Constrained Optimization
Theory of Constrained Optimization

Contents

1 Vocabulary and notations **1**

2 Building intuition toward necessary conditions **2**
2.1 A first practical case . . . 2
2.2 A second practical case . . . 2
2.3 Tangent cone and constraint qualifications . . . 3

3 Optimality conditions **4**
3.1 First-order optimality conditions . . . 4
3.2 Second-order conditions . . . 5
3.2.1 Second-order necessary conditions . . . 5
3.2.2 Second-order sufficient conditions . . . 5
3.3 Lagrande multipliers and sensitivity . . . 5

4 Duality **6**

1 Vocabulary and notations

We consider the general formulation of a constrained optimization problem that is :

$$\min_{x\in\mathbb{R}^n} f(x) \quad \text{subject to} : \begin{cases} c_i(x) = 0 & i \in \mathcal{E} \\ c_i(x) \geq 0 & i \in \mathcal{I} \end{cases} \tag{1}$$

where f and the $(c_i)_{i\in\mathcal{E}\cup\mathcal{I}}$ are smooth, real-valued function on a subset of $\mathbb{R}^n$, and $\mathcal{E}$ and $\mathcal{I}$ are two finit sets of indices. We'll use the following denotation :

$$\begin{cases} f \leftarrow \text{ objective function} \\ (c_i)_{i\in\mathcal{E}} \leftarrow \text{ equality constraints} \\ (c_i)_{i\in\mathcal{I}} \leftarrow \text{inequality constraints} \end{cases} \tag{2}$$

We define the *feasible set* Ω as to be the subset of $\mathbb{R}^n$ of points that satisfy the constraints :

$$\Omega = \{x \in \mathbb{R}^n \,|\, c_i x = 0, \forall i \in \mathcal{E} \text{ and } c_i(x) \geq 0, \forall i \in \mathcal{I}\} \tag{3}$$

The problem (1) can therefore be rewritten more compactly :

$$\min_{x\in\Omega} f(x) \tag{4}$$

We'll define the **active set** $\mathcal{A}(x)$ at any $x \in \Omega$ as the union of equality constraints indices with the saturated inequality constraints :

$$\forall x \in \Omega, \quad \mathcal{A}(x) = \mathcal{E} \cup \{i \in \mathcal{I} \,|\, c_i(x) = 0\} \tag{5}$$

The purpose of the following section is to derive mathematical characterizations of the solutions to the previous equation. We therefore need to insist on the notion of local and global solutions.

- a vector x^* is said to be a local solution to (1) if $x^* \in \Omega$ and there is a neighborhood $V(x^*)$ such that $f(x) \geq f(x^*)$, $\forall x \in \Omega \cap V(x^*)$.
- x^* is a *strict* local solution if the last inequality is strict.
- x^* is an isolated local solution if $x^* \in \Omega$ and there is a neighborhood $V(x^*)$ such that x^* is the only local solution in $\Omega \cap V(x^*)$

2 Building intuition toward necessary conditions

2.1 A first practical case

Let us consider the first case where :

$$\mathcal{E} = \{1\}, \quad \mathcal{I} = \emptyset \tag{6}$$

We are going to examine the first-order Taylor series approximation to the objective and constraint function.To retain feasibility with respect to the condition $c_1(x)$, we require making a small step s so that :

$$\begin{aligned} 0 &= c_1(x+s) \\ &\simeq c_1(x) + s^T \nabla c_1(x) \\ &= s^T \nabla c_1(x) \end{aligned} \tag{7}$$

On the other hand, we want s to produce a decrease of the objective function f, hence :

$$\begin{aligned} 0 &> f(x+s) - f(x) \\ &\simeq s^T \nabla f(x) \end{aligned} \tag{8}$$

Hence the first order conditions require that we find a direction $s \in \mathbb{R}^n$ so that :

$$\begin{cases} (x+s) \in \Omega \\ s^T \nabla c_1(x) = 0 \\ s^T \nabla f(x) < 0 \end{cases} \tag{9}$$

This suggest that if x^* is a local minimizer, there exist no such direction s and therefore $\nabla c_1(x^*)$ and $\nabla f(x^*)$ are *parallel* (if not, take $d = -\left(I - \frac{\nabla c_1(x) \nabla c_1(x)^T}{\|\nabla c_1(x)\|}\right) \nabla f(x)$).

By introducing the **Lagragian function** :

$$\begin{aligned} \mathcal{L} \quad : \quad & \mathbb{R}^n \times \mathbb{R} \to \mathbb{R} \\ & (x, \lambda_1) \quad \to \quad f(x) - \lambda_1 c_1(x) \end{aligned} \tag{10}$$

and noticing that

$$\nabla_x \mathcal{L}(x, \lambda_1) = \nabla f(x) - \lambda_1 \nabla c_1(x) \tag{11}$$

the parallel condition see earlier writes : at the solution x^*, $\exists \lambda_1^*$ sp that :

$$\nabla_x \mathcal{L}(x^*, \lambda_1^*) = 0 \tag{12}$$

This first result suggests that e can search solutions of the equality constrained problem by seeking *stationary points* of the Lagrangian function. The scalar λ_1 is called a Lagrange multiplier for the constraint $c_1(x)$. This reasoning gives out a necessary solutions, clearly not sufficient.

2.2 A second practical case

We now consider the opposite practical case of the single inequality condition :

$$\mathcal{E} = \emptyset, \text{ and } \mathcal{I} = \{1\} \tag{13}$$

As before, we conjecture that a given feasible point x is not optimal if we can find a small step that both retains feasibility and decreases the objective function. This leads to finding $s \in \mathbb{R}^n$ so that

$$\begin{aligned} &\nabla f(x)^T s < 0 \\ &c_1(x) + s^T \nabla c_1(x) \leq 0 \text{ with } c_1(x) \geq 0 \end{aligned} \tag{14}$$

Two cases have to be treated appart :

- x lies *inside* the feasible area. Therefore any step vector small enough that decreases the objective function works. The only time such a step cannot be found (and therefore local optimality has been reached) is when $\nabla f(x) = 0$.
- x lies on the boundary of the feasible set, so that $c_1(x) = 0$. The previous conditions become :

$$\begin{cases} s^T \nabla c_1(x) \geq 0 \\ s^T \nabla f(x) < 0 \end{cases} \tag{15}$$

which defines the intersection of an open-half space and a closed one. With the above drawing, it

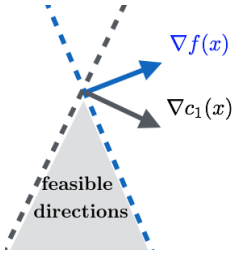

Figure 1: First-order feasible directions

is clear that such a direction does not exist (therefore we reached local optimality) when

$$\nabla f(x) = \lambda_1 \nabla c_1(x), \quad \lambda_1 \geq 0 \tag{16}$$

Such an optimality conditions can be summed up by making reference to the Lagrangian :

$$\nabla_x \mathcal{L}(x^*, \lambda_1^*) = 0 \text{ for some } \lambda_1^* \geq 0 \tag{17}$$

with $\lambda_1^* c_1(x^*) = 0$. This latest equation is known as the complementary condition.

2.3 Tangent cone and constraint qualifications

The approach we had in the latest section (i.e first-order Tayor expansion of the constraint and objective functions to form a problem where both constraints and objective functions are linear) makes sense only when the linearized approximation capture the essential geometric features of the feasible set near the point x considered.

Constraint qualifications are assumptions that ensure similarity of the constraint set Ω and its linearized approximation in a neighborhood of x^*. We will give two definitions, one for the tangent cone and another for set of linearized feasible directions. Both are cones of $\mathbb{R}^n$.

Set of feasible directions

Given a feasible point x and the active constraint set $\mathcal{A}(x)$, the set of linearized feasible directions $\mathcal{F}(x)$ is given by the directions that can be taken to keep x feasible in the first order :

$$\forall x \in \Omega, \quad \mathcal{F}(x) = \left\{ d \,\middle|\, \begin{array}{ll} d^T \nabla c_i(x) = 0, & i \in \mathcal{E} \\ d^T \nabla c_i(x) \geq 0, & i \in \mathcal{I} \cap \mathcal{A}(x) \end{array} \right\} \tag{18}$$

Tangent cone

The vector d is said to be a tangent vector to Ω at $x \in \Omega$ if there is a feasible sequence $(z_k)_k$ approaching x (i.e $z_k \in \Omega, \forall k \in \mathbb{N}$ and $z_k \underset{k\to\infty}{\to} x$) and a sequence of positive scalars $(t_k)_k$ with $t_k \underset{k\to\infty}{\to} 0$ with :

$$d = \lim_{k\to\infty} \frac{z_k - x}{t_k} \tag{19}$$

The set of all tangent directions to Ω in x is called the tangent cone in x and is noted $T_\Omega(x)$.

More precisely, constraint qualifications are conditions underwhich the linearized feasible set $\mathcal{F}(x)$ is similar to the tangent cone $T_\Omega(x)$. They ensure that $\mathcal{F}(x)$ captures the essential features of Ω in the vicinity of x, and they allow to apply the first-order reasoning we held before.

We introduce here the **Linear Independent Constraint Qualification** or LICQ. Given the point $x \in \Omega$ and the set $\mathcal{A}(x)$, we say that the linear independent constraint qualification holds in the set of active constraint gradients :

$$\{\nabla c_i(x) \,|\, i \in \mathcal{A}(x)\} \tag{20}$$

is linearly independent.

3 Optimality conditions

3.1 First-order optimality conditions

Here we state first-order necessary conditions for $x^* \in \Omega$ to be a local minimizer. For the general constrained optimization problem (1) we introduce the corresponding Lagrangian :

$$\begin{array}{rcccl} \mathcal{L} & : & \mathbb{R}^n \times \mathbb{R}^m & \to & \mathbb{R} \\ & & (x, \lambda) & \to & f(x) - \displaystyle\sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i c_i(x) \end{array} \tag{21}$$

with $m = |\mathcal{E}| + |\mathcal{I}|$.

Karush-Kuhn-Tucker first order conditions

Let x^* be a local solution to the problem (1). Assuming that f and $(c_i)_{i \in \mathcal{E} \cup \mathcal{I}}$ are continuously differentiable on $\Omega \subset \mathbb{R}^n$, and that the LICQ holds at x^*, we have the following results.
$\exists \lambda \in \mathbb{R}^m$ such that the following are verified at (x^*, λ^*) :

$$\nabla_x \mathcal{L}(x^*, \lambda^*) = 0 \tag{22}$$

$$c_i(x^*) = 0, \quad \forall i \in \mathcal{E} \tag{23}$$

$$c_i(x^*) \geq 0, \quad \forall i \in \mathcal{I} \tag{24}$$

$$\lambda_i^* \geq 0, \quad \forall i \in \mathcal{I} \tag{25}$$

$$\lambda_i^* c_i(x^*) = 0, \quad \forall i \in \mathcal{E} \cup \mathcal{I} \tag{26}$$

The first condition can be rewritten as :

$$\nabla f(x^*) - \sum_{i \in \mathcal{A}(x)} \lambda_i^* \nabla c_i(x^*) = 0 \tag{27}$$

The proof to the following result is quite complex, but relies on fundamental results of constrained optimization. We'll detail only one of them here, given hereinafter.

Indeed, we have that if x^* is a local solution of the general constrained optimization problem, then :

$$\forall d \in T_\Omega(x), \quad d^T \nabla f(x^*) \geq 0 \tag{28}$$

Proof : *Let $d \in T_\Omega(x)$ such that*

$$d^T \nabla f(x) < 0 \tag{29}$$

Then by definition, $\exists (z_k)_k \in \Omega^{\mathbb{N}}$ and $(t_k)_k \in \mathbb{R}^N$ so that :

$$d = \frac{z_k - x}{t_k} \tag{30}$$

and therefore

$$z_k = x - t_k d \tag{31}$$

Since :

$$\begin{aligned} f(z_k) &= f(x) + (z_k - x)\nabla f(x) + o(\|z_k - x\|) \\ &= f(x) + t_k d \nabla f(x) + o(\|z_k - x\|) \\ &< f(x) + \frac{1}{2} t_k d \nabla f(x) \text{ for } k \text{ sufficiently large } \quad < f(x) \end{aligned} \tag{32}$$

which is in contradiction with the fact that x is a local minimizer, hence proving the result.

3.2 Second-order conditions

What role do the second order derivatives of f and $(c_i)_i$ play in the optimality conditions ? They actually serve as a tie-breaking rule (same principle as the Hessian curvature for unconstrained problems). In the following, we'll assume f and $(c_i)_{i \in \mathcal{E} \cup \mathcal{I}}$ are twice continuously differentiable.

Given $\mathcal{F}(x^*)$ and some Lagrange multipliers λ^* satisfying the KKT conditions, we define the *critical cone* as follows :

$$\mathcal{C}(x^*, \lambda^*) = \left\{ w \in \mathcal{F}(x^*) \,|\, \nabla c_i(x^*)^T w = 0,\, i \in \mathcal{A}(x^*) \cap \mathcal{I} \text{ with } \lambda_i^* > 0 \right\} \tag{33}$$

which describes the directions that would tend to adhere to the active inequality constraints even when we where to make small changes to the objectives, as well as equality constraints.

3.2.1 Second-order necessary conditions

Suppose that x^* is a local solution to (1) and that the LICQ holds. Let λ^* be the Lagrange multiplier for which the KKT conditions are satisfied. Then :

$$w^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) w \geq 0, \qquad \forall w \in \mathcal{C}(x^*, \lambda^*) \tag{34}$$

3.2.2 Second-order sufficient conditions

Let $x^* \in \mathbb{R}^n$ be a feasible point and $\lambda^* \in \mathbb{R}^m$ so that the KKT conditions hold. If :

$$w^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*) w > 0, \qquad \forall x \in \mathcal{C}(x^*, \lambda^*) - \{0\} \tag{35}$$

then x^* is a strict local solution for (1).

3.3 Lagrande multipliers and sensitivity

One can wonder what is the intuitive signification of the Lagrange multipliers. We'll show here that λ_i^* is a sensitivity measure of the optimal objective value $f(x^*)$ to the presence of the constraint c_i.

Let $i \notin \mathcal{A}(x^*)$ such that $c_i(x^*) > 0$. If we perturb c_i by a tiny amount the constraint will still be inactive and x^* will still be a local solution to the constrained problem. Since $i \notin \mathcal{A}(x^*)$, $\lambda_i^* = 0$ truthfully indicates that c_i plays no role in the behavior of the optimal point that is x^*.

Now let $i \in \mathcal{A}(x^*)$. Let us assume that we observe a small perturbation for the constraint c_i :

$$c_i(x) \geq -\varepsilon \|c_i(x^*)\| \tag{36}$$

instead of $c_i(x) \geq 0$. We assume ε to be small enough so that the perturbed solution $x^*(\varepsilon)$ still has the same set of active constraints as initially, as well as the Lagrange multipliers being not affected by the solution. Therefore we have :

$$\begin{aligned} -\varepsilon \|\nabla c_i(x^*)\| &= c_i(x^*(\varepsilon)) - c_i(x^*) \\ &\simeq (x^*(\varepsilon) - x^*)^T \nabla c_i(x^*) \end{aligned} \tag{37}$$

and

$$0 = c_j(x^*(\varepsilon)) - c_j(x^*) \simeq (x^*(\varepsilon) - x^*)^T \nabla c_j(x^*) \tag{38}$$

On the other hand, thanks to the KKT solutions :

$$\begin{aligned} f(x^*(\varepsilon)) - f(x^*) &\simeq (x^*(\varepsilon) - x^*)^T \nabla f(x^*) \\ &= \sum_{j \in \mathcal{A}(x^*)} \lambda_j^* (x^*(\varepsilon) - x^*)^T \nabla c_i(x^*) \\ &= -\varepsilon \lambda_i \|c_i(x^*)\| \end{aligned} \tag{39}$$

which therefore gives us the result :

$$\frac{df(x^*(\varepsilon))}{d\varepsilon} = -\lambda_i \|c_i(x^*)\| \tag{40}$$

Therefore, if $\lambda_i \|c_i(x^*)\|$ is large, the optimal value is more sensitive to the placement of the ith constraint.

4 Duality

This theory is developed here to motivate some important algorithms (augmente Lagrangians, ..). We show here how we construct an alternative problem from the functions and data that define the original problem. In some case, it is easier to solve computationally than the primal problem. Let us rewrite the original problem as :

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t} \quad c(x) \geq 0, \quad \forall i \in \{1, \dots, m\} \tag{41}$$

with $c(x) = (c_1(x), \dots, c_m(x))^T$. This allows us to rewrite the corresponding Lagrangian :

$$\begin{array}{rcccl} \mathcal{L} & : & \mathbb{R}^n \times \mathbb{R}^m & \to & \mathbb{R} \\ & & (x, \lambda) & \to & f(x) - \lambda^T c(x) \end{array} \tag{42}$$

We define the *dual objective function* as :

$$\begin{array}{rcccl} q & : & \mathbb{R}^n & \to & \mathbb{R} \\ & & \lambda & \to & \inf_x \mathcal{L}(x, \lambda) \end{array} \tag{43}$$

and we define the *dual problem* as :

$$\max_{\lambda \in \mathbb{R}^m} q(\lambda) \tag{44}$$

The optimal value of the dual problem gives a lower bound on the optimal objective for the primal problem. This is a consequence of the **weak duality** result :

Weak duality

$\forall x \in \Omega$ and any $\bar{\lambda} \geq 0$, we have that :

$$q(\bar{\lambda}) \leq f(\bar{x}) \tag{45}$$

Such a result can lead in certain case to using the solutions to the dual problem to derive solutions to the primal problem :

> Suppose that f and $(c_i)_{i\in\{1,\dots,m\}}$ are convex and continuoulsy differentiable on $\mathbb{R}^n$. Let $\bar{x} \in \Omega$ be a solution to the primal problem at which the LICQ holds. Suppose that $\hat{\lambda}$ solves the dual problem, and that the infimum in $\inf_x \mathcal{L}(x, \hat{\lambda})$ is reached at $\hat{x}$. if $\mathcal{L}(\cdot, \hat{\lambda})$ is a strictly convex function, then $\hat{x} = \bar{x}$.