Note Title 28-Jun-19

LECTURE-8

CONVEX - CONCAVE MINIMAX OPTIMIZATION

$$\min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} f(x,y)$$

$f(\cdot\,, y)$ is Convex

$f(x, \cdot)$ is Concave

g is Concave
$\triangleq -g$ is Convex

1) $g(\alpha y_1 + (1-\alpha) y_2) \geq \alpha g(y_1) + (1-\alpha) g(y_2)$

2) $g(z) \leq g(y) + \langle \nabla g(y), z - y \rangle$

1) Constrained Optimization : Find x s.t. $f_i(x) \le 0 \quad i=1,\dots m.$ $\equiv$ Find x s.t. $\max_i f_i(x) \le 0$

Each f_i is convex.

$$\min_x \max_{i=1,\dots m} f_i(x) \leftarrow$$

$$|||$$

$$\min_x \max_{\substack{y_i \ge 0 \\ \sum_{i=1}^m y_i = 1}} \sum_{i=1}^m y_i f_i(x)$$

$$f(x,y) = \sum_{i=1}^m y_i f_i(x)$$

$f(\cdot, y)$ is convex $\forall y$

$f(x, \cdot)$ is linear in $\cdot$ $\forall x$. $\rightarrow$ concave

2) Many non-smooth functions can be written as smooth minimax problems.

a) $\|x\|_1 = \max_{-1 \le y_i \le 1} \sum_i y_i x_i$

$$\min_x \frac{1}{2}\|Ax-b\|^2 + \lambda \|x\|_1 = \underbrace{\min_x \underbrace{\max_{-1 \le y_i \le 1} \frac{1}{2}\|Ax-b\|^2 + \lambda \sum_i y_i x_i}}$$

b) Given $a_1, \dots a_m$ find the smallest ball covering $a_1, \dots a_m$.

$\longrightarrow$ Again can be written as smooth minimax problem.

3) Minimax $\equiv$ zero sum games.

$$\underset{x}{\min} \; \underbrace{\underset{y}{\max} \; f(x,y)}_{\triangleq g(x)}$$

$f(\cdot, y)$ is convex $\forall y$
$f(x, \cdot)$ is concave $\forall x$.

<u>Exercise</u>: $g(x)$ is convex

How to evaluate $\nabla g(x)$?

(Informal) <u>Danskin's thm</u>: Under smoothness assumptions on $f(\cdot)$,

$$\nabla g(x) = \text{Conv. hull} \{ \nabla_x f(x,y) : y \in \underset{z}{\arg\max} \; f(x,z) \}.$$

Convex opt.

1) Find $y_t \in \arg\max_y f(x_t, y)$

2) $x_{t+1} = x_t - \eta \nabla_x f(x_t, y_t)$

Subgradient descent on $g(x)$.

Issues: ① Finding y_t takes time

② g could be non smooth $\Rightarrow$ slow convergence rate.

Non smooth : Gradient descent ascent $\longrightarrow O\left(\frac{1}{\sqrt{T}}\right)$

$g(x)$ could still be nonsmooth ← Smooth : Mirror-Prox $\longrightarrow O\left(\frac{1}{T}\right)$.

[f Could be arbitrary] maximin $\leq$ minimax

<u>Weak duality</u> : $\underbrace{\max_{y \in \mathcal{Y}} \min_{x \in \mathcal{X}} f(x,y)}_{LHS} \leq \underbrace{\min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} f(x,y)}_{RHS}$.

$$y^* = \arg\max_{y \in \mathcal{Y}} \left[\min_{x \in \mathcal{X}} f(x,y) \right]$$

$$x^* = \arg\min_{x \in \mathcal{X}} \left[\max_{y \in \mathcal{Y}} f(x,y) \right]$$

$$LHS = \min_{x \in \mathcal{X}} f(x, y^*) \leq f(x^*, y^*) \leq \max_{y \in \mathcal{Y}} f(x^*, y) = RHS.$$

[Sion's minimax thm.]

<u>Strong duality</u>: If $f(x,y)$ is Convex-Concave and $\mathcal{X}$ and $\mathcal{Y}$ are Compact then $\min\limits_{x\in\mathcal{X}} \max\limits_{y\in\mathcal{Y}} f(x,y) = \max\limits_{y\in\mathcal{Y}} \min\limits_{x\in\mathcal{X}} f(x,y)$.

<u>Defn. ϵ-primal dual pair</u>: $(\bar{x},\bar{y})$ is Said to be an ϵ-primal dual pair if $\max\limits_{y} f(\bar{x},y) - \min\limits_{x} f(x,\bar{y}) \leq \epsilon$.

<u>Note</u>: $\max\limits_{y} f(\bar{x},y) \geq f(\bar{x},\bar{y}) \geq \min\limits_{x} f(x,\bar{y})$.

Let $(\bar{x}, \bar{y})$ be an ϵ-primal dual pair

Exercise: Show that $\bar{x}$ is an ϵ-optimal soln. for

$$\min_x g(x) \triangleq \max_y f(x,y)$$

Exercise: $\bar{y}$ is an ϵ-optimal soln. for $\max_y h(y) \triangleq \min_x f(x,y)$.

$\mathrm{diam}(\mathcal{X}), \mathrm{diam}(\mathcal{Y}) \leq R$

Gradient descent ascent:

$f(\cdot, y)$ is convex

$f(x, \cdot)$ is concave

$$x_{t+1} = x_t - \eta \cdot \nabla_x f(x_t, y_t) \to \text{GD for min}$$

$\|\nabla_x f(x,y)\| \leq G$

$$y_{t+1} = y_t + \eta \nabla_y f(x_t, y_t) \to \text{GA for max}$$

$\|\nabla_y f(x,y)\| \leq G$

Thm: GDA has conv. rate $O\left(\frac{GR}{\sqrt{T}}\right)$.

Proof: x is arb.

$$\|x_{t+1} - x\|^2 = \|x_t - x\|^2 - 2\eta \underbrace{\langle \nabla_x f(x_t, y_t), x_t - x \rangle}_{\text{Conv. of } f(\cdot, y_t)} + \eta^2 \underbrace{\|\nabla_x f(x_t, y_t)\|^2}_{G\text{-Lipschitz}}$$

$$\leq \|x_t - x\|^2 - 2\eta\left[f(x_t, y_t) - f(x, y_t)\right] + \eta^2 G^2$$

$$\boxed{y \text{ is arb.}} \quad \|y_{t+1}-y\|^2 = \|y_t - y\|^2 - 2\eta \underbrace{\langle \nabla_y f(x_t, y_t), y - y_t \rangle}_{\text{Concavity of } f(x_t,\cdot)} + \eta^2 \underbrace{\|\nabla_y f(x_t, y_t)\|^2}_{G\text{-Lipschitz}}$$

$$\leq \|y_t - y\|^2 - 2\eta \left[f(x_t, y) - f(x_t, y_t) \right] + \eta^2 G^2$$

$$\|x_{t+1} - x\|^2 + \|y_{t+1} - y\|^2 \leq \|x_t - x\|^2 + \|y_t - y\|^2 - 2\eta \left[f(x_t, y) - f(x, y_t) \right] + 2\eta^2 G^2.$$

$$\frac{1}{T+1} \sum_{t=0}^{T} \left[f(x_t, y) - f(x, y_t) \right] \leq \frac{\|x_0 - x\|^2 + \|y_0 - y\|^2}{2\eta (T+1)} + \eta G^2$$

$$f\Big(\underbrace{\tfrac{1}{T+1}\textstyle\sum_{t=0}^{T} x_t}_{\bar{x}_T}, y\Big) \leq \frac{1}{T+1} \sum_{t=0}^{T} f(x_t, y) \quad \text{and} \quad f\Big(x, \underbrace{\tfrac{1}{T+1}\textstyle\sum_{t=0}^{T} y_t}_{\bar{y}_T}\Big) \geq \frac{1}{T+1} \sum_{t=0}^{T} f(x, y_t)$$

$$f(\bar{x}_T, y) - f(x, \bar{y}_T) \leq \frac{\|x_0 - x\|^2 + \|y_0 - y\|^2}{2\eta(T+1)} + \eta G^2 .$$

$$\max_y f(\bar{x}_T, y) - \min_x f(x, \bar{y}_T) \leq \frac{\max_x \|x_0 - x\|^2 + \max_y \|y_0 - y\|^2}{2\eta(T+1)} + \eta G^2 .$$

$$\leq \frac{R^2}{\eta(T+1)} + \eta G^2$$

$$\eta = \frac{R}{G\sqrt{T+1}} \qquad \leq \frac{2GR}{\sqrt{T+1}} .$$

$(\bar{x}_T, \bar{y}_T)$ is an $\frac{2GR}{\sqrt{T+1}}$ - primal dual pair. ▣

[Prox] Proximal algorithm : $x_{t+1} = \arg\min_{x \in \mathcal{X}} \left[f(x) + \frac{1}{2\eta} \|x - x_t\|^2 \right]$

GD Step : $x_{t+1} = \arg\min_{x \in \mathcal{X}} \left[f(x_t) + \langle \nabla f(x_t), x - x_t \rangle + \frac{1}{2\eta} \|x - x_t\|^2 \right]$

Exercise: Prox has convergence rate $O(\frac{1}{T})$ for [possibly non-smooth] convex functions.

Exercise: Prox step is efficiently implementable for

L-Smooth functions and $\eta \leq \frac{1}{2L}$.

Efficient: Can find ϵ- approximate Soln. to the prox step in $O(\log \frac{1}{\epsilon})$ iterations.

Prox step: Given x, find w^* s.t.

$$w^* = \underset{z \in \mathcal{X}}{\arg\min} \; f(z) + \frac{1}{2\eta} \|x - z\|^2$$

$$\equiv \quad w^* = x - \eta \nabla f(w^*) \longrightarrow \text{Prox.}$$

$\nabla f(x) \longrightarrow$ GD

Algorithm to find Prox Step]. Pick $w_0 = x$

$$w_{t+1} = x - \eta \nabla f(w_t) \longrightarrow (1)$$

(A) Claim: w^* is the unique fixed point of (1)

(B) Claim: (1) is a $\frac{1}{2}$ Contraction.

$$\left.\begin{aligned} w &\longrightarrow w^+ = x - \eta \nabla f(w) \\ \tilde{w} &\longrightarrow \tilde{w}^+ = x - \eta \nabla f(\tilde{w}) \end{aligned}\right] \quad \begin{aligned} \|w^+ - \tilde{w}^+\| &= \eta \, \|\nabla f(w) - \nabla f(\tilde{w})\| \\ &\leq \eta L \, \|w - \tilde{w}\| \\ \text{If } \eta \leq \tfrac{1}{2L} \quad &\leq \tfrac{1}{2} \|w - \tilde{w}\|. \end{aligned}$$

(A) + (B) $\Longrightarrow$ (1) finds ϵ- approximate w^* in $\log \frac{1}{\epsilon}$ iterations.

Conceptual Mirror-Prox :

$$x_{t+1} = \underset{x \in \mathcal{X}}{\arg\min}\ f(x, y_{t+1}) + \frac{1}{2\eta} \|x - x_t\|^2$$

$$y_{t+1} = \underset{y \in \mathcal{Y}}{\arg\max}\ f(x_{t+1}, y) - \frac{1}{2\eta} \|y - y_t\|^2$$

Implementation of each step

$x_t^0 = x_t$

$y_t^0 = y_t$

$i = 1, \dots \underbrace{\log \frac{1}{\varepsilon}}$,

'2' in the actual mirrorProx alg.

$$x_t^i = \underset{x \in \mathcal{X}}{\arg\min}\ f(x, y_t^{i-1}) + \frac{1}{2\eta} \|x - x_t\|^2$$

$$y_t^i = \underset{y \in \mathcal{Y}}{\arg\max}\ f(x_t^{i-1}, y) - \frac{1}{2\eta} \|y - y_t\|^2.$$

<u>Exercise</u>: For conceptual Mirror-Prox, obtain $O(\frac{1}{T})$ convergence rate for L-smooth minimax opt.

<u>Exercise</u>: Show that the algorithm for implementing the Prox step converges in $O(\log \frac{1}{\epsilon})$ iterations.

<u>Exercise</u>: Careful accouting of all the terms to show that 2 inner steps suffice for $O(\frac{1}{T})$ convergence rate.