

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

Suppose  $S$  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  $S$  is closed and convex.

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

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

But by the fact that  $S$  is a subspace,

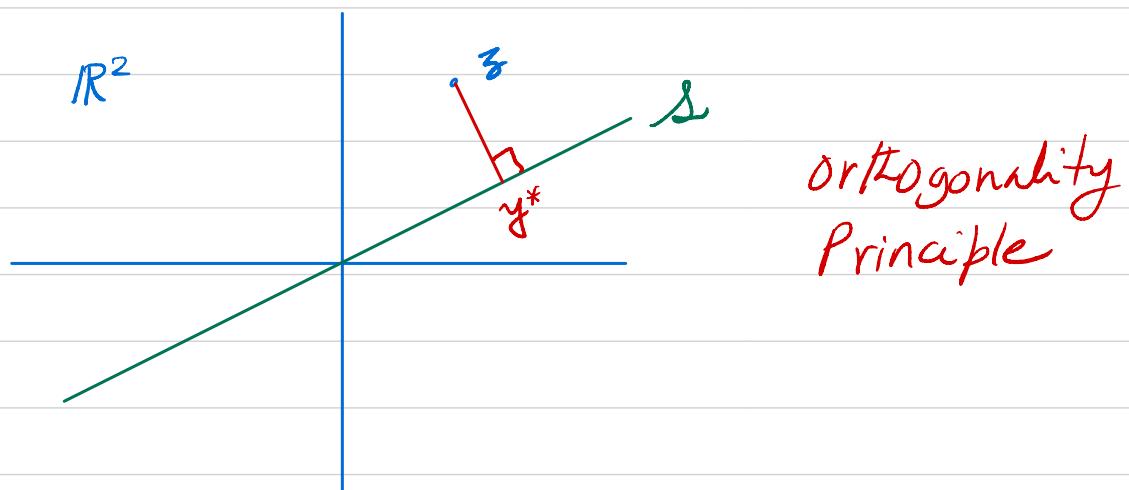
$$y - y^* = x \in S, \quad \forall y \in S.$$

$$\Rightarrow (z - y^*)^T x \leq 0 \quad \forall x \in S \quad — (1)$$

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

$$\Rightarrow (z - y^*)^T x \geq 0 \quad \forall x \in S \quad — (2)$$

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



## Constrained Optimization Example

maximize  $x_1^{a_1} x_2^{a_2} \dots x_n^{a_n}$   
 $x \in S$

$$S = \{x : \sum_{i=1}^n x_i = 1, x_i \geq 0, i=1, 2, \dots, n\} \leftarrow \begin{array}{l} \text{convex} \\ \text{and} \\ \text{closed} \end{array}$$

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

Opt. equivalent to : minimize  $f(x)$   
 $x \in S$

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

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

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

$\Rightarrow f$  strictly convex

$$\begin{aligned} x^* \in S \text{ is (unique) min} &\iff \nabla f(x^*)^T (x - x^*) \geq 0, \forall x \in S \\ &\iff -\sum_{i=1}^n \frac{a_i}{x_i^*} (x_i - x_i^*) \geq 0, \forall x \in S \\ &\iff -\sum_{i=1}^n a_i \frac{x_i}{x_i^*} + \sum_{i=1}^n a_i \geq 0, \forall x \in S \end{aligned}$$

Guess :  $x_i^* = \frac{a_i}{\sum_{i=1}^n a_i} \quad (1)$

$$\begin{aligned} \text{Then} : -\sum_{i=1}^n a_i \frac{x_i}{x_i^*} + \sum_{i=1}^n a_i &= \sum_{i=1}^n a_i \left( -\sum_{i=1}^n \frac{x_i}{a_i} + 1 \right) \\ &= 0 \quad \forall x \in S \end{aligned}$$

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

## Projection Example : Box constraints

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

$\mathcal{S}$  is closed and convex

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

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

If  $z_i < a_i$ , then  $[z]_i^\delta = y_i^* = a_i$

If  $z_i > b_i$ , then  $[z]_i^\delta = y_i^* = b_i$

If  $a_i \leq z_i \leq b_i$ , then  $[z]_i^\delta = y_i^* = z_i$

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

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

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

$$\text{If } z_i < a_i, (z_i - y_i^*) (y_i - y_i^*) = (z_i - a_i) \stackrel{<0}{(y_i - a_i)} \stackrel{\geq 0}{\leq 0} \leq 0$$

$$\text{If } z_i > b_i, (z_i - y_i^*) (y_i - y_i^*) = (z_i - b_i) \stackrel{>0}{(y_i - b_i)} \stackrel{\leq 0}{\leq 0} \leq 0$$

$$\text{If } a_i \leq z_i \leq b_i, (z_i - y_i^*) (y_i - y_i^*) = 0$$

## Gradient Projection Method

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

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

Special Case Fixed step-size, steepest descent

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

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

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

### Result

$$\tilde{x} = [\tilde{x} - \alpha \nabla f(\tilde{x})]^\perp \Leftrightarrow \nabla f(\tilde{x})^\top (x - \tilde{x}) \geq 0, \forall x \in S$$

Proof By necessary and sufficient condition for projection

$$\tilde{x} = [\tilde{x} - \alpha \nabla f(\tilde{x})]^\perp \Leftrightarrow (\tilde{x} - (\tilde{x} - \alpha \nabla f(\tilde{x})))^\top (x - \tilde{x}) \geq 0 \quad \forall x \in S$$

$$\begin{aligned} \alpha > 0 &\Leftrightarrow \alpha \nabla f(\tilde{x})^\top (x - \tilde{x}) \geq 0, \forall x \in S \\ &\Leftrightarrow \nabla f(\tilde{x})^\top (x - \tilde{x}) \geq 0, \forall x \in S \end{aligned}$$

### Corollary

$x^*$  is a local min  $\Rightarrow x^* = [x^* - \alpha \nabla f(x^*)]^\perp$   
of  $f$  on  $S$     i.e.,  $x^*$  is a fixed point of (1)

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,

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

By nec. and suff. condition for projection,

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

$\nwarrow$  Set  $x = x_k$

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

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

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

$$\Rightarrow \|x_{k+1} - x_k\| \xrightarrow[k \rightarrow \infty]{\mathcal{K}} 0, \text{ i.e. } \| [x_k - \alpha \nabla f(x_k)]^\perp - x_k \| \xrightarrow[k \rightarrow \infty]{\mathcal{K}} 0$$

$$\Rightarrow \| [\bar{x} - \alpha \nabla f(\bar{x})]^\perp - \bar{x} \| = 0 \Rightarrow [\bar{x} - \alpha \nabla f(\bar{x})]^\perp = \bar{x}$$

Note that any  $x^* \in \mathcal{S}$  that satisfies  $\nabla f(x^*) = 0$ , also (trivially) satisfies

$$x^* = [x^* - \alpha \nabla f(x^*)]^\perp$$

Therefore, fixed points are necessarily be local minima but they generally tend to be (Capture Theorem)

## Convergence of Gradient Projection Method for convex $f$

Result: For convex  $f$ ,

$x^*$  is a global min  $\Leftrightarrow x^* = [x^* - \alpha \nabla f(x^*)]^{-1}$   
of  $f$  on  $S$  i.e.,  $x^*$  is a fixed point of (1)

Proof Follows from necessary and sufficient condition  
for optimality for convex  $f$  (lec 10).

Therefore, if  $f$  is convex and has  $L$ -Lipschitz gradient  
with  $0 < \alpha < \frac{2}{L}$ , gradient projection:

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

converges to a global min of  $f$  (assuming existence)

Furthermore, it can be shown that

$$f(x_N) \rightarrow f(x^*) \text{ at rate } \frac{1}{N} \quad \begin{pmatrix} \text{Same as unconstrained} \\ \text{case (lec 7)} \end{pmatrix}$$

---

Assume Lipschitz gradients and strong convexity:

$$\|\nabla f(x) - \nabla f(y)\| \leq L \|x - y\|, \quad \forall x, y \in S$$

$$\nabla^2 f(x) \succcurlyeq mI \quad \forall x \in S.$$

Let  $x^*$  be the (unique) min. of  $f$  over  $S$   
(assuming existence)

$$\begin{aligned}
\|x_{k+1} - x^*\|^2 &= \|[\bar{x}_k - \alpha \nabla f(\bar{x}_k)]^\alpha - x^*\|^2 \\
&= \|\bar{x}_k - \alpha \nabla f(\bar{x}_k) - [x^* - \alpha \nabla f(x^*)]^\alpha\|^2 \\
&\stackrel{\text{previous result}}{\leq} \|x_k - \alpha \nabla f(x_k) - (x^* - \alpha \nabla f(x^*))\|^2 \\
&\stackrel{\text{non-expansive}}{=} \|x_k - x^*\|^2 + \alpha^2 \|\nabla f(x_k) - \nabla f(x^*)\|^2 \\
&\quad - 2\alpha (x_k - x^*)^\top (\nabla f(x_k) - \nabla f(x^*)) \\
&\leq \|x_k - x^*\|^2 (1 + \alpha^2 L^2) \\
&\quad + 2\alpha (x^* - x_k)^\top \nabla f(x_k) + 2\alpha f(x_k) \\
&\quad + 2\alpha (x_k - x^*)^\top \nabla f(x^*) - 2\alpha f(x_k)
\end{aligned}$$

By strong convexity,

$$\begin{aligned}
f(x^*) &\geq f(x_k) + (x^* - x_k)^\top \nabla f(x_k) + \frac{1}{2} \|x^* - x_k\|^2 \\
\Rightarrow \|x_{k+1} - x^*\|^2 &\leq \|x_k - x^*\|^2 (1 + \alpha^2 L^2 - \alpha m) + 2\alpha f(x^*) \\
&\quad + 2\alpha (x_k - x^*)^\top \nabla f(x^*) - 2\alpha f(x_k)
\end{aligned}$$

By strong convexity,

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

Thus,

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

$$\Rightarrow \text{if } |1 + \alpha^2 L^2 - 2\alpha m| < 1,$$

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

(Same as unconstrained case - lec 7)

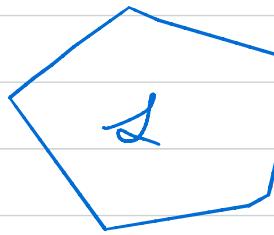
## Remarks :

- 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,

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

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



Then there are efficient ways to implement constrained optimization - see sec 3.4 - 3.6 of text .

- 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 .