

Conditional gradient methods. Projected Gradient Descent. Frank-Wolfe Method. Proximal method. Idea of Mirror Descent

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

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

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

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

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

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

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:

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

Gradient Descent is a great way to solve unconstrained problem

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

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:

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

Gradient Descent is a great way to solve unconstrained problem

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

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:

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

Gradient Descent is a great way to solve unconstrained problem

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

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 $\text{proj}_S(\mathbf{y}) \in S$:

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

- **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 $S \subset \mathbb{R}^n$:

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

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 $\text{proj}_S(\mathbf{y}) \in S$:

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

- **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)

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

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.

¹Non-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:

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

¹Non-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:

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

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

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

¹Non-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

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

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

$$\langle y - \pi(y), \pi(x) - \pi(y) \rangle \leq 0. \quad (2)$$

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

$$\langle x - \pi(x), \pi(y) - \pi(x) \rangle \leq 0. \quad (3)$$

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

$$\langle \pi(x) - x, \pi(x) - \pi(y) \rangle \leq 0. \quad (4)$$

$$\langle y - \pi(y) + \pi(x) - x, \pi(x) - \pi(y) \rangle \leq 0$$

$$\langle y - x + \pi(x) - \pi(y), \pi(x) - \pi(y) \rangle \leq 0$$

$$\langle y - x, \pi(x) - \pi(y) \rangle \leq -\langle \pi(x) - \pi(y), \pi(x) - \pi(y) \rangle$$

$$\langle y - x, \pi(y) - \pi(x) \rangle \geq \|\pi(x) - \pi(y)\|_2^2$$

$$\|(y - x)^\top (\pi(y) - \pi(x))\|_2 \geq \|\pi(x) - \pi(y)\|_2^2$$

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.

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: $\pi = x_0 + R \cdot \frac{y - x_0}{\|y - x_0\|}$

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

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

$$\begin{aligned} \left(x_0 - y + R \frac{y - x_0}{\|y - x_0\|} \right)^T \left(x - x_0 - R \frac{y - x_0}{\|y - x_0\|} \right) &= \\ \left(\frac{(y - x_0)(R - \|y - x_0\|)}{\|y - x_0\|} \right)^T \left(\frac{(x - x_0)\|y - x_0\| - R(y - x_0)}{\|y - x_0\|} \right) &= (y - x_0)^T(x - x_0) \leq \|y - x_0\|\|x - x_0\| \\ \frac{R - \|y - x_0\|}{\|y - x_0\|^2} (y - x_0)^T ((x - x_0)\|y - x_0\| - R(y - x_0)) &= \frac{(y - x_0)^T(x - x_0)}{\|y - x_0\|} - R \leq \frac{\|y - x_0\|\|x - x_0\|}{\|y - x_0\|} - R \\ \frac{R - \|y - x_0\|}{\|y - x_0\|} ((y - x_0)^T(x - x_0) - R\|y - x_0\|) &= \\ (R - \|y - x_0\|) \left(\frac{(y - x_0)^T(x - x_0)}{\|y - x_0\|} - R \right) & \end{aligned}$$

Example: projection on the halfspace

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:

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

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

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

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

$$\begin{aligned}(y + \alpha c - y)^T(x - y - \alpha c) &= \\ \alpha c^T(x - y - \alpha c) &= \\ \alpha(c^T x) - \alpha(c^T y) - \alpha^2(c^T c) &= \\ \alpha b - \alpha(b - \alpha c^T c) - \alpha^2 c^T c &= \\ \alpha b - \alpha b + \alpha^2 c^T c - \alpha^2 c^T c &= 0 \geq 0\end{aligned}$$

Convergence rate for smooth and convex case

Idea

Convergence

Comparison to PGD

Proximal operator

Convergence

Idea

Notations

Convergence