Lecture 16

Last time

- Fixed points
- Set-valued mappings
- Krasnoselskii-Mann iteration.

Today

- Maximal monotone
- Douglas Rachford
- Augmented Lagrangian

Maximal monotone operator

Recall that given a monotone operator $T : E \rightrightarrows E$ we defined the resolvant as $R := (I + \alpha T)^{-1}$.

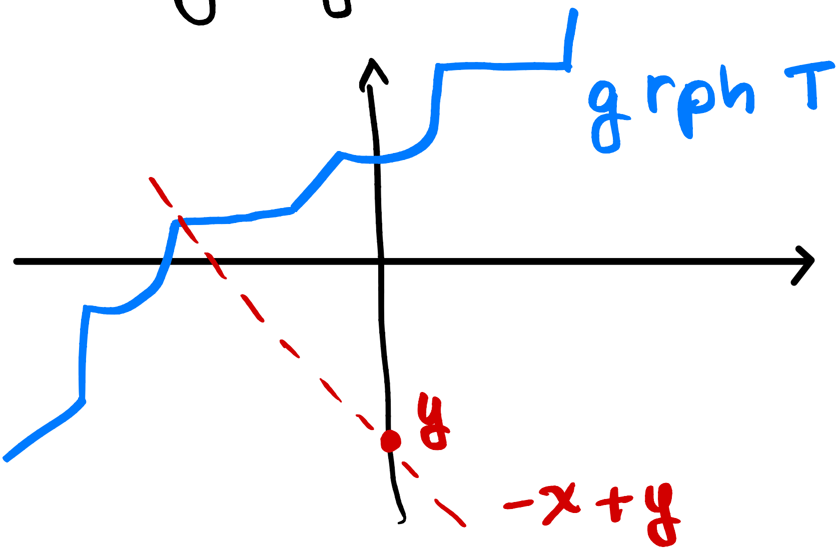
When $T = \partial f$ we showed that R was well-defined.

Q: How do we know $R(x) \neq \emptyset$?

Let's think about this in $E = \mathbb{R}$. Monotonicity in 1D makes the graph

$\operatorname{grph} T = \{(x,y) \mid y \in T(x)\}$

look like a monotonously non-decreasing function (with jumps)



Then, the the following is solvable by

$$y - x \in T(x) \qquad (\because)$$

if, and only if, the graph of $x \mapsto -x+y$ intersects $\operatorname{grph} T$ for all y. Thus, we want T to be as "large" as possible.

Proposition: Let $T : E \rightrightarrows E$ a monotone operator and $\alpha > 0$ be arbitrary. Assume $(\ast)$ is solvable $\forall y$. Then, T is maximal.

i.e., if $T' : E \rightrightarrows$ monotone s.t. $T(x) \subseteq T'(x)$ $\forall x \Rightarrow T = T'$!

Proof: Suppose $y \in T'(x)$, we want to show $y \in T(x)$. Note that $\alpha y \in \alpha T'(x)$ and by assumption

$$(x + \alpha y) - z \in \alpha T(z) \subseteq \alpha T'(z).$$

By monotonicity

$$-\|x - y\|^2 = \langle (x + \alpha y - z) - \alpha y, z - x \rangle \geq 0.$$

Therefore $z = x$, and so
$\alpha y = (x + \alpha y) - x \in \alpha T(x)$.
Proving the result. $\square$

A direct implication of this result is:

Corollary: For a closed, convex, proper $f : E \to \mathbb{R} \cup \{\infty\}$, we have that ∂f is maximal monotone. $\dashv$

In fact the opposite also holds.

Theorem (Minty 1962). Let $T : E \rightrightarrows E$ be a monotone operator and $\alpha > 0$. Then, T is maximal if, and only if, $(:=)$

is solvable for all y. $\dashv$

We will not prove this result. It is usually simpler to show directly that R is well-defined.

Baby steps: Smooth optimization

Consider the problem of minimizing $f: E \to \mathbb{R}$ a smooth convex function with L-Lipschitz gradient. A natural strategy is gradient descent

$$x_{k+1} \leftarrow x_k - \alpha \nabla f(x_k),$$

Theorem: Let f be an L-smooth convex function. Then, we have

$$\langle x-y, \nabla f(x) - \nabla f(y) \rangle \geq \frac{1}{L} \|\nabla f(x) - \nabla f(y)\|^2 \quad (\star)$$
$$\forall x, y$$

Therefore,

$$T_\alpha(x) = x - \alpha \nabla f(x)$$

is non expansive if $\alpha \in [0, 2/L]$ and it is averaged if $\alpha \in (0, 2/L)$.

Thus, if $\alpha \in (0, 2/L)$, gradient descent converges.

Proof: We derived $(\bowtie)$ in Lecture 6 of Nonlinear 1 (see notes).

Then, we can prove nonexpansiveness

$$\|x - \alpha \nabla f(x) - y + \alpha \nabla f(y)\|^2$$

$$= \|x - y\| - 2\alpha \langle x - y, \nabla f(x) - \nabla f(y) \rangle + \alpha^2 \|\nabla f(x) - \nabla f(y)\|^2$$

$$\underset{(\bowtie)}{\le} \|x - y\|^2 - \underbrace{\left(2\frac{\alpha}{L} - \alpha^2\right)}_{\ge 0 \text{ iff } \alpha \in [0, 2/L]} \|\nabla f(x) - \nabla f(y)\|^2$$

Moreover, if $\alpha \in (0, \frac{2}{L})$

$$T_\alpha(x) = \theta x + (1-\theta)\left(x - \frac{\alpha}{(1-\theta)} \nabla f(x)\right)$$

We can pick θ so that $\frac{\alpha}{(1-\theta)} \in [0, 2/L]$.

Convergence follows immediately via the KM iteration result. $\square$

Augmented Lagrangian method

Let us go back to the problem of solving

smooth and convex.

$$\min f(x)$$
$$\text{s.t. } g_i(x) \leq 0 \quad \forall i \in [m].$$

Before presenting the algorithm, let us motivate it. Notice

$$\inf_x \{ f(x) \mid g(x) \leq 0 \}$$
$$= \inf_{\substack{x \\ z \geq 0}} \{ f(x) \mid g(x) + z = 0 \}$$

$$= \inf_{\substack{x \\ z \geq 0}} \left\{ f(x) + \tfrac{1}{2}\|g(x)+z\|^2 \;\middle|\; g(x)+z=0 \right\}$$

$$= \inf_{\substack{x \\ z \geq 0}} \sup_{\lambda} \left\{ f(x) + \tfrac{1}{2}\|g(x)+z\|^2 + \lambda^T(g(x)+z) \right\}$$

$$\geq \sup_{\lambda} \inf_{\substack{x \\ z \geq 0}} \left\{ f(x) + \tfrac{1}{2}\|g(x)+z\|^2 + \lambda^T(g(x)+z) \right\}$$

(why?)

$$= \sup_{\lambda} \inf_{x} \Big\{ \underbrace{f(x) + \tfrac{1}{2}\|(\lambda+g(x))_+\|^2 - \tfrac{1}{2}\|\lambda\|^2}_{\text{Augmented Lagrangian } \bar{L}(x;\lambda)} \Big\}.$$

Recall that we defined, the standard Lagrangian in Lecture 5

$$L(x;z) = f(x) + z^T g(x).$$

Proposition: Suppose $f, g_i : E \to \mathbb{R}$ are convex, and $\lambda \geq 0$, $z = (z+g(x))_+$. Then, x minimizes $\bar{L}(\cdot,\lambda)$ if, and only if, x minimizes $L(\cdot, z)$.

Proof: Exercise. $\square$

This motivates a natural alternating algorithm where we minimize $\bar{L}(\cdot;\lambda)$ and then maximize $\bar{L}(x;\cdot)$. Let's understand how the maximizer of $\bar{L}(x;\cdot)$ looks like via:

Claim: If $h: E \to \mathbb{R}$ convex then so is h_+^2 and $\partial(h_+^2)(x) = 2h_+(x)\,\partial h(x)$. ⊣

For any fixed $x \in E$, the function

$$\lambda_i \longmapsto \frac{1}{2}(\lambda_i + g_i(x))_+^2 - \frac{1}{2}\lambda_i^2$$

is concave and by first order optimality conditions it is maximized at

$$(\lambda_i + g_i(x))_+ - \lambda_i = 0.$$

Thus λ is a fixed point of $T(\lambda) = (\lambda + g(x))_+$. With this we can now introduce the Augmented Lagrangian method.

Augmented Lagrangian Method (ALM)

Pick $x_0 \in E$, $\lambda_0 \in \mathbb{R}^m$.

Loop $k \geq 0$:

$$x_{k+1} \leftarrow \underset{x}{\operatorname{argmin}}\ \bar{L}(x, \lambda_k)$$

$$\lambda_{k+1} \leftarrow (\lambda_k + g(x_{k+1}))_+ .$$

Let's make a couple of remarks:

1) What did we gain? We reduce a constrained problem into a sequence of unconstrained ones. If we have an effective solver for $\bar{L}(\cdot\,; \lambda_k)$, this might be a reasonable algorithm.

2) Notice that in general, x_{k+1} is not well-defined since

the minimizer might not be well defined.

However, if, we assume that $x(\lambda) \leftarrow \operatorname{argmin}_x L(x;\lambda)$ is well-defined for all the λ's we see, then we can analyze this method via the KM iteration. Consider the function

$$\varphi(\lambda) = \begin{cases} -\inf_x L(x,\lambda) & \text{if } y \geq 0, \\ +\infty & \text{otherwise.} \end{cases}$$

Proposition: Suppose that x_{k+1} is well-defined for all λ_k in an execution of ALM. Then,

$$\lambda_{k+1} \leftarrow \operatorname{prox}_\varphi(\lambda_k) \quad \forall k \geq 0.$$

Consequently, λ_{k+1} converges. $\dashv$

The proof is left as an exercise. The final conclusion follows from our KM iteration result since φ is convex. One can also show that any cluster point of the iterates $\{x_k\}$ is a solution to the original problem.

Warning These results are somewhat unsatisfactory since we needed to assume the well-posedness of the primal iterates. Next we will see methods that do not suffer from this issue.