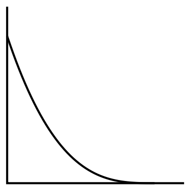
31 Lecture 31

[31.1] **Why talking about the linearizing cone?** Because, it is difficult to draw the pictures. Without them applying FONC is difficult, as the computations of feasible directions $D(a)$ becomes difficult. So, instead, if can find another set which is close (English word) to $D(a)$ and which can be computed a bit mechanically, we will be good. Such a candidate is the linearizing cone. ⌟

But then we have a natural question. Can we extend FONC to $\mathcal{D}(a)$? That is, if a is a point of local minimum, then we know that for each direction $d \in \overline{D}(a)$, we have $D_d f(a) \geq 0$. Can we say this for each $d \in \mathcal{D}(a)$? Why? Because, if the answer is yes, then we could have got a stronger and more helpful necessary condition. (And we would never bother about $D(a)$.) No, unfortunately (and as expected), it is not possible to extend FONC to $\mathcal{D}(a)$, as seen in the following example.

[31.2] **Example: extending FONC to $\mathcal{D}(a)$ is not possible** Consider



$$\begin{array}{ll} \min & f(x) \equiv -x_1 \\ \hline & g_1(x) \equiv (1-x_1)^3 - x_2 \geq 0 \\ \text{s.t.} & g_2(x) \equiv x_1 \geq 0 \\ & g_3(x) \equiv x_2 \geq 0. \end{array}$$

a) The point $a = (1, 0)$ is a point of absolute minimum. (This is the only point with highest first coordinate.)

b) Only g_1 and g_3 are active at a. So

$$\begin{aligned} \mathcal{D}(a) &= \left\{d \mid \nabla g_1(a)^T d \geq 0, \quad \nabla g_3(a)^T d \geq 0\right\} \\ &= \{d \mid -d_2 \geq 0, \quad d_2 \geq 0\} \\ &= \{d \mid d_2 = 0\}. \end{aligned}$$

c) Note that $e_1 \in \mathcal{D}(a)$ and $\langle \nabla f(a), e_1 \rangle = -1 < 0$. So we could not say that $D_d f(a) \geq 0$ for each $d \in \mathcal{D}(a)$.

d) Note that $\overline{D}(a) = \{d \mid d_1 \leq 0, d_2 = 0\}$. So $D_d f(a) \geq 0$ for each $d \in \overline{D}(a)$, as expected.

[31.3] Discussion

a) Imagine the surface $T:\ x^2+y^2+z^2=1$. Take $F \equiv x^2+y^2+z^2-1$.

b) Take the point $a=(1,0,0)$ on it. There is a tangent plane passing through a. What is the plane parallel to it passing through origin?

b1) Is it $D_x f(a)x + D_y F(a)y + D_z F(a)z = 0$?

c) Take the point $a=(0,1,0)$ on it. There is a tangent plane passing through a. What is the plane parallel to it passing through origin?

b1) Is it $D_x F(a)x + D_y F(a)y + D_z F(a)z = 0$?

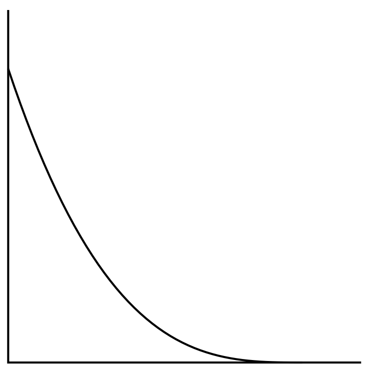
d) Take the point $a=(\frac{1}{2},\frac{1}{2},\frac{1}{\sqrt{2}})$ on it. There is a tangent plane passing through a. What is the plane parallel to it passing through origin? (Here note the radius.)

d1) Is it $D_x F(a)x + D_y F(a)y + D_z F(a)z = 0$?

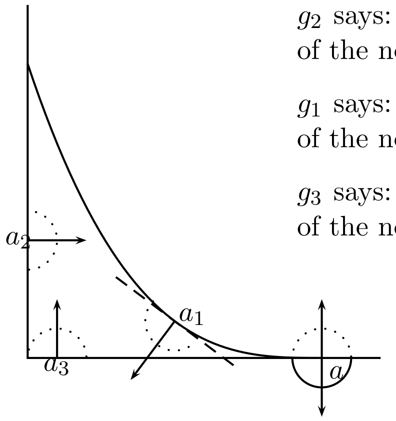
e) If I have a differentiable surface $F(x,y,z)=c$, and a is a point on the surface, then what does $(D_xF(a), D_yF(a), D_zF(a))$ stand for?

[31.4] Re-analyze the previous example We had

$$\begin{array}{ll} \min & f(x) \equiv -x_1 \\ \hline & g_1(x) \equiv (1-x_1)^3 - x_2 \geq 0 \\ \text{s.t.} & g_2(x) \equiv x_1 \geq 0 \\ & g_3(x) \equiv x_2 \geq 0. \end{array}$$



Re-analyze previous example The three constraints tell you three stories.



g_2 says: if you move in direction d_2 within 90° of the normal at a_2 you may stay inside T.

g_1 says: if you move in direction d_1 within 90° of the normal at a_1 you may stay inside T.

g_3 says: if you move in direction d_3 within 90° of the normal at a_3 you may stay inside T.

So, at a, we have $\mathcal{D}(a)$, the intersection of them: the x-axis.

Is $\overline{D}(a) \subsetneq \mathcal{D}(a)$ here?

We will talk about the linearizing cone more. But before that one small observation for sets defined by linear constraints.

[31.5] **Proposition** Consider $\begin{array}{ll} \min & f(x) \\ \text{s.t.} & Ax \geq b \end{array}$. Let a be a feasible point. Then $\mathcal{D}(a) = \overline{D}(a) = D(a)$.

Proof. Notice that $g \equiv A_a x - b_a \geq 0$ are precisely the active constraints at a, for which ∇g^t is precisely A_a. (These are the hyperplanes that pass through a.)

We already know that $D(a) \subseteq \mathcal{D}(a)$. Conversely, let $d \in D(a)$. This means, $A_a d \geq 0$. So $A_a(a + \alpha d) \geq b_a$ for each $\alpha > 0$. Furthermore, $\overline{A}_a a > \overline{b}_a$. Hence, by continuity, there exists $\delta > 0$ such that $\overline{A}_a(a + \theta d) > \overline{b}_a$ for each $\theta \in [0, \delta]$.

So, overall, for each $\theta \in [0, \delta]$, we have $A(a + \alpha d) \geq b$. So $d \in D(a)$. ∎

[31.6] **Practice** Put $f(x) = x^4 \sin(\frac{1}{x})$ for $x \neq 0$ and $f(0) = 0$. Imagine that we have to minimize y in the region of $\mathbb{R}^2$ where $y - f(x) \geq 0$. Compute $D(0,0)$ and $\mathcal{D}(0,0)$

Generalized Lagrange multipliers

[31.7] Discussion

a) As we have seen previously, '$d \in \mathcal{D}(a)$ need not imply $\langle \bigtriangledown f(a), d\rangle \geq 0$ at a local minimum.

b) But, under some additional conditions called the REGULARITY CONDITIONS on the functions (the definition will be given later), we can make $\langle \bigtriangledown f(a), d\rangle \geq 0$ hold true for each $d \in \mathcal{D}(a)$.

c) This is good news, as it gives us a better necessary condition. Plus verification is more mechanical here.

d) That is, if the problem satisfies the regularity conditions, then the set $\boldsymbol{Z(a)}$ defined as

$$\begin{aligned} Z(a) \quad &:= \quad \mathcal{D}(a) \ \cap \ \Big\{ d \mid \langle \bigtriangledown f(a), d\rangle < 0 \Big\} \\ &= \quad \Big\{ d \mid \quad \langle \bigtriangledown g_i(a), d\rangle \geq 0, \forall i \in A(a), \quad \langle \bigtriangledown h_j(a), d\rangle = 0, \forall j, \quad \langle \bigtriangledown f(a), d\rangle < 0 \Big\}, \end{aligned}$$

necessarily becomes $\emptyset$, at a point of local minimum a.

e) So, for such problems (that satisfy the regularity conditions) we will first search for those points a for which $Z(a) = \emptyset$. (Like, the way we used to search for critical points in the unconstrained case.)

f) But, there is a problem. Verifying whether $Z(a)$ is empty or not, may not be easy. Do we have some equivalent criterion? Yes, the generalized Lagrange multipliers.

g) Let $f, g_i, h_j \in \mathcal{C}^1$ and consider the problem

$$(P2) \qquad \begin{array}{ll} \min & f \\ \hline \text{s.t.} & g_i(x) \geq 0, i = 1, \ldots, m, \quad h_j(x) = 0, j = 1, \ldots, p. \end{array}$$

The LAGRANGIAN FUNCTION associated with (P2) is the function L defined as

$$L(x, \lambda, w) := f - \sum \lambda_i g_i - \sum w_j h_j.$$

The factors λ_i and w_j are called the GENERALIZED LAGRANGE MULTIPLIERS.

[31.8] **Theorem** Let a be a feasible point of (P2). TFAE.

a) The set $Z(a) = \emptyset$.

b) There exist $\lambda_i \geq 0$ and $w_j \in \mathbb{R}$, such that $\bigtriangledown L(a, \lambda, w) = 0$ and $\lambda_i g_i(a) = 0$ for each i.

Proof. a)$\Rightarrow$b).

Let $Z(a) = \emptyset$. So for each $d \in \mathcal{D}(a)$ we have $\bigtriangledown f(a)^t d \geq 0$. A vector $d \in \mathcal{D}(a)$ is nothing but a vector that satisfies $B^t d \geq 0$, where

$$B = \begin{bmatrix} \bigtriangledown g_i(a)^t \\ \hline \bigtriangledown h_j(a)^t \\ \hline - \bigtriangledown h_j(a)^t \end{bmatrix},$$

and that the top block of the matrix B corresponds to the constraints in $A(a)$. Thus

$$B^t d \geq 0 \quad \Rightarrow \quad \bigtriangledown f(a)^t d \geq 0.$$

By Farka's lemma, $\exists y \geq 0$ such that $\bigtriangledown f(a)^t = y^t B^t$, that is, $By = \bigtriangledown f(a)$.

That is, $\exists\lambda_i \geq 0$ (these are the y_i corresponding to the active g_i constraints) and w_j (this w_j is the difference of two components of y) such that

$$\bigtriangledown f(a) - \sum_{i \in A(a)} \lambda_i \bigtriangledown g_i - \sum_j w_j \bigtriangledown h_j = 0.$$

Put $\lambda_i = 0$ for $i \notin A(a)$. Then we see that

$$\bigtriangledown f(a) - \sum_i \lambda_i \bigtriangledown g_i - \sum_j w_j \bigtriangledown h_j = 0.$$

Now to show that $\lambda_i g_i(a) = 0$ for each i, notice that if $g_i(a) = 0$, then $\lambda_i g_i(a) = 0$. If $g_i(a) > 0$, then $i \notin A(a)$ and $\lambda_i = 0$ (by our choice). So, in this case too, $\lambda_i g_i(a) = 0$.

b)$\Rightarrow$a).

Assume that b) holds. We want to show that $Z(a) = \emptyset$. That is, $\{d \mid d \in \mathcal{D}(a),\ D_d f(a) < 0\} = \emptyset$.

For that, let $d \in \mathcal{D}(a)$. By definition, for each $i \in A(a)$, we have $\bigtriangledown g_i^t d \geq 0$ and for each j we have $\bigtriangledown h_j^t d = 0$.

As $\lambda_i g_i(a) = 0$ holds for each i by the hypothesis of b), we see that $\lambda_i = 0$ for each $i \notin A(a)$.

Hence, from the hypothesis of b), we get that

$$\begin{array}{rcl}
\bigtriangledown f(a)^t d & = & \sum\limits_{i \in A(a)} \lambda_i \bigtriangledown g_i(a)^t d \quad + \quad \sum\limits_{i \notin A(a)} \lambda_i \bigtriangledown g_i(a)^t d \quad + \quad \sum_j w_j \bigtriangledown h_j(a)^t d \\
 & = & \text{nonnegative (as } d \in \mathcal{D}(a)) \quad + \quad 0 \text{ (as } \lambda_i = 0 \text{ here)} \quad + \quad 0 \text{ (as } d \in \mathcal{D}(a)) \\
 & \geq & 0 \qquad .
\end{array}$$

Thus $Z(a) = \emptyset$. ■