

# ELE 538: Large-Scale Optimization Homework 3

Zachary Hervieux-Moore

Wednesday 25<sup>th</sup> April, 2018

**Exercise 1: Smoothness:**

- a) Show that  $f_\mu(\mathbf{x}) := \sqrt{\|\mathbf{x}\|_2^2 + \mu^2} - \mu$  is  $\frac{1}{\mu}$ -smooth.
- b) Show that  $f_\mu(\mathbf{x}) = \mu \log(\sum_{i=1}^n e^{x_i/\mu}) - \mu \log(n)$  ( $\mathbf{x} \in \mathbb{R}^n$ ) is  $\frac{1}{\mu}$ -smooth.

**Answer:**

- a) Let  $g(\mathbf{x}) = \sqrt{\|\mathbf{x}\|_2^2 + \mu^2}$ , then we have

$$\begin{aligned}\nabla f_\mu(\mathbf{x}) &= \frac{\mathbf{x}}{g(\mathbf{x})} \\ \nabla^2 f_\mu(\mathbf{x}) &= \frac{1}{g(\mathbf{x})} I - \frac{\mathbf{x}\mathbf{x}^T}{g(\mathbf{x})^3}\end{aligned}$$

Since  $I$  and  $\mathbf{x}\mathbf{x}^T$  commute, then the eigenvalues of their sum is simply the sum of their individual eigenvalues. Also,  $\mathbf{x}\mathbf{x}^T$  only has one non-zero eigenvalue, namely  $\|\mathbf{x}\|_2^2$ . Therefore, the eigenvalues have the form

$$\begin{aligned}\frac{1}{g(\mathbf{x})} - \frac{\|\mathbf{x}\|_2^2}{g(\mathbf{x})^3} &= \frac{\mu^2}{g(\mathbf{x})^3} \\ \text{or } \frac{1}{g(\mathbf{x})}\end{aligned}$$

Note that these are always positive. Maximizing either of these can be accomplished via simple calculus, we can treat  $\|\mathbf{x}\|_2^2$  as a univariable,

$$\begin{aligned}\max_{\mathbf{x}} \frac{\mu^2}{(\|\mathbf{x}\|_2^2 + \mu^2)^{3/2}} &= \frac{1}{\mu} \\ \text{or } \max_{\mathbf{x}} \frac{1}{\sqrt{\|\mathbf{x}\|_2^2 + \mu^2}} &= \frac{1}{\mu}\end{aligned}$$

We conclude that  $\nabla^2 f_\mu(\mathbf{x}) \preceq \frac{1}{\mu} I$

- b) Let  $y = \sum_{i=1}^n e^{x_i/\mu}$ , then we have

$$\begin{aligned}[\nabla f_\mu(\mathbf{x})]_i &= \frac{e^{x_i/\mu}}{y} \\ [\nabla^2 f_\mu(\mathbf{x})]_{ij} &= \begin{cases} \frac{\frac{1}{\mu}(e^{x_i/\mu}y - e^{2x_i/\mu})}{y^2}, & \text{if } i = j \\ \frac{-\frac{1}{\mu}e^{x_i/\mu + x_j/\mu}}{y^2}, & \text{if } i \neq j \end{cases}\end{aligned}$$

In matrix form, letting  $\mathbf{z} = [e^{x_1/\mu} \dots e^{x_n/\mu}]$

$$\nabla^2 f_\mu(\mathbf{x}) = \frac{1}{\mu y} \text{diag}(e^{x_1/\mu}, \dots, e^{x_n/\mu}) - \frac{1}{\mu y^2} \mathbf{z} \mathbf{z}^T$$

We again have two commuting matrices with the subtrahend being an outer product. Again, using the facts we used in part a), we get that the eigenvalues have the form

$$\begin{aligned} & \frac{1}{\mu y} e^{x_i/\mu} - \frac{1}{\mu y^2} \|\mathbf{z}\|_2^2 \\ & \text{or } \frac{1}{\mu y} e^{x_i/\mu} \end{aligned}$$

Where we note that  $y^2 \geq \|\mathbf{z}\|_2^2$  and so the eigenvalues are all non-negative. Furthermore, the maximal eigenvalue must come from the second form of eigenvalue. Then maximizing the second line above, with relatively simple calculus, we have that picking  $y = e^{x_i/\mu}$ , that is,  $x_i = c, x_j \rightarrow -\infty \forall i \neq j$  is the maximizer. Thus,

$$\sup_{\mathbf{x}} \frac{1}{\mu y} e^{x_i/\mu} = \frac{1}{\mu}$$

We conclude that  $\nabla^2 f_\mu(\mathbf{x}) \preceq \frac{1}{\mu} \mathbf{I}$

**Exercise 2: Nesterov's coefficient:** Recall that one version of Nesterov's momentum coefficients is generated by

$$\theta_{t+1} = \frac{1 + \sqrt{1 + 4\theta_t^2}}{2}$$

with  $\theta_0 = 1$ .

a) Show that for all  $t \geq 1$ , one has  $\theta_t \geq \frac{t+2}{2}$ .

b) Show that

$$\frac{\theta_t - 1}{\theta_{t+1}} = 1 - \frac{3}{t} + o\left(\frac{1}{t}\right)$$

**Answer:**

a) This is a simple proof by induction. Clearly, it is true for  $\theta_0 = 1 \geq \frac{2}{2} = 1$ . Now for the induction step,

$$\begin{aligned} \theta_{t+1} &= \frac{1 + \sqrt{1 + 4\theta_t^2}}{2} \\ &\geq \frac{1 + \sqrt{1 + (t+2)^2}}{2} && \text{by induction hypothesis} \\ &\geq \frac{\sqrt{(t+2)^2}}{2} && \text{by making the numerator smaller} \\ &\geq \frac{t+2}{2} \end{aligned}$$

b) We begin by just using definitions and manipulating

$$\begin{aligned} \frac{\theta_t - 1}{\theta_{t+1}} &= \frac{2\theta_t - 2}{1 + \sqrt{1 + 4\theta_t^2}} \\ &= \frac{(2\theta_t - 2)(\sqrt{1 + 4\theta_t^2} - 1)}{4\theta_t^2} \\ &= \frac{(\theta_t - 1)(\sqrt{1 + 4\theta_t^2} - 1)}{2\theta_t^2} \end{aligned}$$

Now, calculating the Laurent series of the above yields

$$\frac{\theta_t - 1}{\theta_{t+1}} = 1 - \frac{3}{2\theta_t} + o\left(\frac{1}{\theta_t}\right)$$

Now, by part a), we know that  $\frac{1}{\theta_t} = \frac{2}{t+2} + o\left(\frac{1}{t}\right) = \frac{2}{t} + o\left(\frac{1}{t}\right)$ , which concludes the proof.

**Exercise 3: Infimal convolution:** Define the infimal convolution of two functions  $f$  and  $g$  as

$$(f \square g)(\mathbf{x}) := \inf_{\mathbf{y}} \{f(\mathbf{y}) + g(\mathbf{x} - \mathbf{y})\}$$

with  $\theta_0 = 1$ .

- a) Show that the Fenchel conjugate obeys  $(f \square g)^* = f^* + g^*$
- b) Suppose  $f$  is convex. Use part a) to show that the Moreau envelope  $M_{\mu f}(\mathbf{x}) := \inf_{\mathbf{z}} \{f(\mathbf{z}) + \frac{1}{2\mu} \|\mathbf{x} - \mathbf{z}\|_2^2\}$  is  $\frac{1}{\mu}$ -smooth.

**Hint:** if a function  $f$  is  $\alpha$ -strongly convex, then  $f^*$  is  $\frac{1}{\alpha}$ -smooth.

**Answer:**

- a) By definition of the Fenchel conjugate,

$$\begin{aligned} (f \square g)^*(\boldsymbol{\theta}) &= \sup_{\mathbf{x}} \left\{ \langle \mathbf{x}, \boldsymbol{\theta} \rangle - \inf_{\mathbf{y}} \{f(\mathbf{y}) + g(\mathbf{x} - \mathbf{y})\} \right\} \\ &= \sup_{\mathbf{x}} \left\{ \langle \mathbf{x}, \boldsymbol{\theta} \rangle + \sup_{\mathbf{y}} \{-f(\mathbf{y}) - g(\mathbf{x} - \mathbf{y})\} \right\} \\ &= \sup_{\mathbf{x}, \mathbf{y}} \{ \langle \mathbf{x}, \boldsymbol{\theta} \rangle - f(\mathbf{y}) - g(\mathbf{x} - \mathbf{y}) + \langle \mathbf{y}, \boldsymbol{\theta} \rangle - \langle \mathbf{y}, \boldsymbol{\theta} \rangle \} \\ &= \sup_{\mathbf{x}, \mathbf{y}} \{ \langle \mathbf{y}, \boldsymbol{\theta} \rangle - f(\mathbf{y}) + \langle \mathbf{x} - \mathbf{y}, \boldsymbol{\theta} \rangle - g(\mathbf{x} - \mathbf{y}) \} \\ &= \sup_{\mathbf{y}} \{ \langle \mathbf{y}, \boldsymbol{\theta} \rangle - f(\mathbf{y}) \} + \sup_{\mathbf{z}} \{ \langle \mathbf{z}, \boldsymbol{\theta} \rangle - g(\mathbf{z}) \} \\ &= f^*(\boldsymbol{\theta}) + g^*(\boldsymbol{\theta}) \end{aligned}$$

Where the substitution  $\mathbf{z} = \mathbf{x} - \mathbf{y}$  is valid because  $\mathbf{z}$  can still span all  $\mathbb{R}^n$  for fixed  $\mathbf{y}$ .

- b) Given,  $M_{\mu f}(\mathbf{x}) := \inf_{\mathbf{z}} \{f(\mathbf{z}) + \frac{1}{2\mu} \|\mathbf{x} - \mathbf{z}\|_2^2\}$ , we let  $g(\mathbf{x}) = \frac{1}{2\mu} \|\mathbf{x}\|_2^2$ , then we have precisely that

$$M_{\mu f}(\mathbf{x}) = (f \square g)(\mathbf{x})$$

Using the hint, we calculate the Fenchel conjugate of the above by using part a),

$$M_{\mu f}^*(\mathbf{x}) = f^*(\mathbf{x}) + g^*(\mathbf{x})$$

Now, we calculate  $g^*$ ,

$$g^*(\boldsymbol{\theta}) = \sup_{\boldsymbol{x}} \{ \langle \boldsymbol{x}, \boldsymbol{\theta} \rangle - \frac{1}{2\mu} \|\boldsymbol{x}\|_2^2 \}$$

Taking the gradient and setting to 0 yields

$$\nabla \implies \boldsymbol{\theta} - \frac{1}{\mu} \boldsymbol{x} = 0 \implies \boldsymbol{x} = \mu \boldsymbol{\theta}$$

So,

$$g^*(\boldsymbol{\theta}) = \frac{\mu}{2} \|\boldsymbol{\theta}\|_2^2$$

and

$$M_{\mu f}^*(\boldsymbol{x}) = f^*(\boldsymbol{x}) + \frac{\mu}{2} \|\boldsymbol{x}\|_2^2$$

As  $f^*$  is convex, then we have that  $M_{\mu f}^*(\boldsymbol{x})$  is  $\mu$ -strongly convex by definition. Then, by the hint, we have  $M_{\mu f}(\boldsymbol{x})$  is  $\mu$ -smooth.

**Exercise 4: Moreau envelope:** Let  $f$  be a convex function.

- a) Suppose that the Moreau envelope  $M_f(\mathbf{x}) > -\infty$  for all  $\mathbf{x}$ . Show that  $M_f(\cdot)$  is convex.
- b) Show that  $\nabla M_{\mu f}(\mathbf{x}) = \frac{1}{\mu}(\mathbf{x} - \text{prox}_{\mu f}(\mathbf{x}))$ .
- c) Show that  $\text{prox}_f(\mathbf{x}) = \nabla M_{f^*}(\mathbf{x})$

**Answer:**

- a) Using the standard definition of convexity

$$\begin{aligned} M_f(\lambda \mathbf{x} + (1 - \lambda) \mathbf{y}) &= \inf_{\mathbf{z}} \{f(\mathbf{z}) + \frac{1}{2} \|\lambda \mathbf{x} + (1 - \lambda) \mathbf{y} - \mathbf{z}\|_2^2\} \\ &= \inf_{\mathbf{z}_1, \mathbf{z}_2} \{f(\lambda \mathbf{z}_1 + (1 - \lambda) \mathbf{z}_2) + \frac{1}{2} \|\lambda \mathbf{x} - (1 - \lambda) \mathbf{y} - \lambda \mathbf{z}_1 - (1 - \lambda) \mathbf{z}_2\|_2^2\} \end{aligned}$$

Where the infimum is not changed because one can pick  $\mathbf{z} = \lambda \mathbf{z}_1^* + (1 - \lambda) \mathbf{z}_2^*$ . Using convexity of  $f$  and triangle inequality,

$$\begin{aligned} &\leq \inf_{\mathbf{z}_1, \mathbf{z}_2} \left\{ \lambda f(\mathbf{z}_1) + \frac{1}{2} \|\lambda \mathbf{x} - \lambda \mathbf{z}_1\|_2^2 \right. \\ &\quad \left. + (1 - \lambda) f(\mathbf{z}_2) + \frac{1}{2} \|(1 - \lambda) \mathbf{y} - (1 - \lambda) \mathbf{z}_2\|_2^2 \right\} \end{aligned}$$

Since  $\lambda \in (0, 1)$  we get,

$$\begin{aligned} &\leq \inf_{\mathbf{z}_1, \mathbf{z}_2} \left\{ \lambda f(\mathbf{z}_1) + \frac{\lambda}{2} \|\mathbf{x} - \mathbf{z}_1\|_2^2 + (1 - \lambda) f(\mathbf{z}_2) + \frac{(1 - \lambda)}{2} \|\mathbf{y} - \mathbf{z}_2\|_2^2 \right\} \\ &= \lambda M_f(\mathbf{x}) + (1 - \lambda) M_f(\mathbf{y}) \end{aligned}$$

- b) We have that

$$\begin{aligned} M_{\mu f}(\mathbf{x}) &= \inf_{\mathbf{z}} \left\{ f(\mathbf{z}) + \frac{1}{2\mu} \|\mathbf{x} - \mathbf{z}\|_2^2 \right\} \\ &= -\sup_{\mathbf{z}} \left\{ -f(\mathbf{z}) - \frac{1}{2\mu} \|\mathbf{x} - \mathbf{z}\|_2^2 \right\} \\ &= -\sup_{\mathbf{z}} \left\{ -f(\mathbf{z}) - \frac{1}{2\mu} \|\mathbf{x}\|_2^2 + \frac{1}{\mu} \langle \mathbf{x}, \mathbf{z} \rangle - \frac{1}{2\mu} \|\mathbf{z}\|_2^2 \right\} \\ &= \frac{1}{2\mu} \|\mathbf{x}\|_2^2 - \frac{1}{\mu} \sup_{\mathbf{z}} \left\{ \langle \mathbf{x}, \mathbf{z} \rangle - \mu f(\mathbf{z}) - \frac{1}{2} \|\mathbf{z}\|_2^2 \right\} \end{aligned}$$

Define  $g(\mathbf{z}) = \mu f(\mathbf{z}) + \frac{1}{2}\|\mathbf{z}\|_2^2$  to get

$$M_{\mu f}(\mathbf{x}) = \frac{1}{2\mu}\|\mathbf{x}\|_2^2 - \frac{1}{\mu}g^*(\mathbf{x})$$

Thus, we have

$$\nabla M_{\mu f}(\mathbf{x}) = \frac{1}{\mu}(\mathbf{x} - \nabla g^*(\mathbf{x}))$$

Thus, we just have to show that  $\nabla g^*(\mathbf{x}) = \text{prox}_{\mu f}(\mathbf{x})$ . Since

$$g^*(\mathbf{x}) = \sup_{\mathbf{z}} \{\langle \mathbf{x}, \mathbf{z} \rangle - \mu f(\mathbf{z}) - \frac{1}{2}\|\mathbf{z}\|_2^2\}$$

Suppose that  $\mathbf{z}^*$  is the supremum, then

$$\begin{aligned} g^*(\mathbf{x}) &= \langle \mathbf{x}, \mathbf{z}^* \rangle - \mu f(\mathbf{z}^*) - \frac{1}{2}\|\mathbf{z}^*\|_2^2 \\ &\implies \nabla g^*(\mathbf{x}) = \mathbf{z}^* \\ &= \arg \max_{\mathbf{z}} \{\langle \mathbf{x}, \mathbf{z} \rangle - \mu f(\mathbf{z}) - \frac{1}{2}\|\mathbf{z}\|_2^2\} \\ &= \arg \min_{\mathbf{z}} \{f(\mathbf{z}) + \frac{1}{2\mu}\|\mathbf{z}\|_2^2 - \frac{1}{\mu}\langle \mathbf{x}, \mathbf{z} \rangle\} \end{aligned}$$

Adding the scalar  $\frac{1}{2\mu}\|\mathbf{x}\|_2^2$  does not change the arg max

$$\begin{aligned} g^*(\mathbf{x}) &= \arg \min_{\mathbf{z}} \{f(\mathbf{z}) + \frac{1}{2\mu}\|\mathbf{z}\|_2^2 - \frac{1}{\mu}\langle \mathbf{x}, \mathbf{z} \rangle + \frac{1}{2\mu}\|\mathbf{z}\|_2^2\} \\ &= \{f(\mathbf{z}) + \frac{1}{2\mu}\|\mathbf{x} - \mathbf{z}\|_2^2\} \\ &= \text{prox}_{\mu f}(\mathbf{x}) \end{aligned}$$

Thus, we have shown what we wanted to show.

c) From the previous part, we have

$$\nabla M_{f^*}(\mathbf{x}) = \mathbf{x} - \text{prox}_{f^*}(\mathbf{x})$$

By Moreau's decomposition, we have  $\mathbf{x} - \text{prox}_{f^*}(\mathbf{x}) = \text{prox}_f(\mathbf{x})$ . We conclude

$$\nabla M_{f^*}(\mathbf{x}) = \text{prox}_f(\mathbf{x})$$



**Exercise 5: ADMM and consensus optimization:** The general form ADMM problem is

$$\begin{aligned} \min & f(\mathbf{x}) + h(\mathbf{z}) \\ \text{s.t. } & \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} = \mathbf{b} \end{aligned}$$

with variables  $\mathbf{x} \in \mathbb{R}^n$ ,  $\mathbf{z} \in \mathbb{R}^m$ . Assume that both  $f$  and  $h$  are convex, and assume  $\mathbf{b} \in \mathbb{R}^p$  (i.e., there are  $p$  scalar equality constraints).

A special case is the (*twovariable*) consensus problem, which has the form

$$\begin{aligned} \min & f(\mathbf{x}) + h(\mathbf{z}) \\ \text{s.t. } & \mathbf{x} - \mathbf{z} = \mathbf{0} \end{aligned}$$

Show how to reduce the general form ADMM problem to the consensus problem. (In other words, explain how to solve the general form problem, assuming you can solve the consensus problem). Explain why the functions in your consensus problem are convex. (You do not have to worry about analysis details, such as showing that functions are closed). Explain the relation between ADMM for the general form problem and ADMM for the consensus problem that you form in your reduction.

**Hint:** define  $\tilde{f}(\tilde{\mathbf{x}}) = \inf_{\mathbf{A}\mathbf{x}=\tilde{\mathbf{x}}} f(\mathbf{x})$ .

**Answer:** We transform the general ADMM into the consensus problem by making the substitutions  $\tilde{f}(\tilde{\mathbf{x}}) = \inf_{\mathbf{A}\mathbf{x}=\tilde{\mathbf{x}}} f(\mathbf{x})$  and  $\tilde{h}(\tilde{\mathbf{z}}) = \inf_{\mathbf{b}-\mathbf{B}\mathbf{z}=\tilde{\mathbf{z}}} h(\mathbf{z})$ . These are convex functions because

$$\begin{aligned} \tilde{f}(\lambda\tilde{\mathbf{x}}_1 + (1-\lambda)\tilde{\mathbf{x}}_2) &= \inf_{\mathbf{A}\mathbf{x}=\lambda\tilde{\mathbf{x}}_1+(1-\lambda)\tilde{\mathbf{x}}_2} f(\mathbf{x}) \\ &= \inf_{\mathbf{A}(\lambda\mathbf{x}_1+(1-\lambda)\mathbf{x}_2)=\lambda\tilde{\mathbf{x}}_1+(1-\lambda)\tilde{\mathbf{x}}_2} f(\lambda\mathbf{x}_1+(1-\lambda)\mathbf{x}_2) \end{aligned}$$

Where the substitution  $\mathbf{x} = \lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2$  is valid because picking  $\mathbf{x} = \lambda\mathbf{x}_1^* + (1-\lambda)\mathbf{x}_2^*$  (the respective infima) is feasible in the original infimum. Now using convexity of  $f$

$$\begin{aligned} &\leq \inf_{\mathbf{A}(\lambda\mathbf{x}_1+(1-\lambda)\mathbf{x}_2)=\lambda\tilde{\mathbf{x}}_1+(1-\lambda)\tilde{\mathbf{x}}_2} \lambda f(\mathbf{x}_1) + (1-\lambda)f(\mathbf{x}_2) \\ &\leq \inf_{\substack{\mathbf{A}\mathbf{x}_1=\tilde{\mathbf{x}}_1 \\ \mathbf{A}\mathbf{x}_2=\tilde{\mathbf{x}}_2}} \lambda f(\mathbf{x}_1) + (1-\lambda)f(\mathbf{x}_2) \end{aligned}$$

Where the second inequality holds because the new problem is more restrictive. Finally, we conclude  $\tilde{f}(\tilde{\mathbf{x}})$  is convex with:

$$\begin{aligned} &= \inf_{\mathbf{Ax}_1=\tilde{\mathbf{x}}_1} \lambda f(\mathbf{x}_1) + \inf_{\mathbf{Ax}_2=\tilde{\mathbf{x}}_2} (1-\lambda)f(\mathbf{x}_2) \\ &= \lambda \tilde{f}(\tilde{\mathbf{x}}_1) + (1-\lambda)\tilde{f}(\tilde{\mathbf{x}}_2) \end{aligned}$$

$\tilde{h}(\tilde{\mathbf{z}})$  is convex by similar logic. Using these substitutes, the problem becomes

$$\begin{aligned} &\min \tilde{f}(\tilde{\mathbf{x}}) + \tilde{h}(\tilde{\mathbf{z}}) \\ &\text{s.t. } \tilde{\mathbf{x}} - \tilde{\mathbf{z}} = \mathbf{0} \end{aligned}$$

For completeness, the original ADMM updates are:

$$\begin{aligned} \mathbf{x}^{t+1} &= \arg \min_{\mathbf{x}} \{f(\mathbf{x}) + \frac{\rho}{2} \|\mathbf{Ax} + \mathbf{Bz}^t - \mathbf{b} + \frac{1}{\rho} \boldsymbol{\lambda}^t\|_2^2\} \\ \mathbf{z}^{t+1} &= \arg \min_{\mathbf{z}} \{h(\mathbf{z}) + \frac{\rho}{2} \|\mathbf{Ax}^{t+1} + \mathbf{Bz} - \mathbf{b} + \frac{1}{\rho} \boldsymbol{\lambda}^t\|_2^2\} \\ \boldsymbol{\lambda}^{t+1} &= \boldsymbol{\lambda}^t + \rho(\mathbf{Ax}^{t+1} + \mathbf{Bz}^{t+1} - \mathbf{b}) \end{aligned}$$

While the consensus problem's is:

$$\begin{aligned} \tilde{\mathbf{x}}^{t+1} &= \arg \min_{\tilde{\mathbf{x}}} \{\tilde{f}(\tilde{\mathbf{x}}) + \frac{\rho}{2} \|\tilde{\mathbf{x}} - \tilde{\mathbf{z}}^t + \frac{1}{\rho} \boldsymbol{\lambda}^t\|_2^2\} \\ \tilde{\mathbf{z}}^{t+1} &= \arg \min_{\tilde{\mathbf{z}}} \{\tilde{h}(\tilde{\mathbf{z}}) + \frac{\rho}{2} \|\tilde{\mathbf{x}}^{t+1} - \tilde{\mathbf{z}} + \frac{1}{\rho} \boldsymbol{\lambda}^t\|_2^2\} \\ \boldsymbol{\lambda}^{t+1} &= \boldsymbol{\lambda}^t + \rho(\tilde{\mathbf{x}}^{t+1} - \tilde{\mathbf{z}}^{t+1}) \end{aligned}$$

Notice that the update rules are much simpler on the surface for the consensus problem. However, that is because the  $\mathbf{Ax} = \tilde{\mathbf{x}}$  is hidden in  $\tilde{f}$ . Also, note that the consensus problem is merely the general problem but with  $\mathbf{A} = \mathbf{I}_n$ ,  $\mathbf{B} = \mathbf{I}_m$ ,  $\mathbf{b} = \mathbf{0}$ .

**Exercise 6: Douglas-Rachford splitting for LP feasibility:**

- a) Work out Douglas-Rachford splitting for solving the LP feasibility problem considered in Homework 2, problem 2.
- b) Implement it and report the numerical convergence of  $\|\mathbf{x}^t - \mathbf{x}^{t-\frac{1}{2}}\|_2$  vs.  $t$ . Compare it with alternating projection implemented in HW2.

**Answer:**

- a) To do the Douglas-Rachford splitting, we define the LP feasibility problem as

$$\min_{\mathbf{x}} f(\mathbf{x}) + h(\mathbf{A}\mathbf{x})$$

With  $f(\mathbf{x}) = 1_{\{0 \leq \mathbf{x}\}}$ , the indicator function with  $+\infty$  if outside and 0 inside, and  $h(\mathbf{x}) = \frac{1}{2}\|\mathbf{x} - \mathbf{b}\|_2^2$ . Which gets turned into the saddle-point problem

$$\min_{\mathbf{x}} \max_{\boldsymbol{\lambda}} f(\mathbf{x}) + \langle \boldsymbol{\lambda}, \mathbf{A}\mathbf{x} \rangle + h^*(\boldsymbol{\lambda})$$

Where

$$\begin{aligned} h^*(\boldsymbol{\lambda}) &= \sup_{\mathbf{z}} \left\{ \langle \boldsymbol{\lambda}, \mathbf{z} \rangle - \frac{1}{2}\|\mathbf{z} - \mathbf{b}\|_2^2 \right\} \\ &= \langle \boldsymbol{\lambda}, \mathbf{b} \rangle + \frac{1}{2}\|\boldsymbol{\lambda}\|_2^2 \end{aligned}$$

Where we used the optimality condition  $\mathbf{z} = \mathbf{b} + \boldsymbol{\lambda}$ .

So, all that is left is to compute the proximal operators for  $\eta f$  and  $\eta h^*$ .

$$\begin{aligned} \text{prox}_{\eta f}(\mathbf{x}) &= \arg \min_{\mathbf{z}} \left\{ \eta f(\mathbf{z}) + \frac{1}{2}\|\mathbf{x} - \mathbf{z}\|_2^2 \right\} \\ &= \arg \min_{\mathbf{z}} \left\{ \eta 1_{\{0 \leq \mathbf{z}\}} + \frac{1}{2}\|\mathbf{x} - \mathbf{z}\|_2^2 \right\} \end{aligned}$$

Which is trivially satisfied by  $\mathbf{z} = \mathbf{x}$  if  $\mathbf{x} \geq \mathbf{0}$ . If it is the other case, we must project onto the positive orthant (to get rid of the infinity from the indicator). Thus, we can write it succinctly as

$$\text{prox}_{\eta f}(\mathbf{x})_i = \max(0, x_i) \quad \forall i$$

For  $\eta h^*$  we have

$$\begin{aligned}\text{prox}_{\eta h^*}(\mathbf{x}) &= \arg \min_{\mathbf{z}} \left\{ \eta h^*(\mathbf{z}) + \frac{1}{2} \|\mathbf{x} - \mathbf{z}\|_2^2 \right\} \\ &= \arg \min_{\mathbf{z}} \left\{ \eta \langle \mathbf{z}, \mathbf{b} \rangle + \frac{\eta}{2} \|\mathbf{z}\|_2^2 + \frac{1}{2} \|\mathbf{x} - \mathbf{z}\|_2^2 \right\}\end{aligned}$$

Optimality conditions yields that  $\eta \mathbf{b} + \eta \mathbf{z} + \mathbf{z} - \mathbf{x} = 0$  or  $\mathbf{z} = \frac{\mathbf{x} - \eta \mathbf{b}}{1 + \eta}$ . Thus,

$$\text{prox}_{\eta h^*}(\mathbf{x}) = \frac{\mathbf{x} - \eta \mathbf{b}}{1 + \eta}$$

Putting it all together, the Douglas-Rachford splitting is

$$\begin{aligned}\mathbf{x}^{t+1/2} &= \max(0, p_i^t) \quad \forall i \\ \boldsymbol{\lambda}^{t+1/2} &= \frac{\mathbf{q}^t - \eta \mathbf{b}}{1 + \eta} \\ \begin{bmatrix} \mathbf{x}^{t+1} \\ \boldsymbol{\lambda}^{t+1} \end{bmatrix} &= \begin{bmatrix} \mathbf{I} & \eta \mathbf{A}^T \\ -\eta \mathbf{A} & \mathbf{I} \end{bmatrix}^{-1} \begin{bmatrix} 2\mathbf{x}^{t+1/2} - \mathbf{p}^t \\ 2\boldsymbol{\lambda}^{t+1/2} - \mathbf{q}^t \end{bmatrix} \\ \mathbf{p}^{t+1} &= \mathbf{p}^t + \mathbf{x}^{t+1} - \mathbf{x}^{t+1/2} \\ \mathbf{q}^{t+1} &= \mathbf{q}^t + \boldsymbol{\lambda}^{t+1} - \boldsymbol{\lambda}^{t+1/2}\end{aligned}$$

- b) The code used to generate the figure is attached below. Figure 1 shows the residual error for alternating projection vs. Douglas-Rachford. Figure 2 shows  $\|\mathbf{x}^t - \mathbf{x}^{t+1/2}\|_2$  for both methods. Notice the substantial increase in speed with characteristic ripples

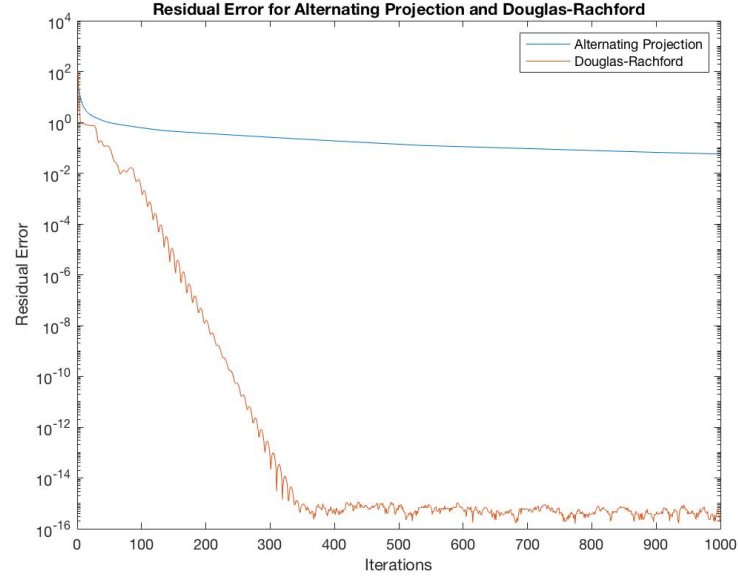


Figure 1: Plot of  $\|Ax^k - b\|$  for both methods

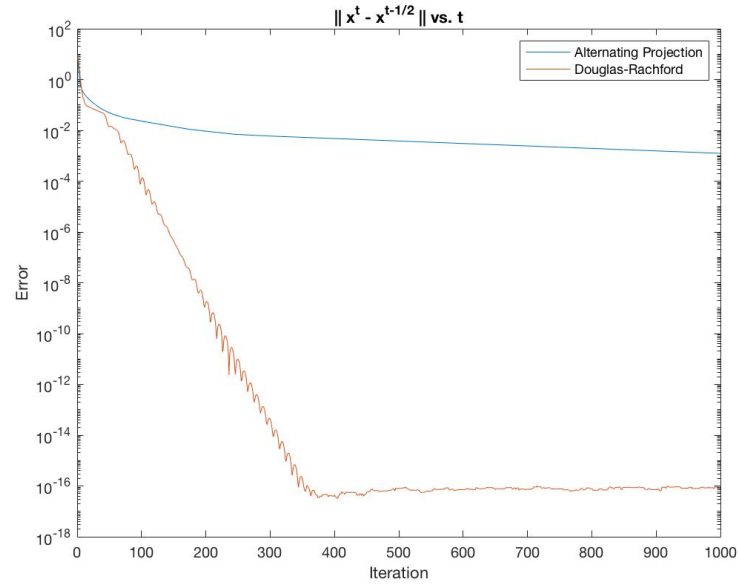


Figure 2: Plot of  $\|x^t - x^{t+1/2}\|_2$  for both methods

## Code Appendix:

```
clear;
clc;

m = 5;
n = 200;

A = rand(m,n);
b = rand(m,1);

% alternating projection
x_ap(:,1) = rand(n,1);
i = 1;
res_ap(1) = norm(x_ap(:,1),2);

while (size(res_ap) < 1000)
    i = i + 1;
    % project into affine set
    x_ap(:,i) = x_ap(:,i-1) + A'/(A*A')*(b-A*x_ap(:,i-1));

    i = i + 1;
    % project into positive orthant
    x_ap(:,i) = max(0, x_ap(:,i-1));

    % res_ap(ceil(i/2)) = norm(A*x_ap(:,i) - b,2);
    res_ap(ceil(i/2)) = norm(x_ap(:,i) - x_ap(:,i-1),2);
end

% Douglas-Rachford
x_dr = rand(n,1);
lambda_dr = rand(m,1);
p_dr = rand(n,1);
q_dr = rand(m,1);
eta = 0.1;
i = 1;
res_dr(1) = norm(x_dr(:,1),2);

while (size(res_dr) < 1000)
    % project onto positive orthant
    x_dr_half(:,i) = max(0, p_dr(:,i));
    % proximal for norm
    lambda_dr_half(:,i) = (q_dr(:,i) - eta*b)/(1+eta);

    % update full step
    tmp = inv([eye(n), eta*A'; -eta*A, eye(m)])*[2*x_dr_half(:,i)-p_dr(:,i); 2*lambda_dr_half(:,i)-q_dr(:,i)];
    x_dr(:,i+1) = tmp(1:n);
    lambda_dr(:,i+1) = tmp(n+1:n+m);

    % update p and q
    p_dr(:,i+1) = p_dr(:,i) + x_dr(:,i+1) - x_dr_half(:,i);
    q_dr(:,i+1) = q_dr(:,i) + lambda_dr(:,i+1) - lambda_dr_half(:,i);

    i = i + 1;
    % res_dr(i) = norm(A*x_dr(:,i) - b,2);
    res_dr(i) = norm(x_dr(:,i) - x_dr_half(:,i-1),2);
end
```

```
semilogy(res_ap)  
hold on  
semilogy(res_dr)
```