SI152: Numerical Optimization

Lecture 12: Gradient Projection and Frank-Wolfe algorithms

Hao Wang
Email: haw309@gmail.com

ShanghaiTech University

November 14, 2024

Outline

We often want to optimize f within a feasible set Ω:

$$\min f(x), \quad \text{s.t. } x \in \Omega.$$

- $x_* \in \Omega$ is a global solution if

$$f(x_*) \leq f(x) \quad \forall x \in \Omega.$$

- $x_* \in \Omega$ is a local solution if there is a neighborhood $\mathcal{N}$ of x_* s.t.

$$f(x_*) \leq f(x) \quad \forall x \in \mathcal{N} \cap \Omega.$$

- $x_* \in \Omega$ is a strict/strong local solution if there is a neighborhood $\mathcal{N}$ of x_* s.t.

$$f(x_*) < f(x) \quad \forall x \in (\mathcal{N} \cap \Omega) \setminus x_*.$$

Consider a general constrained optimization problem over a closed set Ω:

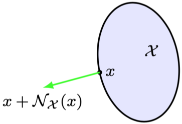
$$\min f(x), \quad \text{s.t. } x \in \Omega.$$

Theorem 1 (Normal Cone)

Given a nonempty convex $\Omega \subset \mathbb{R}^n$ and $x \in \Omega$, the normal cone of Ω at x is

$$\mathcal{N}_\Omega(x) := \{g \mid g^T(\bar{x} - x) \leq 0 \text{ for all } \bar{x} \in \Omega\}.$$

If $x \in \text{int}(\Omega)$, then clearly $\mathcal{N}_\Omega(x) = \{0\}$, but for $x \notin \text{int}(\Omega)$ the normal cone contains at least one halfline.

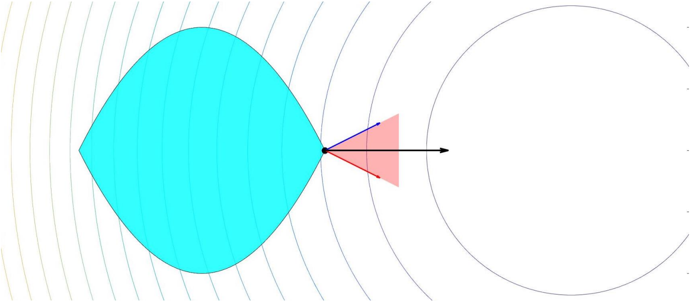


Theorem 2

If x_* is a minimizer of f in Ω, then

$$-\nabla f(x_*) \in \mathcal{N}_\Omega(x_*).$$

That is, if x_* is a minimizer, then the steepest descent direction for f at x_* is in the normal cone of Ω at x_*. (Blue vector denotes $\nabla f(x_*)$.)
In the unconstrained case, $\Omega = \mathbb{R}^n$ and $\mathcal{N}_\Omega(x_*) = \{0\}$.



Definition 3

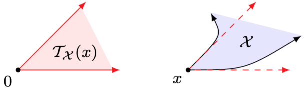
Tangent direction A direction $d \in \mathbb{R}^n$ is tangent to $\Omega \subset \mathbb{R}^n$ at a point $x \in \Omega$ if there exists a sequence of points $\{x_k\} \in \Omega$ and positive scalars $\{\tau_k\}$ such that

$$0 = \lim_{k\to\infty} \tau_k \quad \text{and} \quad d = \lim_{k\to\infty} \frac{1}{\tau_k}(x_k - x).$$

Definition 4 (Tangent cone)

The tangent cone corresponding to a set $\Omega \subset \mathbb{R}^n$ at $x \in \Omega$ is

$$\mathcal{T}_\Omega(x) := \{d \mid d \text{ is tangent to } \Omega \text{ at } x\}.$$



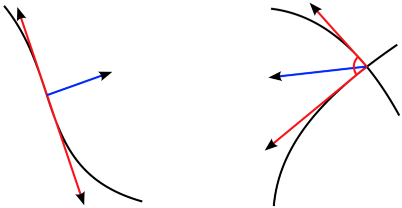
One can verify that for any $\Omega \subset \mathbb{R}^n$ and $x \in \Omega$, the set $\mathcal{T}_\Omega(x)$ is a closed cone.

Theorem 5

If x_* is a minimizer of f in Ω, then

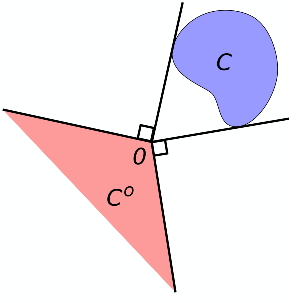
$$\nabla f(x_*)^T d \geq 0, \quad \forall d \in \mathcal{T}_\Omega(x_*).$$

That is, if x_* is a minimizer, then there is no d that is both a descent direction for f at x_* and tangent to Ω at x_*. (Blue vector denotes $\nabla f(x_*)$.)



In the unconstrained case, $\Omega = \mathbb{R}^n$ and $\mathcal{T}_\Omega(x_*) = \mathbb{R}^n$.

For a set $C \subset \mathbb{R}^n$, the polar cone of C is the set $C^\circ = \{y \in \mathbb{R}^n \mid y^T x \leq 0, \forall x \in C\}$



For a convex set Ω, the normal cone $N_\Omega(x)$ is precisely the polar of $T_\Omega(x)$, meaning:

$$N_\Omega(x) = T_\Omega(x)^\circ = \{v \in \mathbb{R}^n \mid v^T d \leq 0 \text{ for all } d \in T_\Omega(x)\}.$$

In this case,

$$\nabla f(x_*)^T d \geq 0, \quad \forall d \in \mathcal{T}_\Omega(x_*) \iff -\nabla f(x_*) \in \mathcal{N}_\Omega(x_*).$$

Outline

Theorem 6

Let $x^0 \in \mathbb{R}^n$ and let $\Omega \subset \mathbb{R}^n$ be a nonempty closed convex set. Then $\bar{x} \in C$ solves the problem

$$\min_x \tfrac{1}{2}\|x - x^0\|_2^2 \quad \textit{s.t. } x \in C$$

if and only if $(\bar{x} - x_0)^T(y - \bar{x}) \geq 0$ for all $y \in C$. Moreover, the solution $\bar{x}$ always exists and is unique.

Proof.

Existence follows from the compactness of the set

$$\{x \in C : \|x - x^0\|_2 \leq \|\hat{x} - x^0\|_2\}$$

where $\hat{x}$ is any element of C. Uniqueness follows from the strong convexity of the 2-norm squared. From the optimality condition $-(\bar{x} - x^0) \in \mathcal{N}_C(x^0)$, meaning

$$(\bar{x} - x^0)^T(y - \bar{x}) \geq 0$$

□

Convex set constrained problem

$$\min_{x} \quad f(x), \quad \text{s.t. } x \in C.$$

- f is $\mathcal{C}^1$
- C is closed convex

Gradient Projection Algorithm:

Set $d^k = P_C(x^k - \nabla f(x^k)) - x^k$

Set λ_k by backtracking Armijo line search

Set $x^{k+1} \leftarrow x^k + \lambda_k d^k$

Proposition 7

Let $x \in C$ *and set* $d = P_C(x - t\nabla f(x)) - x$. *Then*

$$\nabla f(x)^T d \leq -\frac{\|P_C(x - t\nabla f(x)) - x\|^2}{t}.$$

Proof.

Let $z = P_C(x - t\nabla f(x))$. Simply observe that

$$\begin{aligned}
\|P_C(x - t\nabla f(x)) - x\|^2 &= \langle z - x, z - x\rangle \\
&= -t\nabla f(x)^T d + \langle z - (x - t\nabla f(x)), z - x\rangle \\
&\leq -t\nabla f(x)^T d.
\end{aligned}$$

□

Apply the Zoutendijk's result to have

$$\frac{(\nabla f(x^k)^T d^k)^2}{\|d_k\|^2} \to 0$$

Therefore, combining the above Proposition to yield

$$\|d_k\|^2 \to 0$$

Therefore, $P_C(x - \nabla f(x)) - x \to 0$.

Every limit point satisfies

$$x - \nabla f(x) - x \in \mathcal{N}_C(x) \implies -\nabla f(x) \in \mathcal{N}_C(x)$$

Outline

Convex set constrained problem

$$\min_x \quad f(x), \quad \text{s.t. } x \in C.$$

- f is convex and differentiable
- C is closed bounded and convex, e.g., $C = \{x \mid \|x\|_p \leq R\}$

Frank-Wolfe Algorithm:

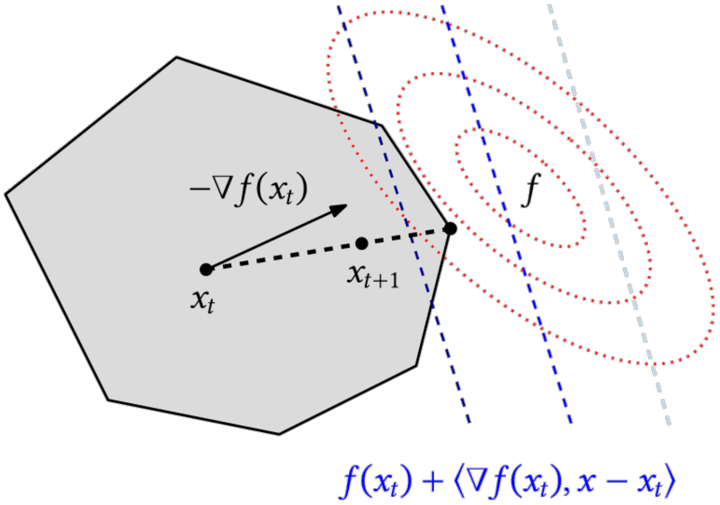
$$s^k \in \arg\min_{s \in C} \nabla f(x^k)^T s$$

$$x^{k+1} \leftarrow (1 - \gamma_k)x^k + \gamma_k s^k$$

Stepsizes: $\gamma_k = 2/(k+2), k = 1, 2, \ldots$ Or, line search like backtracking Armijo.

Note for $\gamma_k \in (0, 1)$, we have $x^{k+1} \in C$ by convexity. Can rewrite update as

$$x^{k+1} \leftarrow x^k + \gamma_k(s^k - x^k)$$


$-\nabla f(x_t)$
f
x_{t+1}
x_t
$f(x_t) + \langle \nabla f(x_t), x - x_t \rangle$

Let L be the L-Lipschitz constant $\geq f(x_0) - f(x_*)$

$$
\begin{aligned}
f(x^{k+1}) &= f((1-\gamma_k)x^k + \gamma_k s^k) \\
&\leq f(x^k) + \gamma_k \langle s^k - x^k, \nabla f(x^k) \rangle + \frac{L}{2}\gamma_k^2 \\
&\leq f(x^k) + \gamma_k \langle x^* - x^k, \nabla f(x^k) \rangle + \frac{L}{2}\gamma_k^2 \\
&= (1-\gamma_k) f(x^k) + \gamma_k (f(x^k) + \langle x^* - x^k, \nabla f(x^k) \rangle) + \frac{L}{2}\gamma_k^2 \\
&\leq (1-\gamma_k) f(x^k) + \gamma_k f(x^*) + \frac{L}{2}\gamma_k^2
\end{aligned}
$$

Hence,

$$
f(x^{k+1}) - f(x^*) \leq (1-\gamma_k)(f(x^k) - f(x^*)) + \frac{L}{2}\gamma_k^2
$$

By choosing $\gamma_k = \frac{2}{k+2}$, it follows from induction that

$$
f(x^k) - f(x^*) \leq (1 - \frac{2}{k+2})\frac{2L}{k+1} + \frac{L}{2}(\frac{2}{k+2})^2 \leq \frac{2L}{k+2}.
$$

Nonconvex cases can combine line search and apply Zoutendjik's results to get

$$\liminf_{k\to\infty} \|\nabla f(x^k)\| = 0$$

Or,

$$\min_{x\in C}\langle \nabla f(x^k), x - x^k\rangle \to 0$$

So that for every limit point:

$$\min_{x\in C}\langle \nabla f(x^*), x - x^*\rangle = 0$$

Or,

$$\langle \nabla f(x^*), x - x^*\rangle \geq 0, \forall x \in C.$$

This means,

$$\nabla f(x^*) \in \mathcal{N}_C(x^*)$$