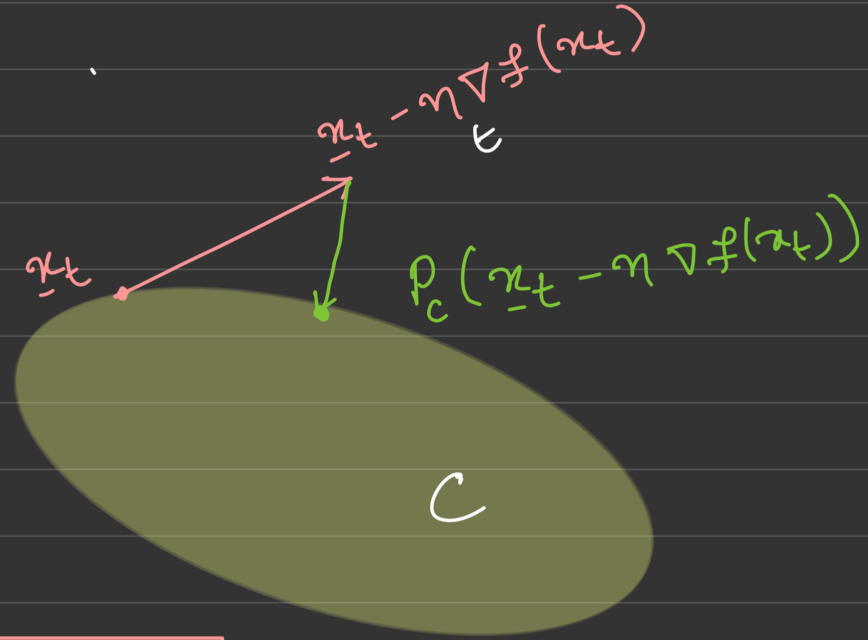
Lecture 11 : Projected and Proximal methods

E1 260

- Projected gradient descent
- Convergence analysis for Lipschitz, Smooth & Strongly convex functions.

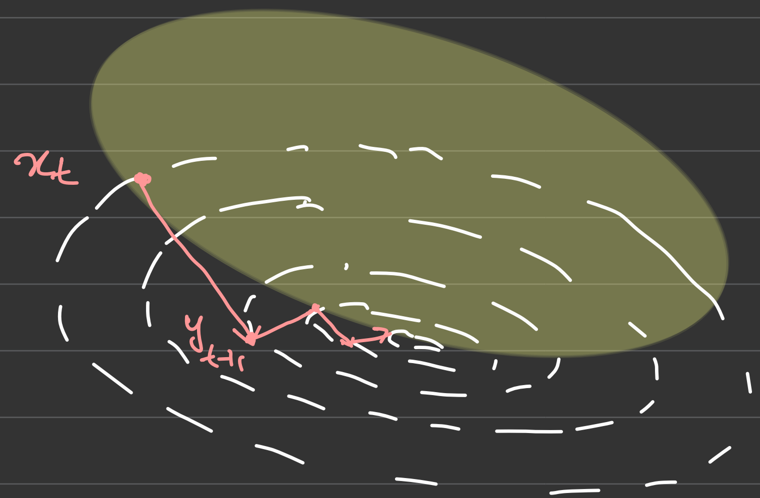
Projected gradient descent

$$\boxed{\begin{array}{ll} \text{minimize} & f(\underline{x}) \\ \text{subject to} & \underline{x} \in C \end{array}}$$



Closed convex set : $C \subseteq \mathbb{R}^d$

$$\boxed{\begin{aligned} \underline{y}_{t+1} &= \underline{x}_t - \eta \nabla f(\underline{x}_t) \\ \underline{x}_{t+1} &= P_c(\underline{y}_{t+1}) \end{aligned}} \notin C$$



$$P_c(x) := \arg\min_{\underline{x} \in C} \; \| \underline{x} - \underline{y}_{t+1} \|^2$$

Euclidean projection onto C

LASSO:

$$\underset{\underline{x}}{\text{minimize}} \quad \| A\underline{x} - \underline{b} \|_2^2 \qquad \text{s.t.} \quad \| \underline{x} \|_1 \le \delta$$

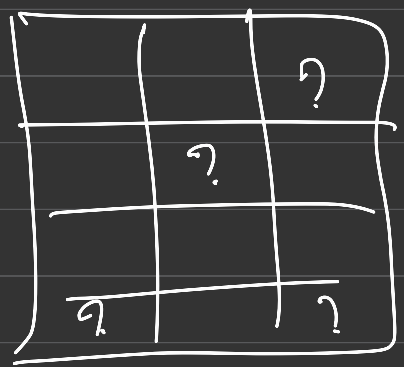
$$\longleftrightarrow \quad \min. \ \| \underline{x} \|_1 \qquad \| A\underline{x} - b \|_2 \le \varepsilon$$

$$\underline{b} = A\underline{x} + \underline{n}$$

Matrix completion:

movies

user



X

$$\underset{X}{\min.} \quad \| Y - P_\Omega(X) \|_F^2$$

$$\text{s.t.} \quad \underline{\| X \|_*} \le R$$

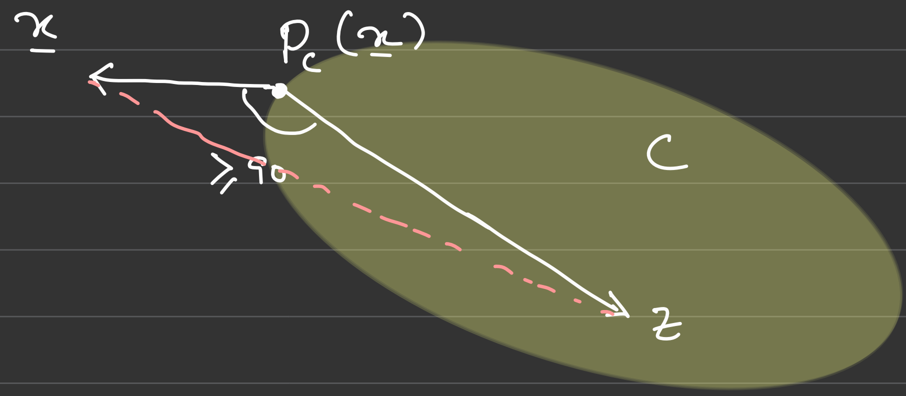
$$\text{rank}(X) \le r$$

Observation: Y, P_Ω : sampling operation

Projection theorem :

- Let C be closed & convex. Then, $P_c(\underline{x})$, is projection of $\underline{x}$ onto C, then

$$(\underline{x} - P_C(\underline{x}))^T (\underline{z} - P_c(\underline{x})) \leq 0, \quad \forall \underline{z} \in C$$



For all z and x_c in C

$$\| \underline{z} - \underline{x} \|^2 = \| \underline{z} - \underline{x}_c \|^2 + \| x_c - x \|^2 - 2(\underline{z} - \underline{x}_c)^T (\underline{x}_c - \underline{x})$$

$$\geq \| \underline{x}_c - x \|^2 - 2(\underline{z} - x_c)^T (\underline{x}_c - \underline{x})$$

So if $\underline{x}_c$ is such that

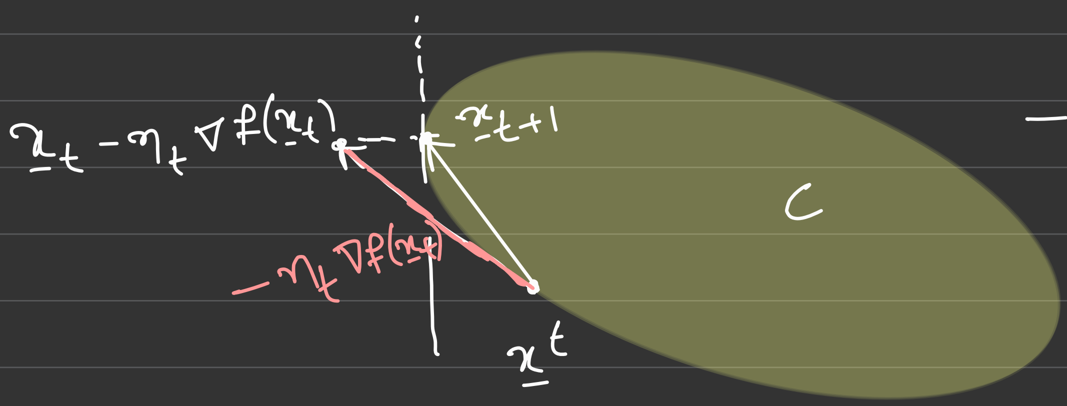
$$(\underline{z} - x_c)^T (\underline{x}_c - \underline{x}) \geq 0 \quad \text{for all } \underline{z} \in C$$

we have $\|\underline{z} - \underline{x}\|^2 \geq \|\underline{x}_c - \underline{x}\|^2 \quad \forall\, \underline{z} \in C$

$$\Rightarrow \quad x_c = P_C(x)$$

as P_C by definition is the minimizer of the differentiable convex function $\|\underline{z} - \underline{x}\|^2$ over C

A direct consequence:

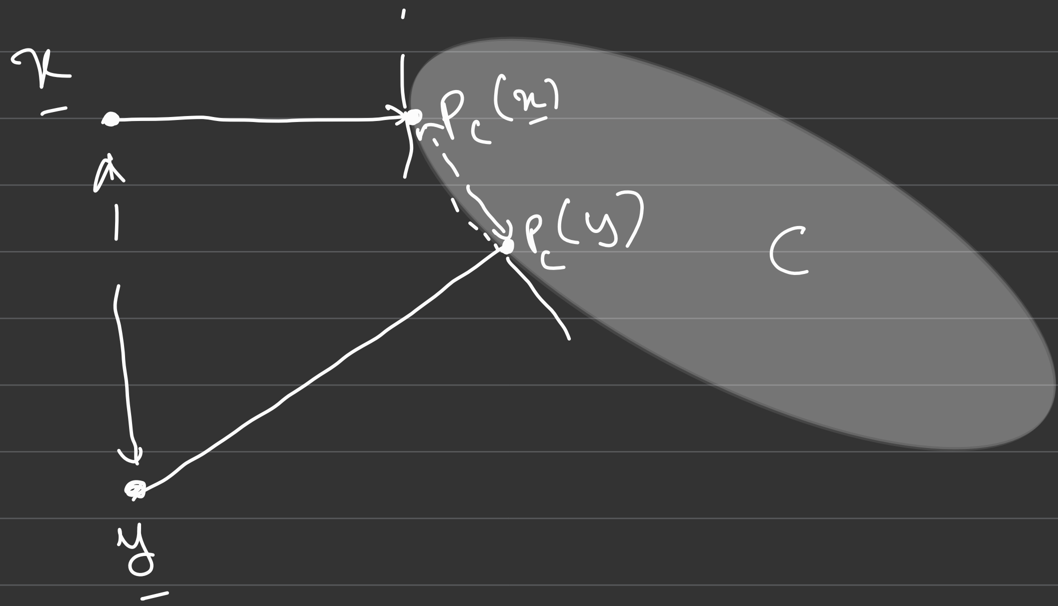


$$-\nabla f^T(x_t)\,(\underline{x}_{t+1} - \underline{x}_t) \geq 0$$

- So $x_{t+1} - \underline{x}_t$ is positively correlated with the steepest descent direction.

Non expansiveness of the projection operator:

$$\| P_C(\underline{x}) - P_C(\underline{y}) \|_2 \leq \| \underline{x} - \underline{y} \|_2$$



Summary: [Recall from gradient descent]

Gradient descent with fixed step size

$$\eta = \frac{1}{L}$$

Bounded gradients	$O(1/\varepsilon^2)$
Smooth	$O(1/\varepsilon)$
Smooth & Strongly convex	$O(\log(1/\varepsilon))$

Do these hold for projected gradient descent as well?

1. Bounded gradients : $O(1/\varepsilon^2)$ with $\underline{x}_0 \in C$

Recall from vanilla analysis:

$$\underline{g}_t = (\underline{y}_{t+1} - \underline{x}_t)/\eta$$

$$\underline{g}_t^T(\underline{x}_t - \underline{x}^*) = \frac{1}{2\eta}\left(\eta^2 \|\underline{g}_t\|^2 + \|\underline{x}_t - \underline{x}^*\|^2 - \|\underline{y}_{t+1} - \underline{x}^*\|^2\right)$$

From projection theorem: (with $\underline{z} = \underline{x}^*$ and $\underline{x} = \underline{y}_{t+1}$)

$$(\underline{y}_{t+1} - \underline{x}_{t+1})^T(\underline{x}^* - \underline{x}_{t+1}) \leq 0,$$

$$\Longrightarrow \quad \|\underline{x}_{t+1} - \underline{x}^*\|^2 \leq \|\underline{y}_{t+1} - \underline{x}^*\|^2$$

So $\underline{g}_t^T(\underline{x}_t - \underline{x}^*) \leq \frac{1}{2\eta}\left(\eta^2 \|\underline{g}_t\|^2 + \|\underline{x}_t - \underline{x}^*\|^2 - \|\underline{x}_{t+1} - \underline{x}^*\|^2\right)$

Remainder of the proof is as before

$$\sum_{t=0}^{T-1} \underline{g}_t^T (\underline{x}_t - \underline{x}^*) \leq \frac{1}{2\eta} \Big(\eta^2 \sum_{t=0}^{T-1} \underbrace{\|g_t\|^2}_{\leq B} + \underbrace{\| \underline{x}_0 - x^* \|^2}_{= R} \Big)$$

From convexity,

$$f(\underline{x}_t) - f(\underline{x}^*) \leq \underline{g}_t^T (\underline{x}_t - \underline{x}^*)$$

$$\Rightarrow \frac{1}{T} \sum_{t=0}^{T-1} f(x_t) - f(x^*) \leq \underbrace{\frac{1}{2\eta} \left[\eta^2 T B + R^2 \right] = q(\eta)}$$

Choosing $\eta = \dfrac{R}{B\sqrt{T}}$ that minimizes

$$\frac{1}{T} \sum_{t=0}^{T-1} f(\underline{x}_t) - f(\underline{x}^*) \leq \frac{RB}{\sqrt{T}} \quad \text{and} \quad T \geq \frac{R^2 B^2}{\varepsilon^2}$$

2. <u>Smooth</u> <u>convex</u> <u>functions</u> : $O(1/\varepsilon)$

$$f(y) \leq f(\underline{x}) + \nabla f^T(x)(\underline{y}-\underline{x}) + \frac{L}{2}\|\underline{x}-\underline{y}\|^2$$

$\forall\ x, y \in C$

Projected gradient descent with $\underline{x}_0 \in C$ and

$\eta = \frac{1}{L}$ satisfies

Sufficient decrease Lemma:

- $f(\underline{x}_{t+1}) \leq f(\underline{x}_t) - \frac{1}{2L}\|\nabla f(\underline{x}_t)\|^2 + \frac{L}{2}\|\underline{y}_{t+1} - \underline{x}_{t+1}\|^2$
- $f(\underline{x}_T) - f(\underline{x}^*) \leq \frac{L}{2T}\|\underline{x}_0 - x^*\|^2$; $T > 0$

<u>Proof:</u> From smoothness, with $\underline{y} = \underline{x}_{t+1}$ and $\underline{x} = \underline{x}_t$

$$f(\underline{x}_{t+1}) \leq f(\underline{x}_t) + \nabla f^T(\underline{x}_t)(\underline{x}_{t+1} - \underline{x}_t) + \frac{L}{2}\|\underline{x}_t - \underline{x}_{t+1}\|^2$$

Recall, $-\frac{1}{L}\nabla f(\underline{x}_t) = \underline{y}_{t+1} - \underline{x}_t$

Then

$$f(\underline{x}_{t+1}) \le f(\underline{x}_t) - L(\underline{y}_{t+1} - \underline{x}_t)^T(\underline{x}_{t+1} - \underline{x}_t) + \frac{L}{2}\|\underline{x}_t - \underline{x}_{t+1}\|^2$$

$$= f(\underline{x}_t) - \frac{L}{2}\left[\|\underline{y}_{t+1} - \underline{x}_t\|^2 + \|\underline{x}_{t+1} - \underline{x}_t\|^2 - \|\underline{y}_{t+1} - \underline{x}_{t+1}\|^2\right] + \frac{L}{2}\|x_t - x_{t+1}\|^2$$

$$= f(\underline{x}_t) - \frac{L}{2}\|\underline{y}_{t+1} - \underline{x}_t\|^2 + \frac{L}{2}\|\underline{y}_{t+1} - \underline{x}_{t+1}\|^2$$

$$= f(\underline{x}_t) - \frac{1}{2L}\|\nabla f(x_t)\|^2 + \frac{L}{2}\|y_{t+1} - x_{t+1}\|^2$$

*

Recall,

$$\underline{g}_t^T(\underline{x}_t - \underline{x}^*) = \frac{1}{2\eta}\Big(\eta^2\|\underline{g}_t\|^2 + \|\underline{x}_t - \underline{x}^*\|^2 - \|\underline{y}_{t+1} - \underline{x}^*\|^2\Big)$$

From Projection theorem,

$$\|\underline{x}_{t+1} - \underline{x}^*\|^2 + \|\underline{y}_{t+1} - \underline{x}_{t+1}\|^2 \le \|\underline{y}_{t+1} - \underline{x}^*\|$$

$$\underline{g}_t^T(\underline{x}_t - \underline{x}^*) \le \frac{1}{2\eta}\Big(\eta^2\|\underline{g}_t\|^2 + \|\underline{x}_t - \underline{x}^*\|^2 - \|\underline{x}_{t+1} - \underline{x}^*\|^2 \cancel{+} \|\underline{y}_{t+1} - \underline{x}_{t+1}\|^2\Big)$$

From convexity,

$$\sum_{t=0}^{T-1} f(\underline{x}_t) - f(\underline{x}^*) \le \sum_{t=0}^{T-1} \underline{g}_t^T(\underline{x} - \underline{x}^*)$$

$$\leq \frac{1}{2L} \sum_{t=0}^{T-1} \|g_t\|^2 + \frac{L}{2} \|x_0 - x^*\|^2$$

$$- \frac{L}{2} \sum_{t=0}^{T-1} \|y_{t+1} - x_{t+1}\|^2$$

From $\circledast$

$$\frac{1}{2L} \sum_{t=0}^{L-1} \|g_t\|^2 \leq \sum_{t=0}^{T-1} \left(f(x_t) - f(x_{t+1}) + \frac{L}{2} \|y_{t+1} - x_{t+1}\|^2 \right)$$

$$= f(x_0) - f(x_T) + \frac{L}{2} \sum_t \|y_{t+1} - x_{t+1}\|^2$$

$$\Rightarrow \sum_{t=1}^{T} f(x_t) - f(x^*) \leq \frac{L}{2} \|x_0 - x^*\|^2$$

$$\leq \frac{L}{2} R^2$$

Since the last iterate is the best

$$f(x_T) - f(x^*) \leq \left(\frac{1}{T} \sum_{t=1}^{T} f(x_t) \right) - f(x^*) \leq \frac{L}{2T} R^2$$

$$\Rightarrow T \geq R^2 L / \varepsilon$$

3. L - Smooth & μ - strongly convex : $O(\log(1/\varepsilon))$

$f : \mathrm{dom}(f) \to \mathbb{R}$ be convex & differentiable with smoothness parameter L and strong convexity parameter μ. Let $C \subseteq \mathrm{dom}(f)$. Then, with $\eta = \frac{1}{L}$ projected gradient descent satisfies

- $$\|\underline{x}_{t+1} - \underline{x}^*\|^2 \le \left(1 - \frac{\mu}{L}\right) \|\underline{x}_t - \underline{x}^*\|^2 \quad t \ge 0$$

- $$f(\underline{x}_T) - f(\underline{x}^*) \le \frac{L}{2}\left(1 - \frac{\mu}{L}\right)^T \|\underline{x}_0 - \underline{x}^*\|^2$$
$$+ \underbrace{\|\nabla f(\underline{x}^*)\|}_{\neq 0 \text{ when } x \notin \mathrm{int}(C)} \left(1 - \frac{\mu}{L}\right)^{T/2} \|\underline{x}_0 - \underline{x}^*\|^2$$

Homework # 2.