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

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

- **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 $\mathsf{proj}_S(\mathbf{y}) \in S$:

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

- **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 $\mathsf{proj}_S(\mathbf{y}) \in S$:

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

- **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 $\mathsf{proj}_S(\mathbf{y}) \in S$:

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

- **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)

Theorem

Let $S \subseteq \mathbb{R}^n$ be closed and convex, $\forall x \in S, y \in \mathbb{R}^n$. Then

$$\langle y - \mathsf{proj}_S(y), \mathbf{x} - \mathsf{proj}_S(y) \rangle \leq 0 \qquad (1)$$

$$\|x - \mathsf{proj}_S(y)\|^2 + \|y - \mathsf{proj}_S(y)\|^2 \leq \|x - y\|^2 \qquad (2)$$

Proof

1. $\mathsf{proj}_S(y)$ is minimizer of differentiable convex function
   $d(y, S, \|\cdot\|) = \|x - y\|^2$ over S. By first-order characterization of
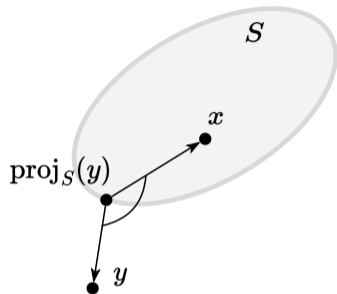   optimality.

Figure 1: Obtuse or straight angle should be for any point $x \in S$

Projection criterion (Bourbaki-Cheney-Goldstein inequality)

Theorem

Let $S \subseteq \mathbb{R}^n$ be closed and convex, $\forall x \in S, y \in \mathbb{R}^n$. Then

$$\langle y - \text{proj}_S(y), \mathbf{x} - \text{proj}_S(y) \rangle \leq 0 \qquad (1)$$

$$\|x - \text{proj}_S(y)\|^2 + \|y - \text{proj}_S(y)\|^2 \leq \|x - y\|^2 \qquad (2)$$

Proof

1. $\text{proj}_S(y)$ is minimizer of differentiable convex function
   $d(y, S, \|\cdot\|) = \|x - y\|^2$ over S. By first-order characterization of
   optimality.

$$\nabla d(\text{proj}_S(y))^T (x - \text{proj}_S(y)) \geq 0$$
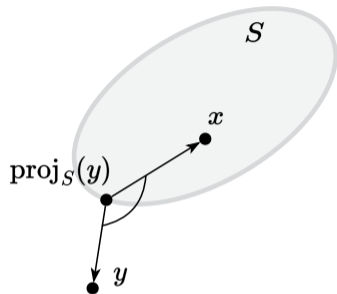

Figure 1: Obtuse or straight angle should be
for any point $x \in S$

Projection criterion (Bourbaki-Cheney-Goldstein inequality)

Theorem

Let $S \subseteq \mathbb{R}^n$ be closed and convex, $\forall x \in S, y \in \mathbb{R}^n$. Then

$$\langle y - \mathsf{proj}_S(y), \mathbf{x} - \mathsf{proj}_S(y) \rangle \leq 0 \qquad (1)$$

$$\|x - \mathsf{proj}_S(y)\|^2 + \|y - \mathsf{proj}_S(y)\|^2 \leq \|x - y\|^2 \qquad (2)$$

Proof

1. $\mathsf{proj}_S(y)$ is minimizer of differentiable convex function
   $d(y, S, \|\cdot\|) = \|x - y\|^2$ over S. By first-order characterization of optimality.

$$\nabla d(\mathsf{proj}_S(y))^T (x - \mathsf{proj}_S(y)) \geq 0$$

$$2 \left(\mathsf{proj}_S(y) - y\right)^T \left(x - \mathsf{proj}_S(y)\right) \geq 0$$
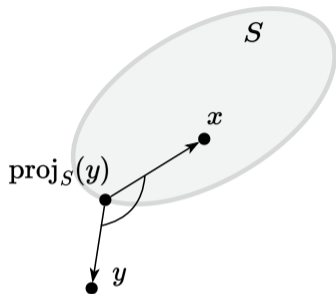

Figure 1: Obtuse or straight angle should be for any point $x \in S$

Projection criterion (Bourbaki-Cheney-Goldstein inequality)

Theorem

Let $S \subseteq \mathbb{R}^n$ be closed and convex, $\forall x \in S, y \in \mathbb{R}^n$. Then

$$\langle y - \mathsf{proj}_S(y), \mathbf{x} - \mathsf{proj}_S(y) \rangle \leq 0 \qquad (1)$$

$$\|x - \mathsf{proj}_S(y)\|^2 + \|y - \mathsf{proj}_S(y)\|^2 \leq \|x - y\|^2 \qquad (2)$$

Proof

1. $\mathsf{proj}_S(y)$ is minimizer of differentiable convex function
   $d(y, S, \|\cdot\|) = \|x - y\|^2$ over S. By first-order characterization of
   optimality.

$$\nabla d(\mathsf{proj}_S(y))^T (x - \mathsf{proj}_S(y)) \geq 0$$

$$2 \left(\mathsf{proj}_S(y) - y \right)^T \left(x - \mathsf{proj}_S(y) \right) \geq 0$$

$$\left(y - \mathsf{proj}_S(y) \right)^T \left(x - \mathsf{proj}_S(y) \right) \leq 0$$
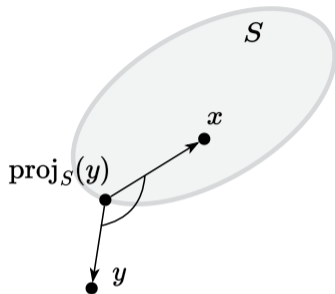

Figure 1: Obtuse or straight angle should be for any point $x \in S$

Projection criterion (Bourbaki-Cheney-Goldstein inequality)

Theorem

Let $S \subseteq \mathbb{R}^n$ be closed and convex, $\forall x \in S, y \in \mathbb{R}^n$. Then

$$\langle y - \mathsf{proj}_S(y), \mathbf{x} - \mathsf{proj}_S(y) \rangle \leq 0 \qquad (1)$$

$$\|x - \mathsf{proj}_S(y)\|^2 + \|y - \mathsf{proj}_S(y)\|^2 \leq \|x - y\|^2 \qquad (2)$$

Proof

1. $\mathsf{proj}_S(y)$ is minimizer of differentiable convex function $d(y, S, \|\cdot\|) = \|x - y\|^2$ over S. By first-order characterization of optimality.

$$\nabla d(\mathsf{proj}_S(y))^T (x - \mathsf{proj}_S(y)) \geq 0$$

$$2 \left(\mathsf{proj}_S(y) - y \right)^T \left(x - \mathsf{proj}_S(y) \right) \geq 0$$

$$\left(y - \mathsf{proj}_S(y) \right)^T \left(x - \mathsf{proj}_S(y) \right) \leq 0$$

2. Use cosine rule $2x^T y = \|x\|^2 + \|y\|^2 - \|x - y\|^2$ with $x = x - \mathsf{proj}_S(y)$ and $y = y - \mathsf{proj}_S(y)$. By the first property of the theorem:
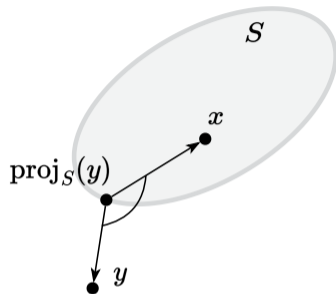

Figure 1: Obtuse or straight angle should be for any point $x \in S$

Projection criterion (Bourbaki-Cheney-Goldstein inequality)

Theorem

Let $S \subseteq \mathbb{R}^n$ be closed and convex, $\forall x \in S, y \in \mathbb{R}^n$. Then

$$\langle y - \mathsf{proj}_S(y), \mathbf{x} - \mathsf{proj}_S(y) \rangle \leq 0 \qquad (1)$$

$$\|x - \mathsf{proj}_S(y)\|^2 + \|y - \mathsf{proj}_S(y)\|^2 \leq \|x - y\|^2 \qquad (2)$$

Proof

1. $\mathsf{proj}_S(y)$ is minimizer of differentiable convex function
   $d(y, S, \|\cdot\|) = \|x - y\|^2$ over S. By first-order characterization of optimality.

   $$\nabla d(\mathsf{proj}_S(y))^T (x - \mathsf{proj}_S(y)) \geq 0$$

   $$2 \left(\mathsf{proj}_S(y) - y \right)^T (x - \mathsf{proj}_S(y)) \geq 0$$

   $$\left(y - \mathsf{proj}_S(y) \right)^T (x - \mathsf{proj}_S(y)) \leq 0$$

2. Use cosine rule $2x^T y = \|x\|^2 + \|y\|^2 - \|x - y\|^2$ with
   $x = x - \mathsf{proj}_S(y)$ and $y = y - \mathsf{proj}_S(y)$. By the first property of the theorem:

   $$0 \geq 2x^T y = \|x - \mathsf{proj}_S(y)\|^2 + \|y + \mathsf{proj}_S(y)\|^2 - \|x - y\|^2$$

Figure 1: Obtuse or straight angle should be for any point $x \in S$

Projection criterion (Bourbaki-Cheney-Goldstein inequality)

Theorem

Let $S \subseteq \mathbb{R}^n$ be closed and convex, $\forall x \in S, y \in \mathbb{R}^n$. Then

$$\langle y - \text{proj}_S(y), \mathbf{x} - \text{proj}_S(y) \rangle \leq 0 \qquad (1)$$

$$\|x - \text{proj}_S(y)\|^2 + \|y - \text{proj}_S(y)\|^2 \leq \|x - y\|^2 \qquad (2)$$

Proof

1. $\text{proj}_S(y)$ is minimizer of differentiable convex function
   $d(y, S, \|\cdot\|) = \|x - y\|^2$ over S. By first-order characterization of optimality.

   $$\nabla d(\text{proj}_S(y))^T (x - \text{proj}_S(y)) \geq 0$$

   $$2\left(\text{proj}_S(y) - y\right)^T (x - \text{proj}_S(y)) \geq 0$$

   $$(y - \text{proj}_S(y))^T (x - \text{proj}_S(y)) \leq 0$$

2. Use cosine rule $2x^T y = \|x\|^2 + \|y\|^2 - \|x - y\|^2$ with
   $x = x - \text{proj}_S(y)$ and $y = y - \text{proj}_S(y)$. By the first property of the theorem:

   $$0 \geq 2x^T y = \|x - \text{proj}_S(y)\|^2 + \|y + \text{proj}_S(y)\|^2 - \|x - y\|^2$$

   $$\|x - \text{proj}_S(y)\|^2 + \|y + \text{proj}_S(y)\|^2 \leq \|x - y\|^2$$
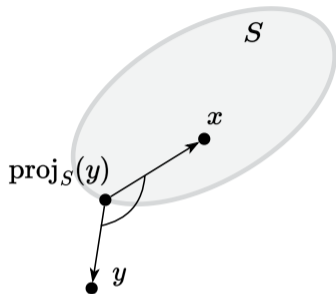

Figure 1: Obtuse or straight angle should be for any point $x \in S$

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 $x, y \in \mathrm{dom} f$,

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

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

- Projection operator is non-expansive:

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

---

[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:

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

- 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 = $ proj and $\pi(x)$ denotes proj(x).

Begins with the variational characterization / obtuse angle inequality

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

Replace x by $\pi(x)$ in Equation 3

Replace y by x and x by $\pi(y)$ in Equation 3

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

$$\langle x - \pi(x), \pi(y) - \pi(x) \rangle \leq 0. \tag{5}$$

(Equation 4)+(Equation 5) will cancel $\pi(y) - \pi(x)$, not good. So flip the sign of (Equation 5) gives

$$\langle \pi(x) - x, \pi(x) - \pi(y) \rangle \leq 0. \tag{6}$$

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

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

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

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

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

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

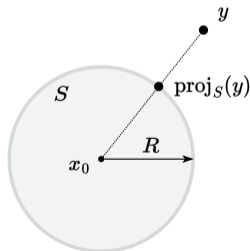Find $\pi_S(y) = \pi$, if $S = \{x \in \mathbb{R}^n \mid \|x - x_0\| \leq R\}$, $y \notin S$

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: $(\pi - y)^T (x - \pi) \geq 0$

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

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

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

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

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

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

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:
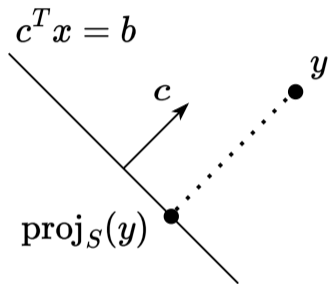

$$c^T x = b$$

Figure 3: Hyperplane

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

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

$$(y + \alpha c - y)^T (x - y - \alpha c) =$$
$$\alpha c^T (x - y - \alpha c) =$$
$$\alpha(c^T x) - \alpha(c^T y) - \alpha^2(c^T c) =$$
$$\alpha b - \alpha(b - \alpha c^T c) - \alpha^2 c^T c =$$
$$\alpha b - \alpha b + \alpha^2 c^T c - \alpha^2 c^T c = 0 \geq 0$$

Idea

$$x_{k+1} = \mathsf{proj}_S \left(x_k - \alpha_k \nabla f(x_k) \right) \qquad \Leftrightarrow \qquad \begin{aligned} y_k &= x_k - \alpha_k \nabla f(x_k) \\ x_{k+1} &= \mathsf{proj}_S \left(y_k \right) \end{aligned}$$
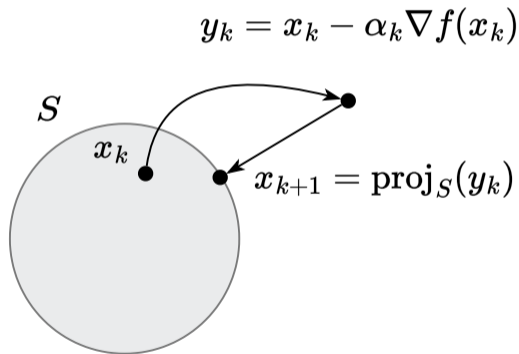

Figure 4: Illustration of Projected Gradient Descent algorithm

Convergence rate for smooth and convex case

Theorem

Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex and differentiable. Let $S \subseteq \mathbb{R}^n$d be a closed convex set, and assume that there is a minimizer x^* of f over S; furthermore, suppose that f is smooth over S with parameter L. The Projected Gradient Descent algorithm with stepsize $\frac{1}{L}$ achieves the following convergence after iteration $k > 0$:

$$f(x_k) - f^* \leq \frac{L\|x_0 - x^*\|_2^2}{2k}$$

Convergence rate for smooth and convex case

Theorem

Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex and differentiable. Let $S \subseteq \mathbb{R}^n$d be a closed convex set, and assume that there is a minimizer x^* of f over S; furthermore, suppose that f is smooth over S with parameter L. The Projected Gradient Descent algorithm with stepsize $\frac{1}{L}$ achieves the following convergence after iteration $k > 0$:

$$f(x_k) - f^* \leq \frac{L\|x_0 - x^*\|_2^2}{2k}$$

Proof

1. Let's prove sufficient decrease lemma, assuming, that $y_k = x_k - \frac{1}{L}\nabla f(x_k)$ and cosine rule
   $2x^T y = \|x\|^2 + \|y\|^2 - \|x - y\|^2$:

(7)

Convergence rate for smooth and convex case

Theorem

Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex and differentiable. Let $S \subseteq \mathbb{R}^n$d be a closed convex set, and assume that there is a minimizer x^* of f over S; furthermore, suppose that f is smooth over S with parameter L. The Projected Gradient Descent algorithm with stepsize $\frac{1}{L}$ achieves the following convergence after iteration $k > 0$:

$$f(x_k) - f^* \leq \frac{L\|x_0 - x^*\|_2^2}{2k}$$

Proof

1. Let's prove sufficient decrease lemma, assuming, that $y_k = x_k - \frac{1}{L}\nabla f(x_k)$ and cosine rule
   $2x^T y = \|x\|^2 + \|y\|^2 - \|x - y\|^2$:

   Smoothness: $\quad f(x_{k+1}) \leq f(x_k) + \langle \nabla f(x_k), x_{k+1} - x_k \rangle + \frac{L}{2}\|x_{k+1} - x_k\|^2$

(7)

Convergence rate for smooth and convex case

> **Theorem**
>
> Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex and differentiable. Let $S \subseteq \mathbb{R}^n$d be a closed convex set, and assume that there is a minimizer x^* of f over S; furthermore, suppose that f is smooth over S with parameter L. The Projected Gradient Descent algorithm with stepsize $\frac{1}{L}$ achieves the following convergence after iteration $k > 0$:
>
> $$f(x_k) - f^* \leq \frac{L\|x_0 - x^*\|_2^2}{2k}$$

Proof

1. Let's prove sufficient decrease lemma, assuming, that $y_k = x_k - \frac{1}{L}\nabla f(x_k)$ and cosine rule $2x^T y = \|x\|^2 + \|y\|^2 - \|x - y\|^2$:

    Smoothness: $\quad f(x_{k+1}) \leq f(x_k) + \langle \nabla f(x_k), x_{k+1} - x_k \rangle + \frac{L}{2}\|x_{k+1} - x_k\|^2$

    Method: $\qquad\qquad = f(x_k) - L\langle y_k - x_k, x_{k+1} - x_k \rangle + \frac{L}{2}\|x_{k+1} - x_k\|^2$

(7)

Convergence rate for smooth and convex case

Theorem

Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex and differentiable. Let $S \subseteq \mathbb{R}^n$d be a closed convex set, and assume that there is a minimizer x^* of f over S; furthermore, suppose that f is smooth over S with parameter L. The Projected Gradient Descent algorithm with stepsize $\frac{1}{L}$ achieves the following convergence after iteration $k > 0$:

$$f(x_k) - f^* \leq \frac{L\|x_0 - x^*\|_2^2}{2k}$$

Proof

1. Let's prove sufficient decrease lemma, assuming, that $y_k = x_k - \frac{1}{L}\nabla f(x_k)$ and cosine rule
   $2x^T y = \|x\|^2 + \|y\|^2 - \|x - y\|^2$:

   Smoothness: $\quad f(x_{k+1}) \leq f(x_k) + \langle \nabla f(x_k), x_{k+1} - x_k \rangle + \frac{L}{2}\|x_{k+1} - x_k\|^2$

   Method: $\qquad\qquad = f(x_k) - L\langle y_k - x_k, x_{k+1} - x_k \rangle + \frac{L}{2}\|x_{k+1} - x_k\|^2$

   Cosine rule: $\qquad = f(x_k) - \frac{L}{2}\left(\|y_k - x_k\|^2 + \|x_{k+1} - x_k\|^2 - \|y_k - x_{k+1}\|^2\right) + \frac{L}{2}\|x_{k+1} - x_k\|^2$ (7)

Convergence rate for smooth and convex case

> Theorem
>
> Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex and differentiable. Let $S \subseteq \mathbb{R}^n$d be a closed convex set, and assume that there is a minimizer x^* of f over S; furthermore, suppose that f is smooth over S with parameter L. The Projected Gradient Descent algorithm with stepsize $\frac{1}{L}$ achieves the following convergence after iteration $k > 0$:
>
> $$f(x_k) - f^* \leq \frac{L\|x_0 - x^*\|_2^2}{2k}$$

Proof

1. Let's prove sufficient decrease lemma, assuming, that $y_k = x_k - \frac{1}{L}\nabla f(x_k)$ and cosine rule
   $2x^T y = \|x\|^2 + \|y\|^2 - \|x - y\|^2$:

   Smoothness: $\quad f(x_{k+1}) \leq f(x_k) + \langle \nabla f(x_k), x_{k+1} - x_k \rangle + \frac{L}{2}\|x_{k+1} - x_k\|^2$

   Method: $\qquad\qquad = f(x_k) - L\langle y_k - x_k, x_{k+1} - x_k \rangle + \frac{L}{2}\|x_{k+1} - x_k\|^2$

   Cosine rule: $\qquad = f(x_k) - \frac{L}{2}\left(\|y_k - x_k\|^2 + \|x_{k+1} - x_k\|^2 - \|y_k - x_{k+1}\|^2\right) + \frac{L}{2}\|x_{k+1} - x_k\|^2 \quad (7)$

   $\qquad\qquad\qquad = f(x_k) - \frac{1}{2L}\|\nabla f(x_k)\|^2 + \frac{L}{2}\|y_k - x_{k+1}\|^2$

Convergence rate for smooth and convex case (proof)

2. Now we do not immediately have progress at each step. Let's use again cosine rule:

$$\left\langle \frac{1}{L}\nabla f(x_k), x_k - x^* \right\rangle = \frac{1}{2}\left(\frac{1}{L^2}\|\nabla f(x_k)\|^2 + \|x_k - x^*\|^2 - \|x_k - x^* - \frac{1}{L}\nabla f(x_k)\|^2 \right)$$

$$\langle \nabla f(x_k), x_k - x^* \rangle = \frac{L}{2}\left(\frac{1}{L^2}\|\nabla f(x_k)\|^2 + \|x_k - x^*\|^2 - \|y_k - x^*\|^2 \right)$$

Convergence rate for smooth and convex case (proof)

2. Now we do not immediately have progress at each step. Let's use again cosine rule:

$$\left\langle \frac{1}{L}\nabla f(x_k), x_k - x^* \right\rangle = \frac{1}{2}\left(\frac{1}{L^2}\|\nabla f(x_k)\|^2 + \|x_k - x^*\|^2 - \|x_k - x^* - \frac{1}{L}\nabla f(x_k)\|^2\right)$$

$$\langle \nabla f(x_k), x_k - x^* \rangle = \frac{L}{2}\left(\frac{1}{L^2}\|\nabla f(x_k)\|^2 + \|x_k - x^*\|^2 - \|y_k - x^*\|^2\right)$$

3. We will use now projection property: $\|x - \mathsf{proj}_S(y)\|^2 + \|y - \mathsf{proj}_S(y)\|^2 \leq \|x - y\|^2$ with $x = x^*, y = y_k$:

$$\|x^* - \mathsf{proj}_S(y_k)\|^2 + \|y_k - \mathsf{proj}_S(y_k)\|^2 \leq \|x^* - y_k\|^2$$

$$\|y_k - x^*\|^2 \geq \|x^* - x_{k+1}\|^2 + \|y_k - x_{k+1}\|^2$$

Convergence rate for smooth and convex case (proof)

2. Now we do not immediately have progress at each step. Let's use again cosine rule:

$$\left\langle \frac{1}{L}\nabla f(x_k), x_k - x^* \right\rangle = \frac{1}{2}\left(\frac{1}{L^2}\|\nabla f(x_k)\|^2 + \|x_k - x^*\|^2 - \left\|x_k - x^* - \frac{1}{L}\nabla f(x_k)\right\|^2 \right)$$

$$\langle \nabla f(x_k), x_k - x^* \rangle = \frac{L}{2}\left(\frac{1}{L^2}\|\nabla f(x_k)\|^2 + \|x_k - x^*\|^2 - \|y_k - x^*\|^2 \right)$$

3. We will use now projection property: $\|x - \mathsf{proj}_S(y)\|^2 + \|y - \mathsf{proj}_S(y)\|^2 \le \|x - y\|^2$ with $x = x^*, y = y_k$:

$$\|x^* - \mathsf{proj}_S(y_k)\|^2 + \|y_k - \mathsf{proj}_S(y_k)\|^2 \le \|x^* - y_k\|^2$$

$$\|y_k - x^*\|^2 \ge \|x^* - x_{k+1}\|^2 + \|y_k - x_{k+1}\|^2$$

4. Now, using convexity and previous part:

Convexity:
$$f(x_k) - f^* \le \langle \nabla f(x_k), x_k - x^* \rangle$$
$$\le \frac{L}{2}\left(\frac{1}{L^2}\|\nabla f(x_k)\|^2 + \|x_k - x^*\|^2 - \|x_{k+1} - x^*\|^2 - \|y_k - x_{k+1}\|^2 \right)$$

Sum for $i = 0, k-1$
$$\sum_{i=0}^{k-1}[f(x_i) - f^*] \le \sum_{i=0}^{k-1}\frac{1}{2L}\|\nabla f(x_i)\|^2 + \frac{L}{2}\|x_0 - x^*\|^2 - \frac{L}{2}\sum_{i=0}^{i-1}\|y_i - x_{i+1}\|^2$$

Convergence rate for smooth and convex case (proof)

5. Bound gradients with sufficient decrease lemma 7:

$$\sum_{i=0}^{k-1} [f(x_i) - f^*] \le \sum_{i=0}^{k-1} \left[f(x_i) - f(x_{i+1}) + \frac{L}{2} \|y_i - x_{i+1}\|^2 \right] + \frac{L}{2} \|x_0 - x^*\|^2 - \frac{L}{2} \sum_{i=0}^{i-1} \|y_i - x_{i+1}\|^2$$

$$\le f(x_0) - f(x_k) + \frac{L}{2} \sum_{i=0}^{i-1} \|y_i - x_{i+1}\|^2 + \frac{L}{2} \|x_0 - x^*\|^2 - \frac{L}{2} \sum_{i=0}^{i-1} \|y_i - x_{i+1}\|^2$$

$$\le f(x_0) - f(x_k) + \frac{L}{2} \|x_0 - x^*\|^2$$

$$\sum_{i=0}^{k-1} f(x_i) - k f^* \le f(x_0) - f(x_k) + \frac{L}{2} \|x_0 - x^*\|^2$$

$$\sum_{i=1}^{k} [f(x_i) - f^*] \le \frac{L}{2} \|x_0 - x^*\|^2$$
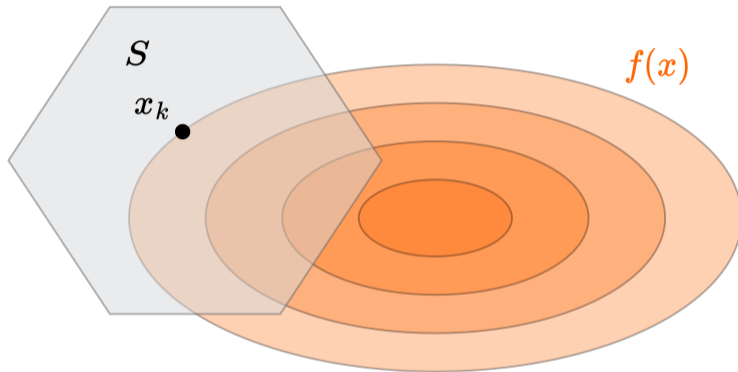
Idea

Figure 5: Illustration of Frank-Wolfe (conditional gradient) algorithm
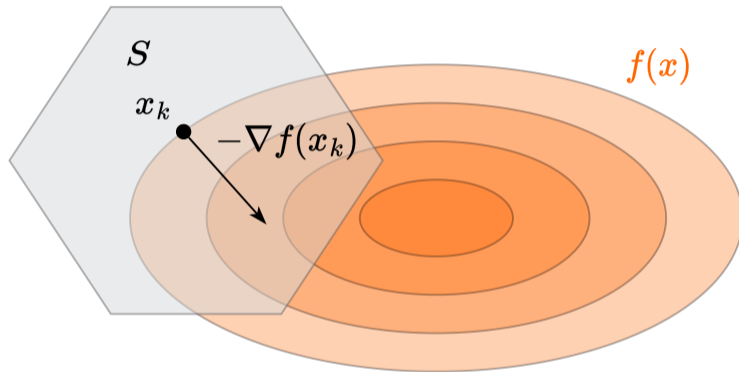
Idea

Figure 6: Illustration of Frank-Wolfe (conditional gradient) algorithm
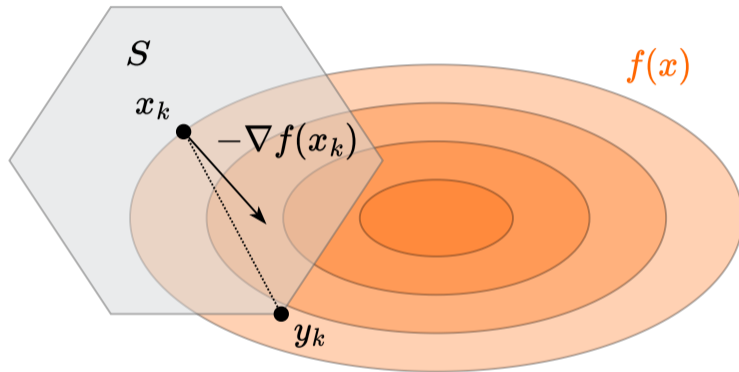
Idea

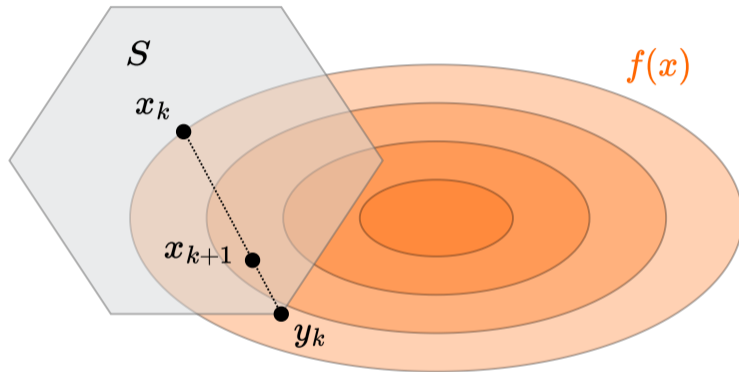Figure 7: Illustration of Frank-Wolfe (conditional gradient) algorithm

Idea

Figure 8: Illustration of Frank-Wolfe (conditional gradient) algorithm
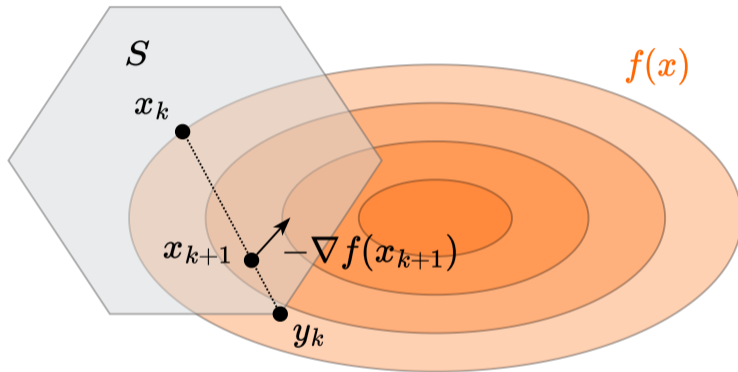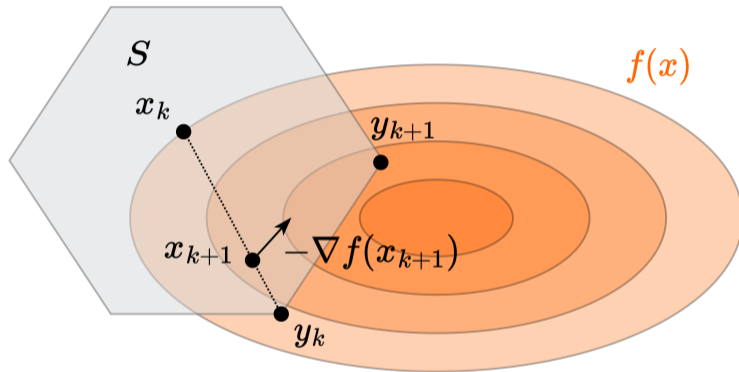
Idea

Figure 9: Illustration of Frank-Wolfe (conditional gradient) algorithm

Idea

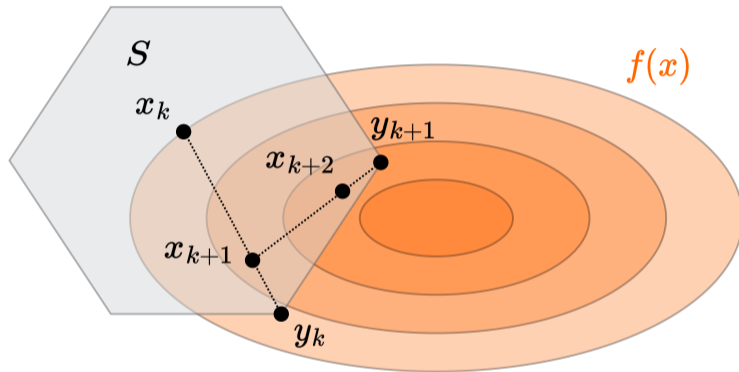Figure 10: Illustration of Frank-Wolfe (conditional gradient) algorithm

Idea

Figure 11: Illustration of Frank-Wolfe (conditional gradient) algorithm

Idea

$$y_k = \arg\min_{x \in S} f_{x_k}^I(x) = \arg\min_{x \in S} \langle \nabla f(x_k), x \rangle$$

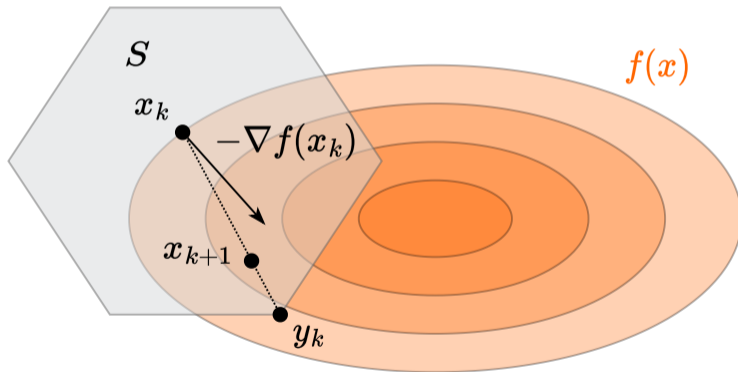$$x_{k+1} = \gamma_k x_k + (1 - \gamma_k) y_k$$

Figure 12: Illustration of Frank-Wolfe (conditional gradient) algorithm

Convergence

Comparison to PGD