Conditional gradient methods. Projected Gradient Descent. Frank-Wolfe Method.

Daniil Merkulov

Optimization methods. MIPT

Constrained optimization

Unconstrained optimization

$$\min_{x \in \mathbb{R}^n} f(x)$$

- Any point $x_0 \in \mathbb{R}^n$ is feasible and could be a solution.

Constrained optimization

$$\min_{x \in S} f(x)$$

Gradient Descent is a great way to solve unconstrained problem

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k) \tag{GD}$$

Is it possible to tune GD to fit constrained problem?

Constrained optimization

Unconstrained optimization

$$\min_{x \in \mathbb{R}^n} f(x)$$

- Any point $x_0 \in \mathbb{R}^n$ is feasible and could be a solution.

Constrained optimization

$$\min_{x \in S} f(x)$$

- Not all $x \in \mathbb{R}^n$ is feasible and could be a solution.

Gradient Descent is a great way to solve unconstrained problem

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k) \tag{GD}$$

Is it possible to tune GD to fit constrained problem?

Constrained optimization

Unconstrained optimization

$$\min_{x \in \mathbb{R}^n} f(x)$$

- Any point $x_0 \in \mathbb{R}^n$ is feasible and could be a solution.

Constrained optimization

$$\min_{x \in S} f(x)$$

- Not all $x \in \mathbb{R}^n$ is feasible and could be a solution.
- The solution has to be inside the set S.

Gradient Descent is a great way to solve unconstrained problem

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k) \tag{GD}$$

Is it possible to tune GD to fit constrained problem?

Constrained optimization

Unconstrained optimization

$$
\min_{x \in \mathbb{R}^n} f(x)
$$

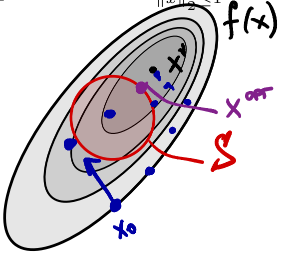
- Any point $x_0 \in \mathbb{R}^n$ is feasible and could be a solution.

Constrained optimization

$$
\min_{x \in S} f(x)
$$

- Not all $x \in \mathbb{R}^n$ is feasible and could be a solution.
- The solution has to be inside the set S.
- Example:

$$
\frac{1}{2}\|Ax - b\|_2^2 \to \min_{\|x\|_2^2 \leq 1}
$$



Gradient Descent is a great way to solve unconstrained problem

$$
x_{k+1} = x_k - \alpha_k \nabla f(x_k) \tag{GD}
$$

Is it possible to tune GD to fit constrained problem?

Constrained optimization

Unconstrained optimization

$$\min_{x \in \mathbb{R}^n} f(x)$$

- Any point $x_0 \in \mathbb{R}^n$ is feasible and could be a solution.

Constrained optimization

$$\min_{x \in S} f(x)$$

- Not all $x \in \mathbb{R}^n$ is feasible and could be a solution.
- The solution has to be inside the set S.
- Example:
$$\frac{1}{2}\|Ax - b\|_2^2 \to \min_{\|x\|_2^2 \leq 1}$$

Gradient Descent is a great way to solve unconstrained problem

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k) \tag{GD}$$

Is it possible to tune GD to fit constrained problem?

Constrained optimization

Unconstrained optimization

$$\min_{x \in \mathbb{R}^n} f(x)$$

- Any point $x_0 \in \mathbb{R}^n$ is feasible and could be a solution.

Constrained optimization

$$\min_{x \in S} f(x)$$

- Not all $x \in \mathbb{R}^n$ is feasible and could be a solution.
- The solution has to be inside the set S.
- Example:

$$\frac{1}{2}\|Ax - b\|_2^2 \to \min_{\|x\|_2^2 \leq 1}$$

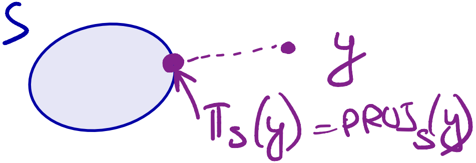
Gradient Descent is a great way to solve unconstrained problem

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k) \tag{GD}$$

Is it possible to tune GD to fit constrained problem?

Yes. We need to use projections to ensure feasibility on every iteration.

Projection



The distance d from point $\mathbf{y} \in \mathbb{R}^n$ to closed set $S \subset \mathbb{R}^n$:

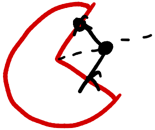
$$d(\mathbf{y}, S, \|\cdot\|) = \inf\{\|x - y\| \mid x \in S\}$$

We will focus on Euclidean projection (other options are possible) of a point $\mathbf{y} \in \mathbb{R}^n$ on set $S \subseteq \mathbb{R}^n$ is a point $\text{proj}_S(\mathbf{y}) \in S$:

$$\boxed{\text{proj}_S(\mathbf{y}) = \frac{1}{2} \underset{\mathbf{x} \in S}{\operatorname{argmin}} \|x - y\|_2^2}$$

- **Sufficient conditions of existence of a projection**. If $S \subseteq \mathbb{R}^n$ - closed set, then the projection on set S exists for any point.

Projection



The distance d from point $\mathbf{y} \in \mathbb{R}^n$ to closed set $S \subset \mathbb{R}^n$:

$$d(\mathbf{y}, S, \|\cdot\|) = \inf\{\|x - y\| \mid x \in S\}$$

We will focus on Euclidean projection (other options are possible) of a point $\mathbf{y} \in \mathbb{R}^n$ on set $S \subseteq \mathbb{R}^n$ is a point $\text{proj}_S(\mathbf{y}) \in S$:

$$\text{proj}_S(\mathbf{y}) = \frac{1}{2} \underset{\mathbf{x} \in S}{\text{argmin}} \|x - y\|_2^2$$

- **Sufficient conditions of existence of a projection**. If $S \subseteq \mathbb{R}^n$ - closed set, then the projection on set S exists for any point.
- **Sufficient conditions of uniqueness of a projection**. If $S \subseteq \mathbb{R}^n$ - closed convex set, then the projection on set S is unique for any point.

Projection

The distance d from point $\mathbf{y} \in \mathbb{R}^n$ to closed set $S \subset \mathbb{R}^n$:

$$d(\mathbf{y}, S, \| \cdot \|) = \inf\{\|x - y\| \mid x \in S\}$$

We will focus on Euclidean projection (other options are possible) of a point $\mathbf{y} \in \mathbb{R}^n$ on set $S \subseteq \mathbb{R}^n$ is a point $\text{proj}_S(\mathbf{y}) \in S$:

$$\text{proj}_S(\mathbf{y}) = \frac{1}{2} \underset{\mathbf{x} \in S}{\text{argmin}} \|x - y\|_2^2$$

- **Sufficient conditions of existence of a projection**. If $S \subseteq \mathbb{R}^n$ - closed set, then the projection on set S exists for any point.
- **Sufficient conditions of uniqueness of a projection**. If $S \subseteq \mathbb{R}^n$ - closed convex set, then the projection on set S is unique for any point.
- If a set is open, and a point is beyond this set, then its projection on this set does not exist.

Projection

The distance d from point $\mathbf{y} \in \mathbb{R}^n$ to closed set $S \subset \mathbb{R}^n$:

$$d(\mathbf{y}, S, \| \cdot \|) = \inf\{\|x - y\| \mid x \in S\}$$

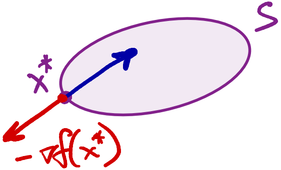
We will focus on Euclidean projection (other options are possible) of a point $\mathbf{y} \in \mathbb{R}^n$ on set $S \subseteq \mathbb{R}^n$ is a point $\text{proj}_S(\mathbf{y}) \in S$:

$$\text{proj}_S(\mathbf{y}) = \frac{1}{2} \underset{\mathbf{x} \in S}{\operatorname{argmin}} \|x - y\|_2^2$$

- **Sufficient conditions of existence of a projection**. If $S \subseteq \mathbb{R}^n$ - closed set, then the projection on set S exists for any point.
- **Sufficient conditions of uniqueness of a projection**. If $S \subseteq \mathbb{R}^n$ - closed convex set, then the projection on set S is unique for any point.
- If a set is open, and a point is beyond this set, then its projection on this set does not exist.
- If a point is in set, then its projection is the point itself.

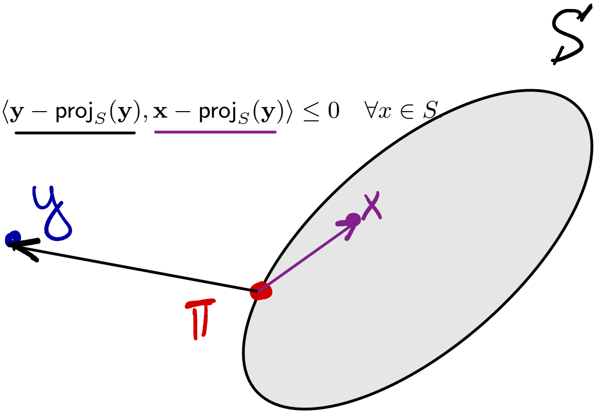
Projection criterion (Bourbaki-Cheney-Goldstein inequality)

$\pi = \mathrm{PROJ}_S(y)$



$$\langle \nabla f(x^*), x - x^* \rangle \geq 0$$

$$\langle \mathbf{y} - \mathsf{proj}_S(\mathbf{y}), \mathbf{x} - \mathsf{proj}_S(\mathbf{y}) \rangle \leq 0 \quad \forall x \in S$$

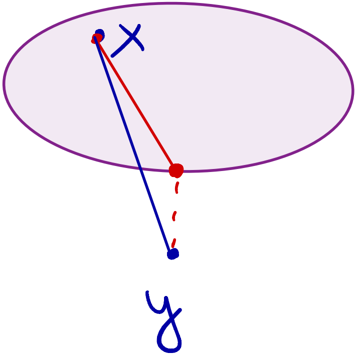


Projection operator is non-expansive

- A function f is called non-expansive if f is L-Lipschitz with $L \leq 1$ [1]. That is, for any two points $x, y \in \text{dom} f$,

$$\|f(x) - f(y)\| \leq L\|x - y\|, \text{ where } L \leq 1.$$

It means the distance between the mapped points is possibly smaller than that of the unmapped points.



[1] Non-expansive becomes contractive if $L < 1$.

Projection operator is non-expansive

- A function f is called non-expansive if f is L-Lipschitz with $L \leq 1$ [1]. That is, for any two points $x, y \in \text{dom} f$,

$$\|f(x) - f(y)\| \leq L\|x - y\|, \text{ where } L \leq 1.$$

It means the distance between the mapped points is possibly smaller than that of the unmapped points.

- Projection operator is non-expansive:

$$\|\text{proj}(x) - \text{proj}(y)\|_2 \leq \|x - y\|_2.$$

[1] Non-expansive becomes contractive if $L < 1$.

Projection operator is non-expansive

S - выпукло и замкнуто

- A function f is called non-expansive if f is L-Lipschitz with $L \leq 1$ [1]. That is, for any two points $x, y \in \text{dom} f$,

$$\|f(x) - f(y)\| \leq L\|x - y\|, \text{ where } L \leq 1.$$

It means the distance between the mapped points is possibly smaller than that of the unmapped points.

- Projection operator is non-expansive:

$$\|\text{proj}(x) - \text{proj}(y)\|_2 \leq \|x - y\|_2.$$

- Next: variational characterization implies non-expansiveness. i.e.,

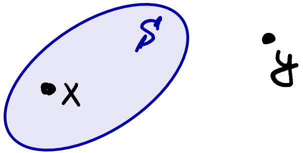
$$\langle y - \text{proj}(y), x - \text{proj}(y)\rangle \leq 0 \quad \forall x \in S \qquad \Rightarrow \qquad \|\text{proj}(x) - \text{proj}(y)\|_2 \leq \|x - y\|_2.$$

[1] Non-expansive becomes contractive if $L < 1$.

Projection operator is non-expansive

Shorthand notation: let $\pi = \text{proj}$ and $\pi(x)$ denotes $\text{proj}(x)$.

Begins with the variational characterization / obtuse angle inequality



$$\langle y - \pi(y), x - \pi(y) \rangle \leq 0 \quad \forall x \in S. \tag{1}$$

Replace x by $\pi(x)$ in Equation 1

$$\langle y - \pi(y), \pi(x) - \pi(y) \rangle \leq 0. \tag{2}$$

Replace y by x and x by $\pi(y)$ in Equation 1

$$\langle x - \pi(x), \pi(y) - \pi(x) \rangle \leq 0. \tag{3}$$

(Equation 2)+(Equation 3) will cancel $\pi(y) - \pi(x)$, not good. So flip the sign of (Equation 3) gives

$$\langle \pi(x) - x, \pi(x) - \pi(y) \rangle \leq 0. \tag{4}$$

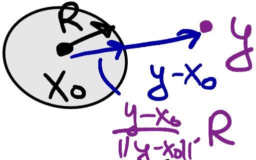
(2)+(4)

$$
\begin{aligned}
\langle y - \pi(y) + \pi(x) - x, \pi(x) - \pi(y) \rangle &\leq 0 \\
\langle y - x + \pi(x) - \pi(y), \pi(x) - \pi(y) \rangle &\leq 0 \\
\langle y - x, \pi(x) - \pi(y) \rangle &\leq -\langle \pi(x) - \pi(y), \pi(x) - \pi(y) \rangle \\
\langle y - x, \pi(y) - \pi(x) \rangle &\geq \|\pi(x) - \pi(y)\|_2^2 \\
\|(y - x)^\top (\pi(y) - \pi(x))\|_2 &\geq \|\pi(x) - \pi(y)\|_2^2
\end{aligned}
$$

By Cauchy-Schwarz inequality, the left-hand-side is upper bounded by $\|y - x\|_2 \|\pi(y) - \pi(x)\|_2$, we get $\|y - x\|_2 \|\pi(y) - \pi(x)\|_2 \geq \|\pi(x) - \pi(y)\|_2^2$. Cancels $\|\pi(x) - \pi(y)\|_2$ finishes the proof.

$$\|\pi(x) - \pi(y)\|_2 \leq \|x - y\|_2$$

Example: projection on the ball



Find $\pi_S(y) = \pi$, if $S = \{x \in \mathbb{R}^n \mid \|x - x_0\| \leq R\}$, $y \notin S$

Build a hypothesis from the figure: $\boxed{\pi = x_0 + R \cdot \frac{y - x_0}{\|y - x_0\|}}$

Check the inequality for a convex closed set: $(\pi - y)^T(x - \pi) \geq 0$

$$\left(x_0 - y + R\frac{y - x_0}{\|y - x_0\|}\right)^T\left(x - x_0 - R\frac{y - x_0}{\|y - x_0\|}\right) =$$

$$\left(\frac{(y - x_0)(R - \|y - x_0\|)}{\|y - x_0\|}\right)^T\left(\frac{(x - x_0)\|y - x_0\| - R(y - x_0)}{\|y - x_0\|}\right) =$$

$$\frac{R - \|y - x_0\|}{\|y - x_0\|^2}(y - x_0)^T\left((x - x_0)\|y - x_0\| - R(y - x_0)\right) =$$

$$\frac{R - \|y - x_0\|}{\|y - x_0\|}\left((y - x_0)^T(x - x_0) - R\|y - x_0\|\right) =$$

$$(R - \|y - x_0\|)\left(\frac{(y - x_0)^T(x - x_0)}{\|y - x_0\|} - R\right)$$

The first factor is negative for point selection y. The second factor is also negative, which follows from the Cauchy-Bunyakovsky inequality:

$$(y - x_0)^T(x - x_0) \leq \|y - x_0\|\|x - x_0\|$$

$$\frac{(y - x_0)^T(x - x_0)}{\|y - x_0\|} - R \leq \frac{\|y - x_0\|\|x - x_0\|}{\|y - x_0\|} -$$

Example: projection on the halfspace

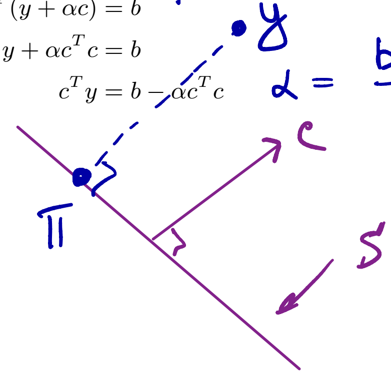
$$\pi = y + \frac{b - c^T y}{c^T c} c$$

Find $\pi_S(y) = \pi$, if $S = \{x \in \mathbb{R}^n \mid c^T x = b\}$, $y \notin S$. Build a hypothesis from the figure: $\pi = y + \alpha c$. Coefficient α is chosen so that $\pi \in S$: $c^T \pi = b$, so:

$$c^T(y + \alpha c) = b$$

$$c^T y + \alpha c^T c = b$$

$$c^T y = b - \alpha c^T c$$



$$\alpha = \frac{b - c^T y}{c^T c}$$

Check the inequality for a convex closed set: $(\pi - y)^T(x - \pi) \geq 0$

$$(y + \alpha c - y)^T(x - y - \alpha c) =$$

$$\alpha c^T(x - y - \alpha c) =$$

$$\alpha(c^T x) - \alpha(c^T y) - \alpha^2(c^T c) =$$

$$\alpha b - \alpha(b - \alpha c^T c) - \alpha^2 c^T c =$$

$$\alpha b - \alpha b + \alpha^2 c^T c - \alpha^2 c^T c = 0 \geq 0$$

Convergence rate for smooth and convex case

$$x^{k+1} = \text{PROJ}_S\left(x^k - \alpha^k \nabla f(x^k)\right)$$

$$\|x^{k+1} - x^*\|^2 = \|x^k - \alpha^k \nabla f(x^k) - x^*\|^2 \le$$

$$= \|x^k - x^*\|^2 + (\alpha^k)^2 \cdot \|\nabla f(x^k)\|^2 - 2\alpha^k \nabla f(x^k)^T (x^k - x^*)$$

$$\|x^{k+1} - x^*\|^2 = \left\|\text{PROJ}_S\left(x^k - \alpha^k \nabla f(x^k)\right) - \text{PROJ}_S(x^*)\right\|_2^2 \le$$

$$\le \left\| x^k - \alpha^k \nabla f(x^k) - x^* \right\|_2^2$$

Idea

Simplex PROJection

$$S = \{ 1^T x = 1, \ x \geq 0 \}$$

НЕТ формулы

~~$Ax = b$~~

есть алгоритмы $O(n)$

Convergence

Сх-ть метода проекции (суб)градиента

$$x^{k+1} = \text{PROJ}_S(x^k - \alpha^k g^k)$$

ТОЧНО ТАКАЯ ЖЕ, как и для

GD суб $\frac{1}{k}$ $\mu = 0$
лин. $\mu > 0$

SD суб $\frac{1}{\sqrt{k}}$

стоимость итерации может быть сильно дороже

Comparison to PGD

Алгоритм Франк-Вульфа

давайте на каждом шаге искать условный минимум $f^I_{x_k}(x)$

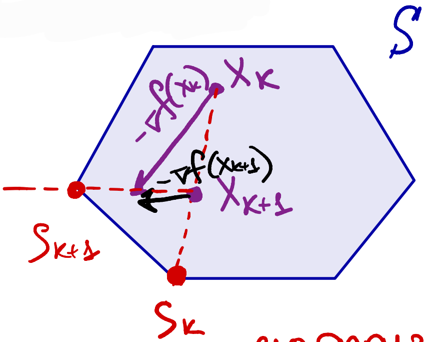
$$f^I_{x_k}(x) = f(x_k) + \langle \nabla f(x_k), x - x_k \rangle$$

$$\rightarrow \quad S_k = \arg\min_{x \in S} f^I_{x_k}(x) = \arg\min_{x \in S} \langle \nabla f(x_k), x \rangle$$

FW:

$$\boxed{x_k = \gamma^k x_{k-1} + (1 - \gamma^k) S_k}$$

$\gamma \in (0;1)$



что по сх-ти ?
для выпуклого гладкого случая

$O\left(\frac{1}{k}\right)$

сюрприз 1: в сильно выпуклом случае $O\left(\frac{1}{k}\right)$

сюрприз 2: не работает в негладком случае (работают модификации)

Comparison to PGD

- более детально вывести сх-ть PGD для выпуклого гладк. сл.
- вывести сх-ть FW для выпуклого гладкого сл.

Comparison to PGD

$$f(x^k) - f^* = \varepsilon$$

$$O\left(\frac{1}{k}\right)$$

$$\frac{1}{k} = \varepsilon \to$$

$$O\left(\frac{1}{\varepsilon}\right)$$

Comparison to PGD

$$t := \beta t$$

$$\alpha = 1$$

$$f(x + tv) \leq f(x) + \alpha t \nabla f(x)^T v$$