Projection on (Linear) subspaces of $\mathbb{R}^n$

Suppose $\mathcal{S}$ is a linear subspace of $\mathbb{R}^n$, e.g., line including 0 in $\mathbb{R}^2$, plane including 0 in $\mathbb{R}^3$, etc.

Note that $\mathcal{S}$ is closed and convex.

Then, for $z \in \mathbb{R}^n$, $[z]^{\mathcal{S}} = y^*$ satisfies:

$$(z - y^*)^T (y - y^*) \leq 0 \quad \forall\, y \in \mathcal{S}.$$

But by the fact that $\mathcal{S}$ is a subspace,

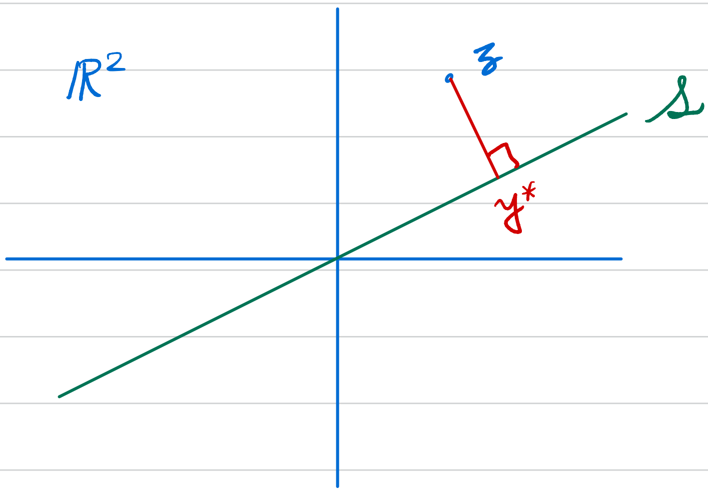
$$y - y^* = x \in \mathcal{S}, \quad \forall\, y \in \mathcal{S}.$$

$$\Rightarrow (z - y^*)^T x \leq 0 \quad \forall\, x \in \mathcal{S} \quad \text{--- (1)}$$

But $x \in \mathcal{S} \Rightarrow -x \in \mathcal{S}$ (since $\mathcal{S}$ is subspace)

$$\Rightarrow (z - y^*)^T x \geq 0 \quad \forall\, x \in \mathcal{S} \quad \text{--- (2)}$$

(1) and (2) $\Rightarrow (z - y^*)^T x = 0, \ \forall\, x \in \mathcal{S}$.



Orthogonality Principle

Constrained Optimization Example

$$\underset{x \in \mathcal{S}}{\text{maximize}} \quad x_1^{a_1} x_2^{a_2} \cdots x_n^{a_n}$$

$$\mathcal{S} = \left\{ x : \sum_{i=1}^{n} x_i = 1 , \; x_i \geq 0 , \; i = 1, 2, \ldots, n \right\} \leftarrow \text{convex and closed}$$

a_i, $i = 1, 2, \ldots, n$ are given positive scalars.

Opt. equivalent to : $\underset{x \in \mathcal{S}}{\text{minimize}} \; f(x)$

with $f(x) = -\sum a_i \ln x_i$

$$\nabla f(x) = \left(-\frac{a_1}{x_1} , -\frac{a_2}{x_2} , \ldots , -\frac{a_n}{x_n} \right)$$

$$\nabla^2 f(x) = \text{diag}\left(\frac{a_1}{x_1^2} , \frac{a_2}{x_2^2} , \ldots , \frac{a_n}{x_n^2} \right) \succ 0$$

$\Rightarrow$ f strictly convex

$x^* \in \mathcal{S}$ is (unique) min $\iff$ $\nabla f(x^*)^T (x - x^*) \geq 0 \quad \forall x \in \mathcal{S}$

$$\iff -\sum_{i=1}^{n} \frac{a_i}{x_i^*} (x_i - x_i^*) \geq 0 , \; \forall x \in \mathcal{S}$$

$$\iff -\sum_{i=1}^{n} a_i \frac{x_i}{x_i^*} + \sum_{i=1}^{n} a_i \geq 0 , \; \forall x \in \mathcal{S}$$

<u>Guess</u> : $x_i^* = \dfrac{a_i}{\sum_{i=1}^{n} a_i}$ —— (1)

Then : $-\sum_{i=1}^{n} a_i \dfrac{x_i}{x_i^*} + \sum_{i=1}^{n} a_i = \sum_{i=1}^{n} a_i \left(-\sum_{i=1}^{n} x_i + 1 \right)$

$$= 0 \quad \forall x \in \mathcal{S}$$

Thus x^* in (1) is unique min.

Projection Example : Box Constraints

$$\mathcal{S} = \{ y : a_i \le y_i \le b_i \} \quad i = 1, \dots, n$$

$\mathcal{S}$ is closed and convex

For $z \in \mathbb{R}^n$,

$$[z]^{\mathcal{S}} = \arg\min_{y \in \mathcal{S}} \|z - y\|^2 = \arg\min_{y \in \mathcal{S}} \sum_{i=1}^{n} (z_i - y_i)^2$$

If $z_i < a_i$, then $[z]_i^{\mathcal{S}} = y_i^* = a_i$

If $z_i > b_i$, then $[z]_i^{\mathcal{S}} = y_i^* = b_i$

If $a_i \le z_i \le b_i$, then $[z]_i^{\mathcal{S}} = y_i^* = z_i$

Can also verify that y^* satisfies necessary and sufficient condition:

$$(z - y^*)^T (y - y^*) \le 0 \quad \forall y \in \mathcal{S}$$

i.e. $$\sum_{i=1}^{n} (z_i - y_i^*)(y_i - y_i^*) \le 0 \quad \forall y \in \mathcal{S}$$

If $z_i < a_i$, $(z_i - y_i^*)(y - y_i^*) = \underbrace{(z_i - a_i)}_{<0} \underbrace{(y_i - a_i)}_{\ge 0} \le 0$

If $z_i > b_i$, $(z_i - y_i^*)(y - y_i^*) = \underbrace{(z_i - b_i)}_{>0} \underbrace{(y_i - b_i)}_{\le 0} \le 0$

If $a_i \le z_i \le b_i$, $(z_i - y_i^*)(y - y_i^*) = 0$

Gradient Projection Method

$$\min_{x \in S} f(x) \qquad S : \text{convex, closed.}$$

$$x_{k+1} = [x_k + \alpha_k d_k]^S$$

Special Case Fixed step-size, steepest descent

$$x_{k+1} = [x_k - \alpha \nabla f(x_k)]^S \quad \text{—— (1)}$$

Definition $\tilde{x}$ is a fixed (stationary) point of iteration in (1) if:

$$\tilde{x} = [\tilde{x} - \alpha \nabla f(\tilde{x})]^S.$$

Result If f has L-Lipschitz gradient and $0 < \alpha < \frac{2}{L}$, every limit point of (1) is a fixed point of (1).

Proof By the Descent Lemma,

$$f(x_{k+1}) \leq f(x_k) + \nabla f(x_k)^T (x_{k+1} - x_k) + \frac{L}{2} \| x_{k+1} - x_k \|^2 \quad \cdots (2)$$

By necc. and suff. condition for projection,

$$(x_k - \alpha \nabla f(x_k) - x_{k+1})^T (x - x_{k+1}) \leq 0, \quad \forall x \in S$$

set $x = x_k$

$$\Rightarrow \alpha \nabla f(x_k)^T (x_{k+1} - x_k) \leq - \| x_k - x_{k+1} \|^2 \quad \cdots (3)$$

$$(2) \& (3) \Rightarrow f(x_{k+1}) - f(x_k) \leq \underbrace{\left(\frac{L}{2} - \frac{1}{\alpha} \right)}_{<0} \| x_k - x_{k+1} \|^2$$

If $\{x_k\}$ has limit point $\bar{x}$, LHS $\xrightarrow[k \in \mathcal{K}]{k \to \infty} 0$

$$\Rightarrow \| x_{k+1} - x_k \| \xrightarrow[k \in \mathcal{K}]{k \to \infty} 0 \Rightarrow [\bar{x} - \alpha \nabla f(\bar{x})]^S = \bar{x}$$

Result If f is convex, then

x^* is a minimizer of f over $\mathcal{S}$

$\Longleftrightarrow$ $x^* = [x^* - \alpha \nabla f(x^*)]^{\mathcal{S}}$

(i.e., x^* is a fixed point of (1))

Proof

optimality condition

x^* minimizes f over $\mathcal{S}$ $\Longleftrightarrow$ $\nabla f(x^*)^T (x - x^*) \geq 0, \ \forall x \in \mathcal{S}$

$\Longleftrightarrow$ $-\alpha \nabla f(x^*)^T (x - x^*) \leq 0 \quad \forall x \in \mathcal{S}$

$\Longleftrightarrow$ $(x^* - \alpha \nabla f(x^*) - x^*)^T (x - x^*) \leq 0 \quad \forall x \in \mathcal{S}$

$\Longleftrightarrow$ $[x^* - \alpha \nabla f(x^*)]^{\mathcal{S}} = x^*$

Projection Theorem

Convergence of Gradient Projection Method for convex f

If f is convex and has L-Lipschitz gradient, it can be shown that for $0 < \alpha < \frac{2}{L}$

$f(x_k) \to f(x^*)$ at rate $\frac{1}{k}$ (Same as unconstrained lec 7)

Assume Lipschitz gradients and strong convexity:

$$\|\nabla f(x) - \nabla f(y)\| \leq M \|x - y\|, \quad \forall x, y \in \mathcal{S}$$

$$\nabla^2 f(x) \succcurlyeq mI \quad \forall x \in \mathcal{S},$$

Let x^* be the (unique) min. of f over $\mathcal{S}$

$$\|x_{k+1} - x^*\|^2 = \|[x_k - \alpha \nabla f(x_k)]^{\mathcal{X}} - x^*\|^2$$

previous result

$$= \|[x_k - \alpha \nabla f(x_k)]^{\mathcal{X}} - [x^* - \alpha \nabla f(x^*)]^{\mathcal{X}}\|^2$$

non-expansive

$$\leq \|x_k - \alpha \nabla f(x_k) - (x^* - \alpha \nabla f(x^*))\|^2$$

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

$$\leq \|x_k - x^*\|^2 (1 + \alpha^2 M^2)$$
$$+ 2\alpha (x^* - x_k)^T \nabla f(x_k) + 2\alpha f(x_k)$$
$$+ 2\alpha (x_k - x^*) \nabla f(x^*) - 2\alpha f(x_k)$$

By strong convexity, $f(x^*) \geq f(x_k) + (x^* - x_k)^T \nabla f(x_k) + \frac{1}{2} \|x^* - x_k\|^2$

Thus,

$$\|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 (1 + \alpha^2 M^2 - \alpha m) + 2\alpha f(x^*)$$
$$+ 2\alpha (x_k - x^*)^T \nabla f(x^*) - 2\alpha f(x_k)$$

By strong convexity,

$$f(x_k) \geq f(x^*) + \nabla f(x^*)^T (x_k - x^*) + \frac{1}{2} m \|x_k - x^*\|^2$$

Thus,

$$|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 (1 + \alpha^2 M^2 - 2\alpha m)$$

$\Rightarrow$ if $|1 + \alpha^2 M^2 - 2\alpha m| < 1$,

$x_N \to x^*$ geometrically as $N \to \infty$

(same as unconstrained case – lec 7)

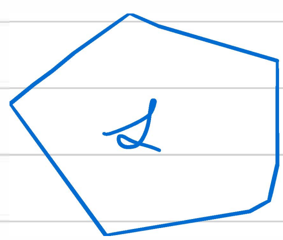
Remarks :

- In order for gradient projection method to be useful in practice, it is necessary that the projection operation is easy to implement

e.g. box constraint set,

$$\mathcal{S} = \{ y : a_i \le y_i \le b_i \} \quad i=1,\dots,n$$

- If constraint set is a polyhedron, i.e., a convex region bounded by hyperplanes



Then there are efficient ways to implement constrained optimization

- There are a number of ad hoc approaches to avoid or simplify projection step, e.g., use adaptive step-size (Armijo rule) to keep x_{k+1} in $\mathcal{S}$.

Convergence of these methods difficult to establish.