7. Prove that denoising objective function is strongly convex.

$$f(x) = \frac{\lambda}{2}\|Dx\|_2^2 + \frac{1}{2}\|x - z_{noisy}\|_2^2 \quad ; D = D_h + iD_v$$

we need to show that

$$f(y) \geq f(x) + \nabla f(x)^T (y-x) + \frac{\mu}{2}\|y-x\|_2^2 \quad \text{for any } x, y \text{ and } \mu > 0 \quad (i)$$

or:

$$y^T \nabla^2 f(x) y \geq \mu \|y\|_2^2 \quad \forall x, y \in \mathbb{R}^n$$

Calculate the derivative and 2^{nd} order derivative

$$\nabla f(x) = \lambda \, \mathrm{real}(D^* D) x + x - z_{noisy}$$

$$= \lambda (D_h^T D_h + D_v^T D_v) + x - z_{noisy}$$

$$\nabla^2 f(x) = \lambda \, \mathrm{real}(D^* D) + I$$

$$= \lambda (D_h^T D_h + D_v^T D_v) + I$$

let z be the value $z = tx + (1-t)y$. $y \in [0, 1]$

According to Taylor's Theorem.

$$f(y) = f(x) + \nabla f(x)(y-x) + \frac{1}{2}(y-x)^T \nabla^2 f(z)(y-x) \quad (ii)$$

from (i) and (ii). we now need to prove

$$\frac{1}{2}(y-x)^T \nabla^2 f(z)(y-x) \geq \frac{\mu}{2}\|y-x\|_2^2 \iff$$

$$\Leftrightarrow \quad (y-x)^T \nabla^2 f(z)(y-x) - \mu (y-x)^T (y-x) \geq 0$$

$$\Leftrightarrow \quad f(x) - \frac{\mu}{2}\|x\|_2^2 \text{ is convex}$$

$$\left(\text{Assume } g(x) = f(x) - \frac{\mu}{2}\|x\|_2^2,\ \nabla^2 g(x) = \nabla^2 f(x) - \mu \right)$$

$$f(x) - \frac{\mu}{2}\|x\|_2^2 = \frac{\lambda}{2}\|Dx\|_2^2 + \frac{1}{2}\|x - z_{noisy}\|_2^2 - \frac{\mu}{2}\|x\|_2^2$$

$$\nabla^2 f(x) - \nabla^2\left(\frac{\mu}{2}\|x\|_2^2\right) = \lambda \, Real(D^* D) + I -$$

$$= \lambda (D_h^T D_h + D_v^T D_v) + I - \mu I$$

$$\Rightarrow \quad \lambda x^T (D_h^T D_h + D_v^T D_v) x + x^T x - \mu x^T x$$

$$= \lambda \|D_h x\|_2^2 + \lambda \|D_v x\|_2^2 + \|x\|_2^2 - \mu \|x\|_2^2 > 0$$

$$\mu < 1 + \frac{\lambda \|D_h x\|_2^2 + \lambda \|D_v x\|_2^2}{\|x\|_2^2}$$

$$\mu < 1 + \frac{\lambda \, real(Dx)}{\|x\|_2}$$

8. Armijo line-Search:

① Termination condition:

$$f(x_k - \alpha \nabla f(x_k)) \le f(x_k) - \alpha \gamma \|\nabla f(x_k)\|_2^2 \qquad \gamma \in (0, \tfrac{1}{2}]$$

$$f(x_k - \alpha \nabla f(x_k)) - f(x_k) + \alpha \gamma \|\nabla f(x_k)\|_2^2 \le 0$$

by FToC: let $y = x - \alpha \nabla f(x)$

$$f(y) \le f(x) + \nabla f(x)^T (y - x) + \frac{L}{2}\|y - x\|_2^2 \quad ; \quad y - x = \alpha \nabla f(x)$$

$$f(x - \alpha \nabla f(x)) \le f(x) + \nabla f(x)^T(-\alpha \nabla f(x)) + \frac{L}{2}\|\alpha \nabla f(x)\|_2^2$$

$$= f(x) - \alpha \|\nabla f(x)\|_2^2 + \frac{L\alpha^2}{2}\|\nabla f(x)\|_2^2$$

Adding $-f(x) + \alpha\gamma \|\nabla f(x)\|_2^2$ to both sides

$$f(x - \alpha \nabla f(x)) - f(x) + \alpha \gamma \|\nabla f(x)\|_2^2$$

$$\le \cancel{f(x)} - \alpha \|\nabla f(x)\|_2^2 + \frac{L\alpha^2}{2}\|\nabla f(x)\|_2^2 - \cancel{f(x)} + \alpha\gamma \|\nabla f(x)\|_2^2$$

$$= \left(\frac{L\alpha^2}{2} - \alpha + \alpha\gamma\right) \|\nabla f(x)\|_2^2$$

The termination condition is

$$f(x - \alpha \nabla f(x)) - f(x) + \alpha \gamma \|\nabla f(x)\|_2^2 < 0, \quad \text{which is}$$

$$\left(\frac{L\alpha^2}{2} - \alpha + \alpha\gamma\right)\|\nabla f(x)\|_2^2 < 0 \quad \text{which is}$$

$$\alpha\left(\frac{L}{2}\alpha + \gamma - 1\right) < 0, \quad \text{Since } \alpha > 0.$$

we have $\quad 0 < \alpha \le \dfrac{2(1-\gamma)}{L}$

② After k iterations, $\alpha_k = (\frac{1}{2})^k \alpha_0$

Assume $\alpha_0 = 1$ at the beginning.

$$\left(\frac{1}{2}\right)^k \leq \frac{2(1-\gamma)}{L}$$

Satisfies the termination condition.

let say $\left(\frac{1}{2}\right)^k = \frac{2(1-\gamma)}{L}$

$$k = \log_{1/2} \left(\frac{1}{2}\right)^k = \log_{1/2} \frac{2(1-\gamma)}{L}$$

$$= \frac{\ln \frac{2(1-\gamma)}{L}}{\ln 1/2}$$

9. gradient descent with armijo line search.

Termination Condition: $f(x_k) - f^* \le \varepsilon$.

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

by FToC: , Assume f is Lipschitz contineous.

$$f(y) \le f(x) + \nabla f(x)^T (y-x) + \frac{L}{2}\|y-x\|_2^2$$

let $y = x_{k+1}$ and $x = x_k$. ; $y - x = -\alpha \nabla f(x)$.

$$f(y) = f(x_k - \alpha \nabla f(x_k)) \le f(x) + \nabla f(x)^T(-\alpha \nabla f(x_k)) + \frac{L}{2}\|-\alpha \nabla f(x_k)\|_2^2$$

$$\alpha\|\nabla^2 f(x_k)\|_2^2 - \frac{L\alpha^2}{2}\|\nabla f(x_k)\|_2^2 \le f(x_k) - f(x_{k+1})$$

$$\|\nabla f(x_k)\|_2^2 \le \frac{2}{2\alpha - L\alpha^2}\left(f(x_k) - f(x_{k+1})\right)$$

by strong convexity:

Lecture 5, Slide 29.

$$f(x_t) - f^* \le \frac{2L\|x_0 - x^*\|_2^2}{t+1}$$

, which is equal to.

$$f(x_k) - f^* \le \frac{1}{2\mu}\|\nabla f(x_k)\|_2^2$$

Together we have.

$$f(x_k) - f^* \le \frac{1}{2\mu}\|\nabla f(x_k)\|_2^2 \le \frac{1}{\mu(2\alpha - L\alpha^2)}\left(f(x_k) - f(x_{k+1})\right)$$

$$f(x_{k+1}) \le -\mu(2\alpha - L\alpha^2)(f(x_k) - f^*) + f(x_k)$$

$$f(x_{k+1}) \le [1 - \mu(2\alpha - L\alpha^2)] f(x_k) + \mu(2\alpha - L\alpha^2) f^*$$

$$f(x_{k+1}) - f^* \le [1 - \mu(2\alpha - L\alpha^2)] f(x_k) + [\mu(2\alpha - L\alpha) - 1] f^*$$

$$\underline{f(x_{k+1}) - f^* \le [1 - \mu(2\alpha - L\alpha^2)] (f(x_k) - f^*)}$$

substitute $f(x_k) - f^*$ by this formula until $k=1$ we have

$$f(x_{k+1}) - f^* = [1 - \mu(2\alpha - L\alpha)]^{k} \cdot f(x_0) - f^*.$$

In order to guarantee the termination condition:

$$f(x_{k+1}) - f^* \le \varepsilon.$$

$$[1 - \mu(2\alpha - L\alpha^2)]^{k} (f(x_0) - f^*) \le \varepsilon$$

$$k \cdot \underline{\ln(1 - \mu(2\alpha - L\alpha^2))} \le \log\left(\frac{\varepsilon}{f(x_0) - f^*}\right)$$

$$1 - \mu(2\alpha - L\alpha^2) < 1$$

$$\ln(1 - \mu(2\alpha - L\alpha^2)) < 0$$

$$\Rightarrow \quad k > \frac{\ln\left(\frac{\varepsilon}{f(x_0) - f^*}\right)}{\ln(1 - \mu(2\alpha - L\alpha^2))}$$

from Q8. we have $0 < \alpha \le \dfrac{2(1-r)}{L}$ $\quad \in (0, \tfrac{1}{2}]$. such is. $\alpha < \dfrac{1}{L}$.

$$1 - (1-r) = r$$

$$\Rightarrow$$

$$2\alpha - L\alpha^2 < \frac{4(1-r)}{L} - \frac{4(1-r)^2}{L} = \frac{4}{L}(1-r)(1-(1-r))$$

$$= \frac{4}{L}(1-r) \cdot r.$$

$$k > \frac{\log\left(\frac{\varepsilon}{f(x)-f^*}\right)}{\log\left(1-\frac{4\mu(1-r)r}{L}\right)}$$

Assume the image can be represented by $n\times n$ matrix

D: $n^2 \times n^2$ $\quad$ X and Z_{noisy}: $n^2 \times 1$

$$f(x) = \frac{\lambda}{2}\|Dx\|_2^2 + \frac{1}{2}\|X - Z_{noisy}\|_2^2$$

$$\nabla f(x) = \underline{\lambda \, \mathrm{real}(D^*(Dx))} + X - Z_{noisy}$$

$\underline{D^*}\underline{(Dx)}$ is the most costly step.

Compute $Dx \sim O(n^4)$ as D is $n^2 \times n^2$. X is $n^2 \times 1$

Compute $D^*Dx \sim O(n^4)$ as D^* is $n^2 \times n^2$, Dx is $n^2 \times 1$

other add. substract are $O(n^2)$

For line-search algothrism. the worst case requires

$\frac{\ln \frac{2(1-r)}{L}}{\ln \frac{1}{2}}$ iterations. So in each step.

the complexity of line-search is $\frac{\ln\frac{2(1-\gamma)}{L}}{\ln \frac{1}{2}} O(n^4)$

Overall complexity for gradient descent is

$$\left[\underbrace{O(n^4)}_{\text{Complexity of Matrix multiplication}} \times \left(\underbrace{\frac{\ln\frac{2(1-\gamma)}{L}}{\ln\frac{1}{2}}}_{\text{line-search's worst case.}} + \underbrace{1}_{\text{gradient descent 1 iteration.}}\right)\right] \cdot \underbrace{\frac{\ln\left(\frac{\varepsilon}{f(x_0)-f^*}\right)}{\ln\left(1-\frac{4\mu(1-\gamma)\gamma}{L}\right)}}_{\text{gradient descent's worst case}}$$

10. Constant Seepsize: $\alpha = 1/L$. Slide 38

That is: $x_{k+1} = x_k - \frac{1}{L}$.

if $f(x)$ is Lipschitz contineous:

$$f(x_{k+1}) \le f(x_k) - \frac{1}{2L}\|\nabla f(x_k)\|_2^2$$

$$\Leftrightarrow \quad \|\nabla f(x_k)\|_2^2 \le 2L\left[f(x_k) - f(x_{k+1})\right]$$

Assume $f(x)$ is bounded by $f^* \le f(x)$, $\forall x \in \mathbb{R}^n$

The termination condition is $f(x_k) - f^* \le \varepsilon$

by strong convexity: lecture 5, slide 29

$$f(x_t) - f^* \le \frac{2L\|x_0 - x^*\|_2^2}{t+1}$$

, which is equal to.

$$f(x_k) - f^* \le \frac{1}{2\mu}\|\nabla f(x_k)\|_2^2 \qquad \text{①}$$

by FToC:

$$f(y) \le f(x) + \nabla f(x)^T(y-x) + \frac{L}{2}\|y-x\|_2^2$$

let $y = x_{k+1}$ and $x = x_k$; $y - x = -\alpha \nabla f(x)$.

$$f(y) = f(x_k - \alpha\nabla f(x_k)) \le f(x) + \nabla f(x)^T(-\alpha\nabla f(x_k)) + \frac{L}{2}\|-\alpha\nabla f(x_k)\|_2^2$$

$$\alpha\|\nabla^2 f(x_k)\|_2^2 - \frac{L\alpha^2}{2}\|\nabla f(x_k)\|_2^2 \le f(x_k) - f(x_{k+1})$$

$$\|\nabla f(x_k)\|_2^2 \le \frac{2}{2\alpha - L\alpha^2}\left(f(x_k) - f(x_{k+1})\right) \qquad \text{②}$$

Let $y = x_k - 1/L = x_{k+1}$, $x = x_k$.

$$f(x_k) - f^* \leq \frac{1}{2\mu} \| \nabla f(x_k) \| \leq \frac{1}{\mu(2\alpha - L\alpha^2)} [f(x_k) - f(x_{k+1})]$$

$$f(x_{k+1}) - f^* \leq [1 - \mu(2\alpha - L\alpha^2)] f(x_k) \leq \varepsilon.$$

$$\Rightarrow f(x_t) - f^* \leq [1 - \mu(2\alpha - L\alpha^2)]^t (f(x_0) - f^*) \leq \varepsilon.$$

$$t \cdot \log [1 - \mu(2\alpha - L\alpha^2)] \leq \frac{\varepsilon}{f(x_0) - f^*}$$

$$t \geq$$

10.

Assume $f(x)$ is Lipschitz contineous.

$$\Rightarrow \begin{cases} x_{k+1} = x_k - \frac{1}{L}\nabla f(x_k) \\ f(y) \leq f(x) + \nabla f(x)^T(y-x) + \frac{L}{2}\|y-x\|_2^2 \qquad \forall\, x, y \in \mathbb{R}^n \text{ — ①} \end{cases}$$

Constant Seep size : $\alpha = 1/L$.

$$① \Rightarrow f(x_{k+1}) \leq f(x_k) + \nabla f(x_k)^T\big(\underbrace{x_{k+1} - x_k}_{}\overset{-\frac{1}{L}\nabla f(x_k)}{}\big) + \frac{L}{2}\|\overset{-\frac{1}{L}\nabla f(x_k)}{x_{k+1} - x_k}\|$$

$$= f(x_k) - \frac{1}{L}\|\nabla f(x_k)\|_2^2 + \frac{1}{2L}\|\nabla f(x_k)\|_2^2$$

$$= f(x_k) - \frac{1}{2L}\|\nabla f(x_k)\|_2^2$$

$$\Rightarrow \|\nabla f(x_k)\|_2^2 \leq 2L\,(f(x_k) - f(x_{k+1})) \text{ —— ①}$$

Strongly convex:

$$f(x_k) - f^* \leq \frac{1}{2\mu}\|\nabla f(x_k)\|_2^2 \text{ —— ②}$$

Together. we have

$$f(x_k) - f^* \leq \frac{1}{2\mu}\|\nabla f(x_k)\|_2^2 \leq \frac{1}{\cancel{2}\mu}\cancel{2}L\left(f(x_k - f(x_{k-1})\right)$$

$$f(x_{k-1}) - f(x_k) \leq \frac{u}{L}[f(x_k) - f^*] - f(x_k)$$ Add $f(x_k) - f^*$ in both side

$$f(x_{k-1}) - f^* \leq \left(1 - \frac{u}{L}\right)\underline{\left(f(x_k) - f^*\right)}$$

Assume termination after t iteration

$$
\begin{aligned}
f(x_t) - f^* &\le \left(1 - \frac{\mu}{L}\right)\left(f(x_{t-1}) - f^*\right) \\
&\le \left(1 - \frac{\mu}{L}\right)\left(1 - \frac{\mu}{L}\right)\left(f(x_{t-2}) - f^*\right) \\
&\;\;\vdots \\
\Rightarrow\quad & \left(1 - \frac{\mu}{L}\right)^t \left(f(x_0) - f^*\right)
\end{aligned}
$$

$$
f(x_t) - f^* \le \underbrace{\left(1 - \frac{\mu}{L}\right)}_{<1}{}^{t} \underbrace{\left[f(x_0) - f^*\right]}_{>0.} \le \varepsilon.
$$

$$
\left(1 - \frac{\mu}{L}\right)^t \le \frac{\varepsilon}{f(x_0) - f^*} \qquad \ln\left(1 - \frac{\mu}{L}\right) < 0
$$

$$
t \ln\left(1 - \frac{\mu}{L}\right) \le \ln \frac{\varepsilon}{f(x_0) - f^*}
$$

$$
t \ge \frac{\ln(\varepsilon) - \ln(f(x_0) - f^*)}{\ln\left(1 - \frac{\mu}{L}\right)} = \frac{\ln(f(x_0) - f^*) - \ln \varepsilon}{-\ln\left(\frac{L-\mu}{L}\right)}
$$

$$
= \frac{\ln\left(\frac{f(x_0) - f^*}{\varepsilon}\right)}{\ln(L) - \ln(L-\mu)}
$$

$$
= \frac{\ln(f(x_0) - f^*)}{\ln(L) - \ln(L-\mu)} + \frac{\ln \frac{1}{\varepsilon}}{\ln(L) - \ln(L-\mu)}
$$

$$
= O\left(1/\varepsilon\right)
$$