested in having the tightest bound possible in order to have a better approximation of the optimal solution value, so we define the *dual Lagrangian problem.*

$$(D) \qquad \bar{\ell} = \max_{u\geq 0, v} \ell(u,v)$$

Theorem 1.13
An optimal solution of the dual problem is a lower bound on the optimal solution of $(\mathcal{P})$, i.e.

$$\bar{\ell} \leq z$$

Proof. It descends from Theorem 1.12 □

The weak duality theorem 1.12 defines the possible relations between the optimal solution value z for the primal and the optimal dual solution $\bar{\ell}$:

(a) in general: $\bar{\ell} \leq z$

(b) if there exists $\bar{x} \in F(\mathcal{P})$ and $(\bar{u},\bar{v}) \in \mathbb{R}^m_+ \times \mathbb{R}^p$ such that $f(\bar{x}) = \ell(\bar{u},\bar{v})$, then $\bar{x}$ and $(\bar{u},\bar{v})$ are optimal solutions for the primal and dual problems

(c) if $z = \infty$ (unbounded primal), then $\ell(u,v) = -\infty \quad \forall (u,v) \in \mathbb{R}^m_+ \times \mathbb{R}^p$

(d) if $\bar{\ell} = \infty$ (unbounded dual), t4hen the primal is infeasible

In case (a), the quantity $z - \bar{\ell}$ is called *optimality gap.*

In case (b), the optimal solution values for the two problems coincide, and then

$$f(\bar{x}) = z = \bar{\ell} = \ell(\bar{u},\bar{v}) = \inf_x \mathcal{L}(x;\bar{u},\bar{v}) \leq f(\bar{x}) + \sum_{i\in I} \bar{u}_i g_i(\bar{x}) + \sum_{j\in E} \bar{v}_j h_j(\bar{x}) \leq f(\bar{x})$$

This chain of equations imposes that

$$\sum_{i\in I} \bar{u}_i g_i(\bar{x}) + \sum_{j\in E} \bar{v}_j h_j(\bar{x}) = 0$$

meaning that

(i) $\bar{u}_i g_i(\bar{x}) = 0 \quad \forall i \in I$

(ii) $\bar{v}_j h_j(\bar{x}) = 0 \quad \forall j \in E$

These constraints are called *orthogonality conditions.* The first class of conditions imposes that each pair of solutions $\bar{x}$ and $(\bar{u}, \bar{v})$ that are optimal for $(\mathcal{P})$ and (D):

- must have $\bar{u}_i = 0$ for all inequalities that are not tight
- must be tight, i.e. such that $g_i(\bar{x}) = 0$ for all inequalities associated to multipliers that are strictly positive.

Observe that $\bar{x}$ satisfies constraints $h_j(\bar{x}) = 0 \quad \forall j \in E$, as it is a feasible solution for $(\mathcal{P})$. Thus, conditions (ii) pose no constraints for what concerns multipliers v_j.

1.4.3 Lagrangian problem and KKT conditions

For the sake of simplicity, we will consider the casae in which $X = \mathbb{R}^n$, meaning that there are no implicit constraints but those that have been relaxed.

Definition 1.11 (saddle point)
A triplet $(\bar{x}, \bar{u}, \bar{v})$ with $\bar{x} \in \mathbb{R}^n, \bar{u} \in \mathbb{R}^m_+, \bar{v} \in \mathbb{R}^p$ is a *saddle point* if, $\forall x \in \mathbb{R}^n, u \in \mathbb{R}^m, v \in \mathbb{R}^p$ we have

$$\mathcal{L}(\bar{x}; u, v) \le \mathcal{L}(\bar{x}; \bar{u}, \bar{v}) \le \mathcal{L}(x; \bar{u}, \bar{v})$$

Theorem 1.14
Let $f, g_i (i = 1, \dots, m)$ and $h_j (j = 1, \dots, p)$ be continuous functions. Let $\bar{x} \in \mathbb{R}^n, \bar{u} \in \mathbb{R}^m, \bar{v} \in \mathbb{R}^p$. If $(\bar{x}; \bar{u}, \bar{v})$ is a saddle point, then

1. $g(\bar{x}) \le 0$ and $h(\bar{x}) = 0$
2. $\bar{u} \ge 0$
3. $\mathcal{L}(\bar{x}; \bar{u}, \bar{v}) = \min_{x \in \mathbb{R}^n} \mathcal{L}(x; \bar{u}, \bar{v})$
4. $\bar{u}_i g_i(\bar{x}) = 0 \quad i = 1, \dots, m$
5. $\bar{x}$ is a global minimum for $(\mathcal{P})$

Proof. 1. By definition, for any choice of multipliers $u \in \mathbb{R}^m_+$ and $\forall v \in \mathbb{R}^p$, we have $\mathcal{L}(\bar{x}; u, v) \le \mathcal{L}(\bar{x}; \bar{u}, \bar{v})$, i.e.,

$$(u - \bar{u})^T g(\bar{x}) + (v - \bar{v})^T h(\bar{x}) = \sum_{i=1}^{m} (u_i - \bar{u}_i) g_i(\bar{x}) + \sum_{j=1}^{p} (v_j - \bar{v}_j) h_j(\bar{x}) \le 0$$

Consider a family of m pairs of multipliers (u, v) defined changing one u multiplier k $(k = 1, \dots, m)$ at a time with respect to $\bar{u}$ as follows: $u_k = \bar{u}_k + 1, u_i = \bar{u}_i$ for $i = 1, \dots, m, i \ne k$ and $v_j = \bar{v}_j$ for $j = 1, \dots, p$. For the k-th pair, the relation above yields $g_k(\bar{x}) \le 0$, and thus we have $g(\bar{x}) \le 0$

□

1.5 Penalty Algorithms

The basic idea is to define an auxiliary problem P_c obtained removing some constraints from the definition of the feasible set F, and optimizing a new objective function that involves these constraints:

$$\min_{x \in \mathbb{R}^n} P(x; c) = f(x) + c\phi(x)$$

where $c > 0$ is a parameter and $\phi : \mathbb{R}^n \to \mathbb{R}$ is a *penalty function* that depends on the violation of the constraints. Ideally, we would like to have a function $\phi(x)$ such that

$$\phi(x) = \begin{cases} 0 & \text{if } x \in F \\ +\infty & \text{otherwise} \end{cases}$$

This requirement produces an unconstrained optimization problem that cannot be solved with classical algorithms, se we relax it as:

- ϕ is continuous;
- $\phi(x) = 0 \quad \forall x \in F$
- $\phi(x) > 0 \quad \forall x \notin F$

The most common choices for the penalty functions are:

- $\phi(x) = \max(0, \max_{i=1,\ldots,m}\{g_i(x)\}, \max_{j=1,\ldots,p\{|h_j(x)|\}})$
- $\phi(x) = \sum_{i=1}^{m} \max(g_i(x)) + \sum_{j=1}^{p} |h_j(x)|$
- $\phi(x) = \sum_{i=1}^{m} [\max(g_i(x), 0)]^2 + \sum_{j=1}^{p} |h_j(x)|^2$

Theorem 1.15
Let $\bar{x}$ be a local minimum for function $P(x; c)$ for some parameter c. If $\phi(\bar{x}) = 0$, then $\bar{x}$ is a local minimum for (P).

Proof. Since $\phi(\bar{x})$, then $\bar{x}$ is a feasible solution and $P(\bar{x}; c) = f(\bar{x})$. As $\bar{x}$ is a local minimum for $P(x; c)$, there exists a positive ρ and a neighbourhood $N(\bar{x}, \rho)$ for which

$$P(x'; c) \geq P(\bar{x}; c) \quad \forall x' \in N(\bar{x}, \rho)$$

Considering only feasible points, for every solution $x' \in N(\bar{x}, \rho) \cap F$, we have

$$f(x') = P(x'; c) \geq P(\bar{x}; c) = f(\bar{x})$$

hence $\bar{x}$ is a local minimum □

Theorem 1.16
Let c_k be a sequence of scalars such that $c_{k+1} > c_k \quad \forall k$. Denote by x^k an optimal solution of each auxiliary problem (P_{c_k}). If $\exists \bar{x} \in F$ such that $f(\bar{x}) = \min_{x \in F} f(x)$ then:

(a) $P(x^k; c_k) \leq f(\bar{x})$

(b) $\phi(x^{k+1}) \leq \phi(x^k)$

(c) $f(x^{k+1}) \geq f(x^k)$

(d) $P(x^k, c_k) \leq P(x^{k+1}; c_{k+1})$

Proof. (a) By definition, for a given value c_k, we have

$$P(x^k; c_k) = \min_{x \in \mathbb{R}^n} P(x; c_k) \leq \min_{x \in F} P(x; c_k) = \min_{x \in F} [f(x) + c_k \phi(x)] = \min_{x \in F} f(x) = f(\bar{x})$$

Thus, at each iteration k, the optimal solution value of the auxiliary problem provides a lower bound on the optimal solution value of $(\mathcal{P})$

(b) By definition of x^k and x^{k+1} we have

$$f(x^k) + c_k \phi(x^k) = P(x^k; c_k) = \min_{x \in \mathbb{R}^n} [f(x) + c_k \phi(x)] \leq [f(x^{k+1}) + c_k \phi(x^{k+1})]$$

$$f(x^{k+1}) + c_{k+1} \phi(x^{k+1}) = P(x^{k+1}; c_{k+1}) = \min_{x \in \mathbb{R}^n} [f(x) + c_{k+1} \phi(x)] \leq [f(x^k) + c_{k+1} \phi(x^k)]$$

summing both sides of these inequalities we get

$$(c_{k+1} - c_k) \phi(x^{k+1}) \leq (c_{k+1} - c_k) \phi(x^k)$$

hence the penalty function is non-increasing with the number of iterations.

(c) Again, by definition of x^k we have

$$f(x^k) + c_k \phi(x^k) \leq f(x^{k+1}) + c_k \phi(x^{k+1}) \leq f(x^{k+1}) + c_k \phi(x^k)$$

where the last inequality derives from (b). This gives $f(x^k) \leq f(x^{k+1})$, i.e. the solution value f is non-improving with the number of iterations.

(d) Finally, we have

$$P(x^k; c_k) \leq P(x^{k+1}; c_k) = f(x^{k+1}) + c_k\phi(x^{k+1}) \leq f(x^{k+1}) + c_{k+1}\phi(x^{k+1}) = P(x^{k+1}; c_{k+1})$$

Thus, the value of the lower bound computed at each iteration is monotonically non-decreasing with the number of iterations.

□

Theorem 1.17
Let c_k be a sequence of scalars that satisfy

$$c_{k+1} > c_k \quad \text{and} \quad \lim_{k\to\infty} c_k = \infty$$

and let $\{x^k\}$ be the sequence of points obtained by the solution of the associated auxiliary problems (P_{c_k}). If all points $\{x^k\}$ belong to a compact set $D \subset \mathbb{R}^n$, then

- $\lim_{k\to\infty} \phi(x^k) = 0$
- $\lim_{k\to\infty} f(x^k) = f(\bar{x})$
- $\lim_{k\to\infty} P(x^k; c_k) = f(\bar{x})$
- any limit point of sequence $\{x^k\}$ is an optimal solution for $(\mathcal{P})$
- $\lim_{k\to\infty} c_k\phi(x^k) = 0$

Chapter 2

Convex Optimization

Definition 2.1 (convex combination)
Let $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^n$. The *convex combination* of x and y is the set of points $z = \lambda x + (1 - \lambda)y$ for any $\lambda \in [0, 1]$.
When $\lambda \in (0, 1)$ we will refer to a *strict* convex combination

Definition 2.2 (convex set)
Let $F \subseteq \mathbb{R}^n$. Set F us convex if it contains the convex combination of every pair of points x and y that belong to F, i.e. if

$$\lambda x + (1 - \lambda)y \in F \quad \forall x, y \in F \text{ and } \forall \lambda \in [0, 1]$$

Definition 2.3 (cone)
A set $F \subseteq \mathbb{R}^n$ is a cone if $x \in F \implies \alpha x \in F \forall \alpha \geq 0$
A cone is a convex set

Definition 2.4 (hyperplane)
A hyperplane is the set $\{x \in \mathbb{R}^n : \alpha^T x = \alpha_0\}$ for some $\alpha \in \mathbb{R}^n$ and $\alpha_0 \in \mathbb{R}$
A hyperplane is a convex set

Definition 2.5 (halfspace)
A halfspace is the set $\{x \in \mathbb{R}^n : \alpha^T x \leq \alpha_0\}$ for some $\alpha \in \mathbb{R}^n$ and $\alpha_0 \in \mathbb{R}$
A halfspace is a convex set

Definition 2.6 (convex function)
Let $F \subseteq \mathbb{R}^n$ be a convex set, A funtion $f : F \to \mathbb{R}$ is *convex* if

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (\lambda)f(y) \quad \forall x, y \in F \text{ and } \forall \lambda \in [0, 1]$$

A function f is *concave* if function $-f$ is convex.
A function f is *strictly convex/concave* if a strict inequality holds in the previous definition for all $\lambda \in (0, 1)$

Lemma 2.1
Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. Then, the level set $F = \{x \in \mathbb{R}^n : f(x) \leq t\}$ is a convex set for any scalar t.

Proof. Let $x, y \in F$, i.e. such that $f(x) \leq t$ and $f(y) \leq t$. We now show that $f(z) \leq t$ for point $z = \lambda x + (1 - \lambda)y$ associated with any $\lambda \in [0, 1]$. By convexity of function f we have $f(z) = f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y) \leq \lambda t + (1 - \lambda)t = t$, i.e. $z \in F$ □

Theorem 2.1 (alternate definition for convexity)
Let $F \subseteq \mathbb{R}^N$ be a convex set and let $f : F \to \mathbb{R}$ be a continuously differentiable function. if f is convex, then

$$f(y) \geq f(x) + \nabla f(x)^T (y - x) \quad \forall x, y \in X$$

Proof. Let x and y be any two points in F. Consider their convex combination $\lambda x + (1-\lambda)y$ using coefficient $\lambda = 1-\epsilon$ for some $\epsilon > 0$. We have $\lambda x+(1-\lambda)y = (1-\epsilon)x+\epsilon y = x+\epsilon(y-x)$ and, similarly, $\lambda f(x)+(1-\lambda)f(y) = f(x) + \epsilon[f(y) - f(x)]$

By convexity of f we have

$$f\left(x + \epsilon(y-x)\right) = f\left(\lambda x + (1-\lambda)y\right) \leq \lambda f(x) + (1-\lambda)f(y) = f(x) + \epsilon[f(y) - f(x)] \tag{2.1}$$

consider now the first order expansion of the function in point x with displacement $h = \epsilon(y-x)$:

$$f\left(x + \epsilon(y-x)\right) = f(x) + \epsilon \nabla f(x)^T (y-x) + R_1(x, \epsilon\|y-x\|) \tag{2.2}$$

Comparing (2.1) and (2.2) we have

$$f(x) + \epsilon[f(y) - f(x)] \geq f(x) + \epsilon \nabla f(x)^T (y-x) + R_1(x, \epsilon\|y-x\|)$$

for sufficiently small values of ϵ, we have

$$\lim_{\epsilon \to 0} \frac{R_1(x, \epsilon\|y-x\|)}{\epsilon\|y-x\|} = 0$$

This yields

$$f(y) - f(x) \geq \nabla f(x)^T (y-x)$$

□

Theorem 2.2
Let $F \subseteq \mathbb{R}^n$ be a convex set and let $f : F \to \mathbb{R}$ be a continuously differentiable function. Assume that the Hessian matrix $\nabla^2 f(x)$ exists and is positive semidifinite for a point $x \in F$. Then, there exists a neighbourhood of x where f is convex.

Proof. Let $x \in F$ be such that $\nabla^2 f(x)$ exists and is positive semidifinite. Let $y \in F$ be another point that is sufficiently close to x. The second order approximation of the function is:

$$f(y) = f(x) + \nabla f(x)^T (y-x) + \frac{1}{2}(y-x)^T \nabla^2 f(x)(y-x) + R_2(x, \|y-x\|)$$

Thus, for sufficiently small $\|y-x\|$, condition $\nabla^2 f(x) \succeq 0$ implies

$$f(y) \geq f(x) \nabla f(x)^T (y-x)$$

and the application of Theorem 2.1 proves that f is convex. □

Relevant convex functions

- $f : \mathbb{R}^n \to \mathbb{R}, x \to \|x\|$ is convex for any norm $\|\cdot\|$
- a quadratic function $f : \mathbb{R}^n \to \mathbb{R}, x \to \frac{1}{2}x^T Q x + c^T x + r$ is convex iff $Q \geq 0$. Indeed, for this function we have $\nabla^2 f(x) = Q$
- Let $f_1, \ldots, f_m$ be m convex functions. Then, their linear combination $f(x) = \sum_{i=1}^m \lambda_i f_i(x)$ is convex if $\lambda_i \geq 0 \forall i$
- Let $f_1(x), \ldots, f_m(x)$ be m convex functions from $\mathbb{R}^n$ to $\mathbb{R}$. Then, the function $f(x) : \mathbb{R}^n \to \mathbb{R}, x \to \max\{f_1(x), \ldots, f_m(x)\}$ is a convex function.

  Proof. Let $x, y \in \mathbb{R}^n$ and $\lambda \in [0,1]$. We have

$$f\left(\lambda x + (1-\lambda)y\right) = \max_{i=1,\ldots,m} f_i\left(\lambda x + (1-\lambda)y\right) \leq \max_{i=1,\ldots,m} [\lambda f_i(x) + (1-\lambda)f_i(y)] \leq$$
$$\lambda \max_{i=1,\ldots,m} f_i(x) + (1-\lambda) \max_{i=1,\ldots,m} f_i(y) \leq \lambda f(x) + (1-\lambda)f(y)$$

□

- Let $f : \mathbb{R}^{n+m} \to \mathbb{R}, (x, y) \to f(x, y)$ be a function that, $\forall y \in \mathbb{R}^m$, is convex wrt x. Then, the function $g : \mathbb{R}^n \to \mathbb{R}, x \to \max_{y \in \mathbb{R}^m} f(x, y)$ is convex.

  Proof. Let $x^1 \in \mathbb{R}^n$ and $x^2 \in \mathbb{R}^n$, and consider their combination with multiplier $\lambda \in [0, 1]$. We have

$$g\left(\lambda x^1 + (1-\lambda)x^2\right) = \max_{y \in \mathbb{R}^m} f\left(\lambda x^1 + (1-\lambda)x^2, y\right) \leq \max_{y \in \mathbb{R}^m} [\lambda f(x^1, y) + (1-\lambda) f(x^2, y)] \leq$$
$$\max_{y \in \mathbb{R}^m} [\lambda f(x^1, y)] + \max_{y \in \mathbb{R}^m} [(1-\lambda) f(x^2, y)] = \lambda g(x^1) + (1-\lambda) g(x^2)$$

□

A consequence of this result is that the dual lagrangian problem is a convex optimization problem, regardless of the nature of the original objective function f and of the constraint functions g_i and h_j. It is sufficient to observe that, for some multipliers u, v the lagrangian relaxation is defined as

$$\ell(u, v) = \min_{x \in \mathbb{R}^n} \mathcal{L}(x; u, v)$$

and the lagrangian function is

$$\mathcal{L}(x; u, v) = f(x) + u^T g(x) + v^T h(x)$$

i.e., it is linear wrt multipliers u and v.

It follows that function

$$\ell : \mathbb{R}^{m+p} \to \mathbb{R}, (u, v) \to \min_{x \in \mathbb{R}^n} \mathcal{L}(x; i, v)$$

is concave in the space $\mathbb{R}^{m+p}$ for any f, g and h.

The dual lagrangian problem asks to maximize function $\ell(u, v)$, i.e. to minimize the convex function $-\ell(u, v)$. The feasible set for this problem is defined only by non-negativity constraints for u multipliers, hence the feasible set is a convex set.

2.1 Main properties of convex programming

A convex optimization problem is defined as

$$(P^c) \qquad \min f(x)$$
$$x \in F$$

where the feasible set $F \subseteq \mathbb{R}^n$ is a convex set, and the objective function $f : F \to \mathbb{R}$ is convex.

Theorem 2.3
Let $F \subseteq \mathbb{R}^n$ be a convex set and $f : F \to \mathbb{R}$ be a convex function. If $\bar{x} \in F$ is a local minimum for f over F, then $\bar{x}$ is also a global minimum for f over F.

Proof. The proof is by contradiction: let $\bar{x}$ be a local minimum for f over F, and assume there exists another point $y \in F$ such that $f(y) < f(\bar{x})$.

By definition of local minimum, there exists a scalar $\rho > 0$ such that $f(x) \geq f(\bar{x}) \quad \forall x \in B(\bar{x}.\rho)$.

Now, consider a convex combination of $\bar{x}$ and y defined as $z = \lambda\bar{x} + (1-\lambda)y$ for a sufficiently small value of $\lambda \in (0, 1)$ such that $z \in B(\bar{x}, \rho)$. By convexity of function f, we have $f(z) \leq \lambda f(\bar{x}) + (1-\lambda) f(y)$. As $f(y) < f(\bar{x})$, we get $f(z) < f(\bar{x})$, which contradicts the initial assumption that $\bar{x}$ is a minimum over $B(\bar{x}, \rho)$. □

Theorem 2.4
Let $F \subseteq \mathbb{R}^n$ be a convex set and let $f : F \to \mathbb{R}$ be a convex function. Let $S = \{x \in F : f(x) \leq f(y) \forall y \in F\}$ be the set of points in F for which f attains a minimum. Then, S is a convex set.

Proof. As local and global minimum coincide, we denote by f^* the value taken by the objective function f on any point in S. Let $x \in S$ and $y \in S$ be two optimal solutions, and let $z = \lambda x + (1-\lambda)y$ eb their convex combination with coefficient $\lambda \in [0, 1]$. Then we have

$$f(z) \leq \lambda f(x) + (1-\lambda) f(y) = \lambda f^* + (1-\lambda) f^* = f^*$$

As f^* is the minimum of f over F, we cannot have $f(z) < f^*$, hence $f(z) = f^*$, i.e. $z \in S$. □

Theorem 2.5
Let $F \subseteq \mathbb{R}^n$ be a convex set and let $f : F \to \mathbb{R}$ be a strictly convex function. Then, if $\bar{x}$ is a global minimum, it is the only global minumum.

Proof. By contradiction assume there exists a point $y \in F, y \neq x$ such that $f(y) = f(\bar{x})$. For any $\lambda \in (0,1)$ it should be $f(\lambda\bar{x} + (1-\lambda)y) < \lambda f(\bar{x}) + (1-\lambda)f(y) = f(\bar{x})$, hence $\bar{x}$ would not be a global minimum. □

Remark. In the proofs of the results above we only required the objective function be convex. If $f \in C^1$, one can use the first order conditions to distinguish global minima.

Theorem 2.6
Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function, $f \in C^1$. A solution $\bar{x}$ is a global minimum for f iff $\nabla f(\bar{x}) = 0$

Proof. If $\bar{x}$ is a global minimum, it is also a local minimum; hence, first order necessary conditions for unconstrained optimization impose that $\nabla f(\bar{x}) = 0$

Conversely, if f is convex and $\nabla f(\bar{x}) = 0$, theorem 2.1 impllies $f(y) \geq f(\bar{x}) \quad \forall y \in F$ □

Theorem 2.7
Let $F \subseteq \mathbb{R}^n$ be a convex set and let $f : F \to \mathbb{R}$ be a convex function, $f \in C^1$. A solution $\bar{x} \in F$ is a global minimum for f over F iff $\nabla f(\bar{x})^T(y - \bar{x}) \geq 0$ for all $y \in F$.

Proof. If the condition holds, Theorem 2.1 implies that $\bar{x}$ is a minimum.

Conversely, assume by contradiction that $\bar{x}$ is a minimum and the condition is not satisfied, i.e. there exists $y \in F$ such that $\nabla f(\bar{x})^T(y - \bar{x}) < 0$.

Consider a convex combination of $\bar{x}$ and y defined as $z = \bar{x} + \lambda(y - \bar{x})$ for a sufficiently small value of $\lambda \in (0,1)$, and observe that $z \in F$ by convexity of the feasible set. Using the first order approximation of the objective function we have:

$$f(z) = f(\bar{x}) + \nabla f(\bar{x})^T(z - \bar{x}) + R_1(\bar{x}, \lambda\|z - \bar{x}\|) = f(\bar{x}) + \lambda\nabla f(\bar{x})^T(y - \bar{x}) + R_1(\bar{x}, \lambda\|z - \bar{x}\|)$$

For $\lambda \to 0$ the residual R_1 can be ignored. This yields $f(z) < f(\bar{x})$ and contradicts the hypotheis. □

2.1.1 Lagrangian relaxation

Theorem 2.8
Let (P^c) be a convex optimization problem. Then $(\bar{x}, \bar{u}, \bar{v})$ is a saddle point iff $\bar{x}$ is a KKT point with multipliers $\bar{u}$ and $\bar{v}$

Proof. Let $(\bar{x}, \bar{u}, \bar{v})$ be a saddle point. Then, $\bar{x}$ is a minimum for the lagrangian function $\mathcal{L}(x; \bar{u}, \bar{v})$, and hence $\nabla_x[\mathcal{L}(x; \bar{u}, \bar{v})]|_{x=\bar{x}} = 0$, i.e.

$$\nabla f(\bar{x}) + \sum_{i=1}^{m} \bar{u}_i \nabla g_i(\bar{x}) + \sum_{j=1}^{p} \bar{v} \nabla h_j(\bar{x}) = 0$$

As $(\bar{x}, \bar{u}, \bar{v})$ is a saddle point, we also have $g(\bar{x}) \leq 0, h(\bar{x}) = 0, \bar{u} \geq 0$ and $\bar{u}^T g(\bar{x}) = 0$, which shows that $\bar{x}$ is a KKT point with multipliers $\bar{u}$ and $\bar{v}$.

Conersely, let $\bar{x}$ be a KKT point with multipliers $\bar{u} \in \mathbb{R}^m_+$ and $\bar{v} \in \mathbb{R}^p$, i.e. such that

$$\nabla f(\bar{x}) + \sum_{i=1}^{m} \bar{u}_i \nabla g_i(\bar{x}) + \sum_{j=1}^{p} \bar{v} \nabla h_j(\bar{x}) = 0 \qquad \bar{u}^T g(\bar{x}) = 0 \qquad \bar{u} \geq 0$$

By convexity of f and g_i, for each $x \in \mathbb{R}^n$, we have

$$\begin{aligned} f(x) &\geq f(\bar{x}) + \nabla f(\bar{x})(x - \bar{x}) \\ g_i(x) &\geq g_i(\bar{x}) + \nabla g_i(\bar{x})(x - \bar{x}) \end{aligned}$$

As for the linear function h_j we have

$$h_j(x) = h_j(\bar{x}) + \nabla h_j(\bar{x})(x - \bar{x})$$

summing the three equations above (weighted with the respective multipliers), we get

$$\mathcal{L}(x;\bar{u},\bar{v}) = f(x) + \sum_{i=1}^{m} \bar{u}_i g_i(x) + \sum_{j=1}^{p} \bar{v}_j h_j(x) \geq$$

$$f(\bar{x}) + \nabla f(\bar{x})^T(x-\bar{x}) + \sum_{i=1}^{m} \bar{u}_i[g_i(\bar{x}) + \nabla g_i(\bar{x})^T(x-\bar{x})] + \sum_{j=1}^{p} \bar{v}_j[h_j(\bar{x}) + \nabla h_j(\bar{x})^T(x-\bar{x})] =$$

$$\mathcal{L}(\bar{x};\bar{u},\bar{v}) + [\nabla f(\bar{x}) + \sum_{i=1}^{m} \bar{u}_i \nabla g_i(\bar{x}) + \sum_{j=1}^{p} \bar{v}_j \nabla h_j(\bar{x})](x-\bar{x})$$

i.e., $\mathcal{L}(x;\bar{u},\bar{v}) \geq \mathcal{L}(\bar{x};\bar{u},\bar{v}) + \nabla_x[\mathcal{L}(\bar{x};\bar{u},\bar{v})]^T(x-\bar{x})$. For a KKT point it must be $\nabla_x \mathcal{L}(\bar{x};\bar{u},\bar{v}) = 0$, then

$$\mathcal{L}(x;\bar{u},\bar{v}) \geq \mathcal{L}(\bar{x};\bar{u},\bar{v}) \quad \forall x \in \mathbb{R}^n$$

In addition, as $\bar{x}$ is a KKT point with multipliers $\bar{u}$ and $\bar{v}$, we have $g(\bar{x}) \leq 0, h(\bar{x}) = 0, \bar{u} \geq 0$ and $\bar{u}^T g(\bar{x}) = 0$. Thus, $\forall u \in \mathbb{R}^m_+$ and $v \in \mathbb{R}^p$ we have $\mathcal{L}(\bar{x};u,v) = f(\bar{x}) + u^T g(\bar{x}) + v^T h(\bar{x}) \leq f(\bar{x}) + \bar{u}^T g(\bar{x}) + \bar{v}^T h(\bar{x}) = \mathcal{L}(\bar{x};\bar{u},\bar{v})$ and hence $(\bar{x},\bar{u},\bar{v})$ is a saddle point. □

A relevant consequence of this theorem is that in convex optimization there cannot be an optimality gap. Indeed, let $(\bar{x},\bar{u},\bar{v})$ be a saddle point. The lagrangian relaxation with multipliers $\bar{u},\bar{v}$ produces a lower bound with value

$$\min_x \mathcal{L}(x;\bar{u},\bar{v}) = \mathcal{L}(\bar{x};\bar{u},\bar{v}) = f(\bar{x})$$

which is the same value of the objective function evaluated in point $\bar{x}$

2.2 Interior point methods

We consider a convex optimization problem of the form

$$\begin{array}{rl} (P^c) & \min f(x) \\ & Ax = b \\ & g_i(x) \leq 0 \quad i \in I = \{1,\ldots,m\} \\ & x \in \mathbb{R}^n \end{array}$$

We will denote by $S = \{x \in \mathbb{R}^n : g_i(x) \leq 0 \quad (i \in I)\}$ the set of points that satisfy all inequalities.

We start from a point inside S and at each iteration find a new feasible solution that stays in the interior of S. Thus, a basic assumptio for interior points methods is that there exists a feasible solution in the interior of S, i.e.

$$\exists \bar{x} \in \mathbb{R}^n : A\bar{x} = b \text{ and } g(\bar{x}) < 0$$

In addition we assume that the problem is bounded, i.e. $\min_{x \in F} > -\infty$. Under these hypotheses, the Slater constraint qualification conditions are satisfied, hence a local minimum is a KKT point.

The logarithmic barrier method is based on the definition of an auxiliary problem

$$\begin{array}{rl} (P_{aux}) & \min p(x) \\ & x \in \mathbb{R}^n Ax = b \end{array}$$

In this problem the objective function is given by

$$p(x) = f(x) + t\phi(x)$$

Where $t > 0$ is a parameter, and the penalty function $\phi(x)$ is defined as

$$\phi(x) = \sum_{i \in I} -\log(-g_i(x))$$

The objective function $p(x)$ is convex and differentiable. Thus, the auxiliary problem can be efficiently solved using e.g. a variant of Newton's method for optimization under linear constraints.

For a given value of $t > 0$, let us denote byu $x^*(t)$ an optimal solution of the auxiliary problem. By definition, this solution has

$$Ax^*(t) = b \qquad \text{and} \qquad g_i(x^*(t)) < 0 \quad i = 1, \ldots, m$$

As the auxiliary problem is convex, solution $x^*(t)$ is a KKT point, i.e. there exist multipliers $\tilde{v} \in \mathbb{R}^p$ such that

$$\nabla p(x^*(t)) + A^T \tilde{v} = 0$$

Given that $\nabla p(x) = \nabla f(x) + t\nabla\phi(x)$ and that $\nabla\phi(x) = \sum_{i=1}^{m} \frac{1}{-g_i(x)} \nabla g_i(x)$, the KKT conditions for x^* are

$$\nabla f(x^*(t)) + t \sum_{i=1}^{m} \frac{1}{-g_i(x^*(t))} \nabla g_i(x^*(t)) + A^T \tilde{v} = 0$$

Defining

$$\lambda_i^*(t) = -t \frac{1}{g_i(x^*(t))} < 0 \quad (i = 1, \ldots, m) \qquad \text{and} \qquad v^*(t) = \tilde{v}$$

these conditions can be rewritten as

$$\nabla f(x^*(t)) + \sum_{i=1}^{m} \lambda_i^*(t) \nabla g_i(x^*(t)) + A^T v^*(t) = 0 \tag{2.3}$$

Let us define the lagrangian relaxation of the original problem using multipliers $\lambda_i^*(t)$ and $v^*(t)$. Observe that $\lambda_i^*(t) > 0$ as $g_i(x^*(t)) < 0$

$$\mathcal{L}(x; \lambda^*(t), v^*(t)) = f(x) + \sum_{i=1}^{m} \lambda_i)^*(t) g_i(x) + {v^*}^T(t)(Ax - b)$$

Imposing the gradient of the lagrangian function to be equal to zero, we get (2.3); hence $x^*(t)$ is a minimum for the lagrangian problem. The quality of the associated lower bound is determined by the following theorem

Theorem 2.9
The duality gap associated with $x^*(t)$ is at most mt

Proof. because the problem is convex, and $(x^*(t), \lambda^*(t), v^*(t))$ is a KKT point, due to theorem 2.8 it is also a saddle point. The associated lagrangian lower bound is

$$\ell(\lambda^*(t), v^*(t)) = \inf_{x \in \mathbb{R}^n} \mathcal{L}(x; \lambda^*(t), v^*(t)) = \mathcal{L}(x^*(t); \lambda^*(t), v^*(t)) = f(x^*(t)) - mt$$

□

Given a required accuracy ϵ, it is sufficient to solve the auxiliary problem (P_{aux}) with $t = \epsilon/m$. However, small values of ϵ produce small values of t, which makes optimization of function $p(x)$ hard. Thus, it is preferrable to use an iterative scheme that starts with a "large" value of t, and reduces it at each iteration as

$$t^{k+1} = \mu t^k$$

with $\mu < 1$. At each iteration k, the algorithm has $t^k = \mu^k t^0$, and the associated gap is $m\mu^k t^0$, thus to get accuracy ϵ in k iterations, we require

$$m\mu^k t^0 \leq \epsilon$$

i.e.

$$k \log \mu \leq \log\left(\frac{\epsilon}{t^0 m}\right)$$

because $\mu < 1 \implies \log\mu < 0$ the duality gap is below the threshold after at most

$$\left\lceil \frac{\log\left(\frac{\epsilon}{t^0 m}\right)}{|\log \mu|} \right\rceil$$

2.2.1 Computing an initial solution

One can consider an additional problem of the form

$$\min\{s : AX = b \quad g_i(x) \leq s \quad (i = 1, \ldots, m)\}$$

and check whether the corresponding optimal value is negative. If this is not the case, there exists no solution x such that $Ax = b$ and $g(x) < 0$, hence the initial assumption is violated and the method must be halted. Otherwise, a feasible solution which is in the interior of S is available, and the method must determine an initial value of parameter t^0. A value of t^0 that is too large will produce an auxiliary problem that is hard to solve. Typically, the first phase produces a dual solution, which allows to estimate the optimality gap. In this case, the value of t^0 is fixed so that the optimality gap at the first iteration is comparable with that of the initial solution.

2.2.2 Central path

The set of points $\{x^*(t^k)\}_k$ that are produced by the algorithm is called *central path.* The central path conditions can be interpreted as relaxed KKT conditions, as it obeys conditions

- $Ax = b, \quad g(x) \leq 0$
- $\lambda \geq 0$
- $\nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x) + A^T v = 0$
- $-\lambda_i g_i(x) = t \quad i = 1, \ldots, m$

The following result holds for convex optimization:

Theorem 2.10
Let f, g_i $(i = 1, \ldots, m)$ and ϕ be continuous and convex functions in the interior of S. Then, any point induced by sequence $\{x^*(t^k)\}_k$ produced by a barrier method is a global minimum for the problem.

Proof. Let $\bar{x}$ denote the point to which sequence $\{x^*(t^k)\}_k$ converges. At each k we have $x^*(t^k) \in S$, thus $\bar{x} \in S$. In addition, we have

$$\lim_{k\to\infty} \left[f(x^*(t^k)) + t^k\phi(x^*(t^k))\right] = f(\bar{x}) + \lim_{k\to\infty} \left[t^k\phi(x^*(t^k))\right] = f(\bar{x})$$

By contradiction, let $x^* \in S$ be a global minimum such that $f(x^*) < f(\bar{x})$. By continuity of function f, $\exists \tilde{x} \in \text{int } S$ such that $f(x^*) < f(\tilde{x}) < f(\bar{x})$. At each iteration we have

$$f(x^*(t^k)) + t^k\phi(x^*(t^k)) \leq f(\tilde{x}) + t^k\phi(\tilde{x})$$

hence for $k \to \infty$

$$\lim_{k\to\infty}[f(x^*(t^k)) + t^k\phi(x^*(t^k))] \leq \lim_{k\to\infty}[f(\tilde{x}) + t^k\phi(\tilde{x})] = f(\tilde{x}) + \lim_{k\to\infty}[t^k\phi(\tilde{x})] = f(\tilde{x})$$

combining the last two equations we get

$$f(\tilde{x}) \geq \lim_{k\to\infty}[f(x^*(t^k)) + t^k\phi(x^*(t^k))] = f(\bar{x})$$

which contradicts the definition of $\tilde{x}$, and the existence of x^*. □