HW2-sol-Optimization in machine learning

Tian QIU

August 22, 2024

1 AGD Monotonically Decreases The Hamiltonian

By smooth:

$$
\begin{aligned}
f(x_{t+1}) - f(y_t) &\leq \frac{l}{2}\|x_{t+1} - y_t\|^2 + < \nabla f(y_t), x_{t+1} - y_t > \\
&\leq \frac{\eta}{2}\|\nabla f(y_t)\|^2 - \eta\|\nabla f(y_t)\|^2 \\
&= -\frac{\eta}{2}\|\nabla f(y_t)\|^2
\end{aligned}
$$

$$
\begin{aligned}
E_{t+1} - E_t &= f(x_{t+1}) + \frac{\|x_{t+1} - x_t\|^2}{2\eta} - f(x_t) - \frac{\|x_t - x_{t-1}\|^2}{2\eta} \\
&\leq f(y_t) - \frac{\eta}{2}\|\nabla f(y_t)\|^2 + \frac{\|x_{t+1} - x_t\|^2}{2\eta} - f(x_t) - \frac{\|x_t - x_{t-1}\|^2}{2\eta} \\
\text{(By convex)} &\leq < \nabla f(y_t), y_t - x_t > + \frac{\|x_{t+1} - x_t\|^2}{2\eta} - \frac{\|x_t - x_{t-1}\|^2}{2\eta} - \frac{\eta}{2}\|\nabla f(y_t)\|^2
\end{aligned}
$$

Decompose the norm:

$$
\begin{aligned}
\|x_{t+1} - x_t\|^2 &= \|r(x_t - x_{t-1}) - \eta\nabla f(y_t))\|^2 \\
&= [\gamma^2\|x_t - x_{t-1}\|^2 + \eta^2\|\nabla f(y_t)\|^2 - 2\eta\gamma < x_t - x_{t-1}, \nabla f(y_t) >] \\
&= [\gamma^2\|x_t - x_{t-1}\|^2 + \eta^2\|\nabla f(y_t)\|^2 - 2\eta < y_t - x_t, \nabla f(y_t) >]
\end{aligned}
$$

Plug in the decomposition:

$$
E_{t+1} - E_t \leq \frac{\gamma^2\|x_t - x_{t-1}\|^2}{2\eta} - \frac{\|x_t - x_{t-1}\|^2}{2\eta} \tag{1}
$$

In conclusion:

$$
E_{t+1} \leq E_t - \frac{(1 - \gamma^2\|x_t - x_{t-1}\|^2)}{2\eta} \tag{2}
$$

2 Lower Bounds for Smooth and Strongly Convex Functions

2.1 a

Calculate the Hessian of the function:

$$\nabla f(x) = \frac{l-\alpha}{8}[2(e_1^T x - 1)e_1 + \sum_{i=1}^{2t-1} 2(e_i - e_{i+1})^T x(e_i - e_{i+1}) + 2\zeta(e_{2t}^T x)e_{2t}] + \alpha x$$

$$\nabla^2 f(x) = \frac{l-\alpha}{4}[e_1 e_1^T + \sum_{i=1}^{2t-1}(e_i - e_{i+1})(e_i - e_{i+1})^T + \zeta e_{2t} e_{2t}^T] + \alpha I$$

The Hessian is in this form:

$$\frac{l-\alpha}{4}\begin{pmatrix} 2 & -1 & 0 & \ddots & \ddots & \ddots \\ -1 & 2 & -1 & \ddots & \ddots & \ddots \\ 0 & -1 & \ddots & \ddots & \ddots & \ddots \\ \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ \ddots & \ddots & \ddots & \ddots & 2 & -1 \\ \ddots & \ddots & \ddots & \ddots & -1 & \zeta+2 \end{pmatrix} + \alpha I \tag{3}$$

According to Gerschgorin's Theorem,the inner Matrix's eigenvalue:

$$\begin{aligned} &|\lambda_1^* - 2| \leq 1 \\ &|\lambda_i^* - 2| \leq 2 \quad (2 \leq i \leq 2t-1) \\ &|\lambda_{2t}^* - (\zeta+2)| \leq 1 \quad (\zeta \in [0,1]) \end{aligned}$$

So:

$$\begin{aligned} 1 \leq &\lambda_1^* \leq 3 \\ 0 \leq &\lambda_i^* \leq 4 \quad (2 \leq i \leq 2t-1) \\ 1 \leq \zeta + 1 \leq &\lambda_{2t}^* \leq \zeta + 3 \leq 4 \\ \Rightarrow \text{all} \quad &\lambda^* \in [0,4] \end{aligned}$$

In the final Hessian,all eigenvalue are :

$$\lambda = \frac{l-\alpha}{4}\lambda^* + \alpha \tag{4}$$

$$\Rightarrow \lambda(\nabla^2 f(x)) \in [\alpha, l] \tag{5}$$

which satisfies the l-smooth and α strong convex

2.2 b

The overall process is only to let $\nabla f(x) = 0$ and plug in $x^\star$ to calculate ζ The final ans is:

$$\zeta = \frac{2}{\sqrt{\kappa}+1} \tag{6}$$

2.3 c

To be done

3 Mirror Descent for Smooth Convex Functions

3.1 a

By ρ-strong convex of Φ:

$$\Phi(x) - \Phi(y) - < \nabla\Phi(y), x - y > \geqslant \frac{\rho}{2}||y - x||^2$$

which means:

$$D_\Phi(x, y) \geqslant \frac{\rho}{2}||y - x||^2$$

and similarly:

$$D_\Phi(y, x) \geqslant \frac{\rho}{2}||y - x||^2$$

So we easily get:

$$\frac{l}{2}||x_{s+1} - x_s||^2 = \frac{\rho}{2\eta}||x_{s+1} - x_s||^2 \leq \frac{1}{\eta}D_\Phi(x_{s+1}, x_s) \tag{7}$$

By l-smooth of function f:

$$f(x_{s+1}) - f(x_s) \leq < g_{x_s}, x_{s+1} - x_s > + \frac{l}{2}||x_{s+1} - x_s||^2 \tag{8}$$

For simplicity,the problem considers unconstrained case,with:

$$\nabla\Phi(x_{s+1}) = \nabla\Phi(x_s) - \eta g_{x_s} \tag{9}$$

Plug (9) in (8)

$$f(x_{s+1}) - f(x_s) \leq \frac{1}{\eta} < \nabla\Phi(x_s) - \nabla\Phi(x_{s+1}), x_{s+1} - x_s > + \frac{l}{2}||x_{s+1} - x_s||^2 \tag{10}$$

By definition of D_Φ:

$$< \nabla\Phi(x_s) - \nabla\Phi(x_{s+1}), x_{s+1} - x_s >= -D_\Phi(x_{s+1}, x_s) - D_\Phi(x_s, x_{s+1}) \tag{11}$$

Plug in (11) and (7)

$$f(x_{s+1}) - f(x_s) \leq -\frac{1}{\eta}(D_\Phi(x_{s+1}, x_s) + D_\Phi(x_s, x_{s+1})) + \frac{1}{\eta} D_\Phi(x_{s+1}, x_s) \quad (12)$$
$$(13)$$

Finally,we prove the descent lemma:

$$f(x_{s+1}) - f(x_s) \leq -\frac{1}{\eta} D_\Phi(x_s, x_{s+1}) \quad (14)$$

3.2 b

First,use the three-point lemma to simplify the right hand,then we want to prove:

$$f(x_s) - f(x^\star) \leq \frac{1}{\eta} < \nabla\Phi(x_{s+1}) - \nabla\Phi(x_s), x^\star - x_s > \quad (15)$$

Plug in the MD:

$$f(x_s) - f(x^\star) \leq \frac{1}{\eta} < -\eta g_{x_s}, x^\star - x_s > \quad (16)$$
$$f(x_s) - f(x^\star) \leq - < g_{x_s}, x^\star - x_s > \quad (17)$$

This is obviously correct by convexity of function f

3.3 c

Add up the lemma in section a and section b:

$$f(x_{s+1}) - f(x^\star) \leq \frac{1}{\eta}(D_\Phi(x^\star, x_s) - D_\Phi(x^\star, x_{s+1})) \quad (18)$$

Telescope:

$$\sum_{s=2}^{t} [f(x_s) - f(x^\star)] \leq \frac{1}{\eta}[D_\Phi(x^\star, x_1) - D_\Phi(x^\star, x_t)] \quad (19)$$

We know that $\eta = \rho/l$ and $D_\Phi \geq 0$, so we prove that:

$$\frac{1}{t-1} \sum_{s=2}^{t} [f(x_s) - f(x^\star)] \leq \frac{l}{\rho(t-1)} D_\Phi(x^\star, x_1) \quad (20)$$

3.4 d

From lemma in section a,we know that $f(x_s)$ monotonically decreases with s.So the left hand of 20 can be further limited:

$$\frac{1}{t-1}\sum_{s=2}^{t}[f(x_t)-f(x^\star)] \leq \frac{1}{t-1}\sum_{s=2}^{t}[f(x_s)-f(x^\star)] \tag{21}$$

That equals to: (22)

$$f(x_t)-f(x^\star) \leq \frac{1}{t-1}\sum_{s=2}^{t}[f(x_s)-f(x^\star)] \tag{23}$$

In conclusion:

$$f(x_t)-f(x^\star) \leq \frac{l}{\rho(t-1)} D_\Phi(x^\star, x_1) \tag{24}$$