Lecture 25

Last time

- Ekeland's variational principle
- Inverse Problems.

Today

- Sufficient conditions for metric regularity.
- Algorithmic implications

Sufficient conditions for metric regularity

Lemma (Ioffe '79) Suppose Φ is given by

$$\Phi(x) = \begin{cases} F(x) & x \in S \\ \emptyset & \text{otherwise} \end{cases}$$

where F is continuous and $S \subseteq E$ is a closed set. Suppose that Φ is not metrically regular at $\bar{x} \in S$. Then, there is an arbitrarily small $\delta > 0$, a y close to $\bar{y} = F(\bar{x})$ and x close to $\bar{x}$ minimizing

$$\|F(\cdot) - y\| + \delta \|\cdot - x\| \text{ over } S,$$

but $F(x) \neq y$. $\dashv$

Proof: Since metric regularity doesn't hold, $\exists (x_n)_n \subseteq S$ and $(y_n)_n \subset Y$ s.t.

$x_n \to \bar{x}$, $y_n \to \bar{y}$ and

$$(\text{☾}) \quad \operatorname{dist}(x_n, S \cap F^{-1}(y_n)) = \operatorname{dist}(x_n, \Phi^{-1}(y_n)) > n \, \|F(x_n) - y_n\|.$$

Next, we apply Ekeland's priciple to $f_n(\cdot) = \|F(\cdot) - y_n\| + z_S(\cdot)$.

We take $\varepsilon_n = \|F(x_n) - y_n\|$ and $\delta_n = \max\{\sqrt{\varepsilon_n}, \frac{1}{r}\}$ therefore $\exists\, \bar{x}_n \in S$ that minimizes

$$f_n(\cdot) + \delta_n \|\cdot - x_n\|$$

that is, it minimizes

$\|F(\cdot) - y_n\| + \delta_n \|\cdot - \bar{x}_n\|$ over S.

By construction $\delta_n \to 0$ as $n \to \infty$. Moreover,

$$\|\bar{x}_n - x_n\| \le \varepsilon_n / \delta_n$$

Further, we claim $F(\bar{x}_n) \neq y_n$. Otherwise we would have $\bar{x}_n \in S \cap F^{-1}(y_n)$, and (☾) would give

$$\|\bar{x}_n - x_n\| > n\, \varepsilon_n \ge \frac{\varepsilon_n}{\delta_n},$$

which yields a contradiction. $\square$

Next we will use this lemma to construct conditions under which metric regularity holds.

Theorem (Identifying metric regularity): Suppose that $F: E \to \mathbf{R}$ is Lipschitz, $S \subseteq E$ and

$$\Phi(x) = \begin{cases} \{F(x)\} & \text{if } x \in S, \\ \emptyset & \text{otherwise.} \end{cases}$$

Then, Φ is metrically regular at $\bar{x}$ provided that the following implication holds

$$0 \in \partial\langle w, F(\cdot)\rangle(\bar{x}) + \partial \iota_S(\bar{x}) \Rightarrow w = 0. \quad \dashv$$

Intuition

This condition replaces the one we used for the Inverse Function Theorem. Indeed if F is smooth, this reduces to $\operatorname{Ker} \nabla F(\bar{x}) = \{0\}$.

Proof: Assume searching contradiction that metric regularity fails. Then, there exist sequences $x_n \to \bar{x}$, $y_n \to \bar{y}$

$= F(\bar{x})$, $\delta_n \to 0$ such that x_n minimizes $\|F(\cdot) - y_n\| + \delta_n \|\cdot - x_n\| + z_S(\cdot)$, which implies thanks to the sum rule that

$$0 \in \partial \|F(\cdot) - y_n\|(x_n) + \delta_n B_n + \partial z_S(x_n)$$
$$= \partial \langle w_n, F(\cdot)\rangle (x_n) + \delta_n B_n + \partial z_S(x_n)$$

where $w_n = (F(x_n) - y_n)/\|F(x_n) - y_n\|$ and the second line follows from the next lemma.

Lemma: Suppose $f(x) = h \circ F(x)$ with F Lipschitz at $\bar{x}$ and h smooth around $F(\bar{x})$. Then

$$\partial (h \circ F)(\bar{x}) = \partial(\langle \nabla h(F(\bar{x})), F(\cdot)\rangle)(\bar{x})$$

Proof of the lemma: By assumption near $\bar{x}$

$$h \circ F(x) = h(F(\bar{x})) + \langle \nabla h(F(\bar{x})), F(x) - F(\bar{x})\rangle + o(\|F(x) - F(\bar{x})\|)$$

$$\overset{F \text{ Lipschitz}}{=} h(F(\bar{x})) + \langle \nabla h(F(\bar{x})), F(x) - F(\bar{x})\rangle + o(\|x - \bar{x}\|).$$

Thus, the two functions are equal up to constant terms and $o(\|x - \bar{x}\|)$. A simple computation yields they have the same subdifferentials. $\square$

WLOG, we can assume $w_n \to w$ and so

$$0 \in \partial\langle w, F(\cdot)\rangle(x_n) + \partial\langle w_n - w, F(\cdot)\rangle(x_n) + \delta_n \mathcal{B} + \partial \tau_g(x_n).$$

Note that $\langle w_n - w, F(\cdot)\rangle$ is an L_n-Lipschitz function with $L_n \to 0$ as $n \to \infty$, and further $\delta_n \to 0$. Then, there exist sequences

$$u_n \in \partial\langle w, F(\cdot)\rangle(x_n) \quad \text{and} \quad v_n \in \partial \tau_s(x_n)$$

with $u_n + v_n \to 0$. Once more, we can assume that both sequences converge. Therefore, $u_n \to u \in \partial\langle w, F(\cdot)\rangle(\bar{x})$ and $v_n \to v \in \partial \tau_s(\bar{x})$ (why? Use that $\mathrm{gph}\,\partial \tau_s$ is closed). Therefore, we established

$$0 \in \partial\langle w, F(\cdot)\rangle(\bar{x}) + \partial \tau_s(\bar{x}) \quad \text{and} \quad w \neq 0,$$

which is a contradiction. $\square$

Algorithmic consequences.

The presense of regularity often leads to faster convergence (akin to strong convexity).

Theorem: Suppose that we are interested in finding a solution to $0 \in \Phi(X)$ with $\Phi : E \rightrightarrows E$. Further assume

$$(\heartsuit) \quad \operatorname{dist}(x, \Phi^{-1}(0)) \leq \kappa \operatorname{dist}(\Phi(x), 0)$$

for all $x \in E$, and we have an algorithm that generates a sequence via

$$\mathcal{A} : E \to E$$

$$x_{t+1} \leftarrow \mathcal{A}(x_t)$$

and guarantees

$$(\downarrow) \quad \operatorname{dist}(\Phi(x_t), 0) \leq \mu \frac{\operatorname{dist}(x_0, \Phi^{-1}(0))}{\sqrt{t}}.$$

Then, we have

$$\operatorname{dist}(x_t, \Phi^{-1}(0)) \leq \frac{\mu}{\kappa} e^{-\frac{1}{2}\left(\frac{t-1-\rho}{\rho}\right)} \operatorname{dist}(x_0, \Phi^{-1}(0)),$$

with $\rho = e\sqrt{\mu/\kappa}$. $\dashv$

Remarks

- Condition $(\heartsuit)$ is known as metric subregularity (note that we only vary x and not y).
- The convergence rate $(\downarrow)$ is satisfied by several algorithms (GD,

PPM, PDHG, ADMM*).

- this result boost sublinear convergence of $\mathrm{dist}(\Phi(x_t), 0)$ to linear convergence of $\mathrm{dist}(x_t, \Phi^{-1}(0))$.

Recap

We covered a number of topics

- Convex Analysis
- Duality
- Classical algorithms for LP
  - Simplex
  - Interior Point Methods
- First order Methods
  - KM iteration
  - Examples: PPM, DR, ADMM, PDHG.
- Intro to variational analysis.

Optimization is a very broad field, there are several topics we didn't mention. But hopefully you now have the tools to explore them on your own :)