**Variational principles.** We want to extend the characterization of minimizers as critical points to situations where minimizers are absent and only "almost minimizing" points are there.

Ekeland's "variational principle" delivers some version:

**Theorem.** Let \( f: \mathbb{R}^n \to \mathbb{R} \) is continuous, Gateaux differentiable and bounded from below. Then for any \( a \in \mathbb{R}^n \), \( f(a) = \inf_{x \in \mathbb{R}^n} f + \epsilon \), there exists a point \( x_\epsilon \) such that:

- \( f(x_\epsilon) = f(a) \)
- \( \text{dist}(a, x_\epsilon) \leq \epsilon \)
- \( \|\nabla f(a)\| \leq \epsilon. \)

In other words, for any point where value of \( f \) is close to best possible (but unattainable, in general) value \( \inf f \), there is point \( x_\epsilon \) not far from \( a \), which has even smaller value of \( f \), and whose gradient of \( f \) is small.

The proof is essentially contained in the following picture:

Let \( \inf f =: c \geq -\infty. \)
We consider the family of functions $x \mapsto h \equiv \|x - x^*_x\|$. For $h \leq c$, this function is below $f(x)$. For $h = f(x^*_x)$ it is equal to at least at $x^*_x$. There is smallest $h^*_h$ for which $h \equiv \|x - x^*_x\|$ equals $h$ somewhere, but still never goes above it. Let $x^*_h$ be the point where $h^*_h - \|x^*_x - x^*_h\| = f(x^*_h)$.

As $h^*_h \leq f(x^*_h)$, we get the first inequality. Also, as $f(x^*_x) \leq C + \varepsilon - \varepsilon \|x^*_x - x^*_h\|$,

$$\|x^*_x - x^*_h\| \leq f(x^*_h)$$

Lastly, as $f(x^*_x + \delta) \geq h^*_h - \|x^*_x - (x^*_x + \delta)\| \geq f(x^*_h) - \|x^*_x - x^*_h\|$

we see that $(\nabla f, \delta) \geq -\|\delta\|$ for any triangle inequality.

**Some remarks:** One can work with functions which are just lower semicontinuous: Sometimes, one can work with discontinuous functions as long as discontinuities are of right kind.

Recall: $f$ is continuous at $x_*$ if for any $\varepsilon > 0$, there is $\delta > 0$: $|f(x) - f(x^*_x)| \leq \varepsilon$ if $\text{dist}(x, x^*_x) \leq \delta$. Lower semicontinuity pays attention only to lower bound in the inequality:

$\boxed{\text{Def } f \text{ is lower semicontinuous to } x \text{ if for any } \varepsilon > 0, \text{ there is } \delta > 0: f(x) \geq f(x^*_x) - \varepsilon \text{ for all } x: \text{dist}(x, x^*_x) \leq \delta}$
So, all of the reasoning above work also for lower semi-continuous functions:

- One can expand the theory beyond \( \mathbb{R}^n \) to \( \infty \)-dim. vector spaces with norm (Banach spaces) or even metric spaces. Won't be really used in this course...

Applications of Ekeland's variational principle.

Our critically important result:

Gordan's lemma: Let \( a_1, ... , a_m \) be vectors in \( \mathbb{R}^n \).

Then one has dichotomy:

- either there exists \( x \in \mathbb{R}^n \): \( (a_k, x) < 0, k=1, ... ,m \),
- or for some \( \lambda_1, ... , \lambda_m, \sum \lambda_k = 1, \lambda_k \geq 0 \) (but not both).

Proof: relies on examining the function

\[
\log \sum_{k=1}^m e^{(a_k, x)} =: f(x).
\]

It is clear that if there is no \( x \): \( (a_k, x) < 0, k=1, ... ,m \), then at least one of \( (a_k, x) \geq 0 \) for any \( x \in \mathbb{R}^n \), and therefore, \( f(x) \) is nonnegative (i.e. bounded from below).
If $f$ is bounded from below, by Ekeland's theorem, for any $\varepsilon > 0$, one can find $x_\varepsilon : \| Df (x_\varepsilon) \| \leq \varepsilon$.

But $Df = \sum a_k e^{(a_k, x)} = \sum a_k \lambda_k (x)_0$, where

$$\lambda_k (x) = \frac{e^{(a_k, x)}}{\sum e^{(a_k, x)}} > 0, \quad \sum \lambda_k (x) = 1.$$  

As such a collection $\{ \lambda_k (x) \}_k = 1 \cdots \infty$ exists for all $\varepsilon > 0$, and the set of $\{ \lambda_k \}_k = 1 \cdots \infty$, $\sum \lambda_k = 1$ is closed and bounded $\Rightarrow$ compact, one can choose a converging subsequence, $\{ \lambda_k (x) \}_k = 1 \cdots \infty$,

$$\lambda (x) \rightarrow \lambda_* = (\lambda_*^1, \ldots, \lambda_*^\infty): \lambda_*^k \geq 0, \quad \sum \lambda_*^k = 1.$$  

Along this subsequence, $\| \sum a_k \lambda_k (x) \| \rightarrow 0$, so that by continuity of $Df$, $\| \sum a_k \lambda_k \| \rightarrow 0$, i.e. $\sum \lambda_*^k a_k = 0$. I.e., if no $x$ satisfies all inequalities $\langle a_k, x \rangle < 0$, then some convex combination of $\lambda_*^k$ vanishes.

Conversely (easy part of Gordon's lemma), if $\sum \lambda_*^k a_k = 0$

$$\lambda_*^k \geq 0, \quad \sum \lambda_*^k = 1,$$

then $0 = \langle 0, x \rangle = \langle \sum \lambda_*^k a_k, x \rangle = \sum \lambda_*^k (a_k, x)$, and one cannot have all terms $(a_k, x) < 0$ simultaneously.  \[\square\]