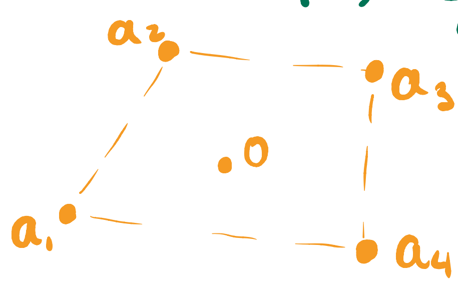
Lecture 5

Last time	Today
▷ Subgradients	▷ Gordon's Theorem of alternatives
▷ Normals	▷ Optimality conditions with functional constraints.
▷ Optimality conditions with convex sets	

Gordon's Theorem of Alternatives

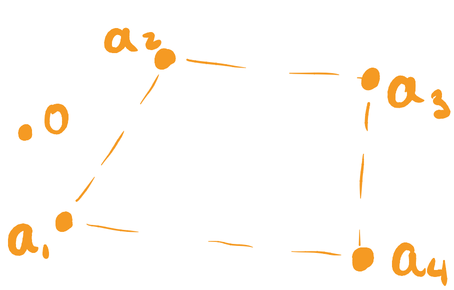
Theorem (Gordon). For any collection $a_1, \dots, a_m \in E$, exactly one of the following is true

(i) $\exists \lambda \in \mathbb{R}^m_+$, $\sum_{i=1}^m \lambda_i a_i = 0$, $\sum_{i=1}^m \lambda_i = 1$.



(0 is in the convex hull of $\{a_i\}$.)

(ii) $\exists x \in E$, $\langle a_i, x \rangle < 0$ $\forall i \in [m]$.



(0 is not in the convex hull of $\{a_i\}$.)

This is particularly useful to derive "certificates of infeasiability," we will use it to derive optimality conditions.

To prove this result we will prove an auxiliary theorem.

Theorem (ε) If $f: E \to \mathbb{R}$ is differentiable and bounded below, then, there exist $x_i \in E$ s.t. $\nabla f(x_i) \to 0$.

Proof of Theorem (ε): Fix $\varepsilon > 0$, then

$h_\varepsilon(\cdot) = f(\cdot) + \varepsilon \|\cdot\|$ has bounded sublevel sets and it is continuous

$\Rightarrow$ There exists x_ε a minimizer of h_ε.

For $v = -\nabla f(x_\varepsilon)$ and $t > 0$ we have

$$\frac{f(x_\varepsilon + td) - f(x_\varepsilon)}{t} = \frac{1}{t}\Big(f(x_\varepsilon + td) + \varepsilon \|x_\varepsilon + td\| - (f(x_\varepsilon) + \varepsilon \|x_\varepsilon\|) - \varepsilon \|x_\varepsilon + td\| + \varepsilon \|x_\varepsilon\| \Big)$$

x_ε is a minimizer

$$\geq \frac{\varepsilon}{t}\left(\|x_\varepsilon\| - \|x_\varepsilon + td\| \right)$$

triangle ineq.

$$\geq -\varepsilon \|d\|.$$

Note that the lower bound is independent of t. Thus taking a limit:

$$-\varepsilon\|\nabla f(x_\varepsilon)\| \le \lim_{t\downarrow 0} \frac{f(x_\varepsilon + td) - f(x_\varepsilon)}{t}$$

$$= \langle \nabla f(x_\varepsilon), d\rangle \quad \leftarrow -\nabla f(x_\varepsilon).$$

$$= -\|\nabla f(x_\varepsilon)\|^2$$

Thus, $\|\nabla f(x_\varepsilon)\| \le \varepsilon$. $\square$

Proof of Gordon's Theorem: We will show that the following are equivalent

(1) $f(x) = \log\left(\sum_{i=1}^m \exp(\langle a_i, x\rangle)\right)$ is bounded below. $\leftarrow$ softmax

(2) System (i) is solvable

(3) System (ii) is not solvable.

In HW 2 you'll prove (2) $\Rightarrow$ (3) $\Rightarrow$ (1).
We prove (1) $\Rightarrow$ (2). Since f is bounded from below Treorem (ε) guarantees we can find x_k with $\nabla f(x_k) \to 0$.
Computing gradients gives

$$\nabla f(x_k) = \sum_{i=1}^{m} \lambda_i^k a_i \quad \text{with} \quad \lambda_i^k = \frac{\exp(\langle a_i, x_k \rangle)}{\sum_{j=1}^{m} \exp(\langle a_j, x_k \rangle)}.$$

Clearly, $\lambda^k \in [0,1]^n$, thus WLOG we can assume $\lambda^k \to \lambda$ for some $\lambda \geq 0$ such that

$$\sum \lambda_i = 1, \quad \sum \lambda_i a_i = \lim_{k=0} \nabla f(x_k) = 0.$$

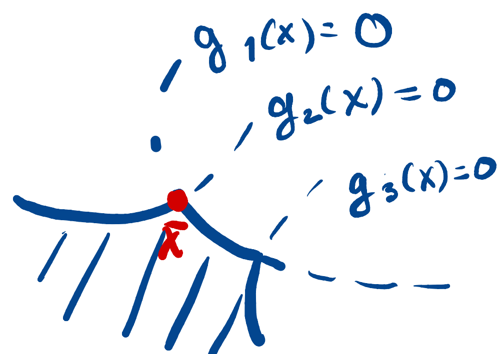
$\square$

Optimality conditions with functional constraints

Recall our second problem of interest

$$(\heartsuit) \quad \begin{aligned} \min \ & f(x) \\ \text{s.t.} \ & g_i(x) \leq 0 \end{aligned}$$

Differentiable



A way to understand optimality conditions is by making it "unconstrained". To do this we penalize the objectives with the constrains via the Lagrangian

$$L(x; \lambda) := f(x) + \sum_{i=1}^{m} \lambda_i g_i(x)$$

$\lambda \geq 0$

Def: Given $\bar{x}$ satisfying $g_i(x) \leq 0$ $\forall i \in [m]$. We say that $\lambda \geq 0$ is a **Lagrange multiplier vector** for $\bar{x}$ if

KKT conditions

1) $\bar{x}$ is a critical point of $x \mapsto L(x;\lambda)$, i.e.,

$$\nabla f(\bar{x}) + \sum \lambda_i \nabla g_i(\bar{x}) = 0.$$

2) Complementary slackness holds:

$$\lambda_i g_i(\bar{x}) = 0 \quad \forall i \in [m].$$

$\dashv$

The following theorems show that these vectors exist.

Theorem (Fritz John): If $\bar{x}$ is a local minimizer of (P) with f and g_i differentiable at $\bar{x}$. Then, $\exists (\lambda_0, \lambda) \in \mathbb{R}_+ \times \mathbb{R}_+^m$ nonzero s.t. complementary slackness holds and

$$\lambda_0 \nabla f(\bar{x}) + \sum \lambda_i \nabla g_i(\bar{x}) = 0.$$

$\dashv$

Warning!

If we had $\lambda_0 > 0$, then $\frac{\lambda}{\lambda_0}$ would be a Lagrange Multiplier vector. But in general λ_0 could be zero. We will come back to this problem after the proof.

Proof: Define the set of active constraints

$$I(\bar{x}) = \{ i \in [m] \mid g_i(x) = 0 \}.$$

Let

$$h(x) = \max \left\{ f(x) - f(\bar{x}), \max_{i \in I(\bar{x})} \{ g_i(x) \} \right\}.$$

Note that for x near $\bar{x}$ s.t. $g_i(x) \leq 0 \ \forall i$,

$$h(x) \geq f(x) - f(\bar{x}) \geq 0.$$

Note that near $\bar{x}$, points can only violate constraints in $I(\bar{x})$. Thus, if x near $\bar{x}$ has $g_i(x) > 0$, then $i \in I(\bar{x})$ and

$$h(x) = g_i(x) > 0.$$

Thus $\bar{x}$ is a local minimizer of h without any constraints.

Claim (HW2): For any $v \in E$:

$$h'(x;v) = \max\left\{ \langle \nabla f(x), v\rangle, \max_{i \in I(\bar{x})} \{ \langle \nabla g_i(x), v\rangle \} \right\}.$$

Then, by the optimality condition we proved last time

$$h'(\bar{x}, v) \geq 0 \quad \forall v \in E.$$

This is equivalent to

$$\nexists v \quad \text{s.t.} \quad \langle \nabla f(\bar{x}), v\rangle < 0$$
$$\langle \nabla g_i(\bar{x}), v\rangle < 0 \quad \forall i \in I(\bar{x}).$$

By Gordon's Theorem of alternatives we have

$$\exists \lambda_0, \lambda \in \mathbb{R}_+ \times \mathbb{R}^m_+ \quad \text{s.t.} \quad \lambda_0 + \sum_{i \in [m]} \lambda_i = 1$$

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

we simply set $\lambda_i = 0$ if $g_i(x) < 0$.

This completes the proof since complementary slackness holds by construction. ◻

Question: How to guarantee $\lambda_0 > 0$?

Notice that if $\lambda_0 = 0$, then we have $\exists \lambda \geq 0$ nonzero s.t.

$$\sum_{i \in I(\bar{x})} \lambda_i \nabla g_i(\bar{x}) = 0$$

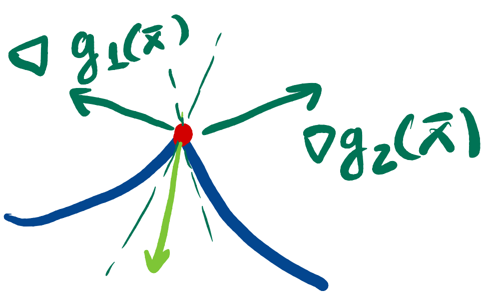
or equivalently (after normalization)

$$\lambda \geq 0, \quad \sum \lambda_i = 1 \quad \text{and} \quad \sum_{i \in I(\bar{x})} \lambda_i \nabla g_i(\bar{x}) = 0.$$

Once again we can prevent this from happening using Gordon's Theorem if we impose

$$\boxed{\exists v \in E^* \text{ s.t. } \langle \nabla g_i(\bar{x}), v \rangle < 0 \quad \forall i \in I(\bar{x}).}$$

Mangasarian-Fromovitz Constrained Qualification



The MFCQ condition is requiring a direction

with instantaneous decrease of active constraints.
We have proven the following.

Theorem (Karush-Kuhn-Tucker (KKT) conditions)
If $\bar{x}$ is a local minimizer, f, g_i are differentiable at $\bar{x}$ and MFCQ holds. Then, there exists a Lagrange Multiplier vector. $\dashv$

Lemma: For convex, differentiable g_i. MFCQ at $\bar{x}$ if, and only if, there is a point x_s s.t. $g_i(x_s) < 0$.
Slater point.

Proof: Suppose MFCQ holds. Then, for small $t>0$, $\bar{x}+tv$ is strictly feasible. (why?). If x_s exists take $d = x_s - \bar{x}$

$$\Rightarrow \quad \frac{g_i(\bar{x}+tv) - g_i(\bar{x})}{t} < 0 \quad \text{for } t=1, i \in I(\bar{x})$$

since this function is non decreasing

we conclude $\langle \nabla g_i(\bar{x}), v \rangle = g_i'(\bar{x}; v) < 0.$

$\square$