Network Science
IST-CMU PhD course
Fall 2016
Instructor: `jxavier@isr.ist.utl.pt`
TA: João Martins, `joaoa@andrew.cmu.edu`

Important: the homework is due the November 26. Send a scanned pdf file with your answers (or typed in LaTeX, if you prefer) to the TA's email.

Homework 3

1. *Brief analysis of ADMM.* Consider n agents linked by a static, undirected and connected communication graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and wishing to solve the optimization problem

$$\underset{x \in \mathbf{R}^d}{\text{minimize}} \quad \textstyle\sum_{v \in \mathcal{V}} f_v(x) \tag{1}$$

where $f_v : \mathbf{R}^d \to \mathbf{R} \cup \{+\infty\}$ is a convex function, only known at agent v.

In class, we saw that the Alternate Direction Method of Multipliers (ADMM) leads to the distributed algorithm:

$$x_v(t+1) = \arg\min_{x_v} f_v(x_v) + \lambda_v(t)^T x_v + \frac{\rho}{2} \sum_{u \sim v} \left\| x_v - \frac{x_v(t) + x_u(t)}{2} \right\|^2 \tag{2}$$

$$\lambda_v(t+1) = \lambda_v(t) + \rho \sum_{u \sim v} x_v(t+1) - x_u(t+1), \tag{3}$$

for $t \geq 0$. The variables $x_v(t)$ and $\lambda_v(t)$ are stored at agent v and are initialized as $\lambda_v(0) = 0$ and arbitrary $x_v(0)$. Recall that the notation $u \sim v$ means that u is a neighbor of v in the communication graph $\mathcal{G}$.

It is known that, under mild assumptions, the algorithm (2)–(3) converges to a solution of the problem (1): all iterates $x_v(t)$ converge to a point $x^\star$ that is a solution of (1). That proof is a bit involved. In this problem, you will prove a weaker statement: you will prove that *if* algorithm (2)–(3) converges, then it converges to the solution of (1). You can consider the scalar case $d = 1$ and assume that the convex functions f_v are finite-valued everywhere ($f_v : \mathbf{R} \to \mathbf{R}$) and continuously differentiable.

(a) Assume the algorithm (2)–(3) converges, that is, $x_v(t) \to x_v^\star$ and $\lambda_v(t) \to \lambda_v^\star$ for some $x_v^\star$ and $\lambda_v^\star$, for all $v \in \mathcal{V}$. Show that the limits $x_v^\star$, $v \in \mathcal{V}$, are all equal:

$$x_v^\star = x^\star$$

for some $x^\star$.

Hint: start by taking the limit $t \to \infty$ in (3) and use basic properties of the laplacian.

(b) Show that $x^\star$ is a solution of (1). Since all functions are convex and differentiable everywhere this boils down to show that

$$\sum_{v \in \mathcal{V}} \dot{f}_v(x^\star) = 0,$$

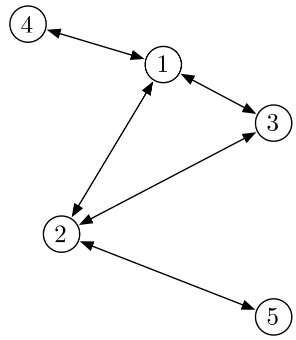

Figure 1: Undirected communication graph $\mathcal{G}(\mathcal{V},\mathcal{E})$ with nodes $\mathcal{V} = \{1,2,3,4,5\}$ and edges $\mathcal{E} = \{\{1,2\},\{1,3\},\{1,4\},\{2,3\},\{2,5\}\}$

.

where $\dot{f}_v$ is the derivative of f_v.

Hint: since each f_v is differentiable, note that equation (2) implies that $x_v(t+1)$ satisfies

$$\dot{f}_v\left(x_v(t+1)\right) + \lambda_v(t) + \rho \sum_{u \sim v} \left(x_v(t+1) - \frac{x_u(t) + x_v(t)}{2}\right) = 0$$

2. *Distributed algorithm via ADMM.* In class, we used the ADMM to build a distributed optimization algorithm for functions of the form (1). In this problem, you will use ADMM to derive a distributed algorithm for functions of the form

$$\underset{\{x_v\}_{v\in\mathcal{V}}}{\text{minimize}} \sum_{v\in\mathcal{V}} f_v\left(x_v\right) + \sum_{a\in\mathcal{A}} g_a\left(x_{S(a)}, x_{T(a)}\right). \tag{4}$$

Here, $\mathcal{A}$ denotes a set of arcs of the undirected communication graph $\mathcal{G} = (\mathcal{V},\mathcal{E})$ once we arbitrarily assign a direction to each edge in $\mathcal{E}$. For example, consider the undirected communication graph $\mathcal{G} = (\mathcal{V},\mathcal{E})$ in figure 1. By arbitrarily assigning a direction to each edge, we transform each edge into an arc and create a set of arcs $\mathcal{A}$ as in figure 2. Note that the underlying communication graph $\mathcal{G}$ is still undirected (agents can communicate in both directions over each edge). The set of arcs does not represent physical channels; it is just a useful device to express the cost function in (4). For each arc $a \in \mathcal{A}$, $S(a)$ is its source and $T(a)$ its sink; for example, in figure (2), we have $S(1,3) = 1$, $T(1,3) = 3$, $S(2,1) = 2$, $T(2,1) = 1$, and so on. In (1), all functions are assumed to be convex.

The structure (1) differs from the structure (4) in several aspects. In (1), all agents share the same variable x; in (4), each agent v has its own variable x_v. This means that, in (1), agents collaborate to agree on a solution of (1), say, $x^\star$; in particular, the solution $x^\star$ will eventually be known by all agents. In (4), agents collaborate to build a solution of (4), say, $\{x_v^\star\}_{v\in\mathcal{V}}$. But it is not required for agent v to know the optimal $x_u^\star$ for $u \neq v$: agent v only cares about knowing its optimal assignement $x_v^\star$. So, solving (4) in a distributed way means that agents should build a solution of (4) exchanging messages over the communication graph $\mathcal{G}$ and each agent v should eventually know an optimal $x_v^\star$.

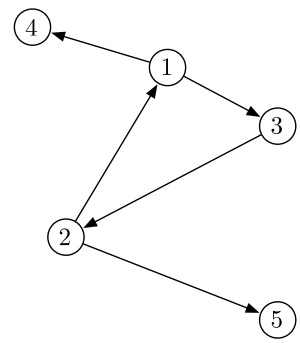

Figure 2: Graph from figure 1 after we transformed each edge into an arc. The arc set is $\mathcal{A} = \{(1,3),(1,4),(2,1),(3,2),(2,5)\}$.
.

We now motivate the structure (4). Consider a set of n robots linked by an undirected communication graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. Robot v wants to collect data (say, video images or chemical concentrations of some pollutants) from a source located at position $p_v \in \mathbf{R}^2$. The quality of the information collected by each robot depends on its distance from the source—the closer, the better. More precisely, let $x_v \in \mathbf{R}^2$ be the position of robot v. Assume that the quality of the data (say, signal-to-noise ratio) measured by robot v degrades as

$$\frac{1}{2}\left(x_v - p_v\right)^T \Sigma_v \left(x_v - p_v\right), \tag{5}$$

where Σ_v is a positive definite matrix that allows us to encode preferred spatial directions for collecting information from source v. We also want each robot to maintain its communication channels with neighbors in the graph $\mathcal{G}$. Assume that the channel $\{u,v\} \in \mathcal{E}$ exists only if the robots u and v are within each other's wireless range: if $\|x_u - x_v\| \le \min\{r_u, r_v\}$ where r_v is the wireless range of robot v. The problem of deciding optimal positions x_v for all the robots translates into the optimization problem

$$\underset{\{x_v\}_{v\in\mathcal{V}}}{\text{minimize}} \quad \sum_{v\in\mathcal{V}} \underbrace{\frac{1}{2}\left(x_v - p_v\right)^T \Sigma_v \left(x_v - p_v\right)}_{f_v(x_v)} + \sum_{a\in\mathcal{A}} \underbrace{\iota_{B(0,R_a)}\left(x_{S(a)} - x_{T(a)}\right)}_{g_a\left(x_{S(a)},x_{T(a)}\right)} \tag{6}$$

where $R_a = \min\{r_{S(a)}, r_{T(a)}\}$, and $\iota_{B(0,R)}$ is the indicator function of the set $B(0,R) = \left\{x \in \mathbf{R}^2 \,:\, \|x\| \le R\right\}$:

$$\iota_{B(0,R)}(x) = \begin{cases} 0 & , \text{ if } \|x\| \le R \\ +\infty & , \text{ otherwise.} \end{cases}$$

In (6), the first term attracts each robot to its source, and the second term enforces a maximum distance between robots that are linked by communication channels. Problem (6) is an instance of (4).

To apply the ADMM, we reformulate (4) as:

$$\begin{array}{lll} \underset{\{x_v\}_{v\in\mathcal{V}},\{y_a\}_{a\in\mathcal{A}},\{z_a\}_{a\in\mathcal{A}}}{\text{minimize}} & \sum_{v\in\mathcal{V}} f_v\left(x_v\right) + \sum_{a\in\mathcal{A}} g_a\left(y_a, z_a\right) & \\ \text{subject to} & x_{S(a)} = y_a, \quad a \in \mathcal{A} & \\ & x_{T(a)} = z_a, \quad a \in \mathcal{A} & \end{array} \tag{7}$$

and associate the lagrange multipliers $\{s_a\}_{a\in\mathcal{A}}$ and $\{t_a\}_{a\in\mathcal{A}}$ to the two set of constraints: the augmented lagrangian function is

$$\begin{aligned} & L\left(\{x_v\}_{v\in\mathcal{V}}, \{y_a\}_{a\in\mathcal{A}}, \{z_a\}_{a\in\mathcal{A}}; \{s_a\}_{a\in\mathcal{A}}, \{t_a\}_{a\in\mathcal{A}}\right) = \\ & \quad \sum_{v\in\mathcal{V}} f_v\left(x_v\right) + \sum_{a\in\mathcal{A}} g_a\left(y_a, z_a\right) \\ & \quad + \sum_{a\in\mathcal{A}} s_a^T \left(x_{S(a)} - y_a\right) + \frac{\rho}{2}\left\|x_{S(a)} - y_a\right\|^2 \\ & \quad + \sum_{a\in\mathcal{A}} t_a^T \left(x_{T(a)} - z_a\right) + \frac{\rho}{2}\left\|x_{T(a)} - z_a\right\|^2 \end{aligned}$$

where $\rho > 0$.

Applying the ADMM to (7) leads to

$$\{x_v(t+1)\}_{v\in\mathcal{V}} = \arg\min_{\{x_v\}_{v\in\mathcal{V}}} L\left(\{x_v\}_{v\in\mathcal{V}}, \{y_a(t)\}_{a\in\mathcal{A}}, \{z_a(t)\}_{a\in\mathcal{A}}; \{s_a(t)\}_{a\in\mathcal{A}}, \{t_a(t)\}_{a\in\mathcal{A}}\right) \tag{8}$$

$$\{y_a(t+1), z_a(t+1)\}_{a\in\mathcal{A}} = \arg\min_{\{y_a,z_a\}_{a\in\mathcal{A}}} L\left(\{x_v(t+1)\}_{v\in\mathcal{V}}, \{y_a\}_{a\in\mathcal{A}}, \{z_a\}_{a\in\mathcal{A}}; \{s_a(t)\}_{a\in\mathcal{A}}, \{t_a(t)\}_{a\in\mathcal{A}}\right) \tag{9}$$

$$s_a(t+1) = s_a(t) + \rho\left(x_{S(a)}(t+1) - y_a(t+1)\right), \quad a \in \mathcal{A}, \tag{10}$$

$$t_a(t+1) = t_a(t) + \rho\left(x_{T(a)}(t+1) - z_a(t+1)\right), \quad a \in \mathcal{A}. \tag{11}$$

You can assume that the optimization problems in (8) and (9) have unique solutions; so, the updates (8) and (9) are well-defined.

(a) Show that the update (8) boils down to

$$\begin{aligned} x_v(t+1) = {} & \arg\min_{x_v} f_v\left(x_v\right) + \left(\sum_{a\in\mathcal{S}(v)} s_a(t) + \sum_{a\in\mathcal{T}(v)} t_a(t)\right)^T x_v \\ & + \frac{\rho}{2}\sum_{a\in\mathcal{S}(v)} \|x_v - y_a(t)\|^2 + \frac{\rho}{2}\sum_{a\in\mathcal{T}(v)} \|x_v - z_a(t)\|^2 \end{aligned} \tag{12}$$

for all $v \in \mathcal{V}$, where $\mathcal{S}(v) = \{a \in \mathcal{A} : S(a) = v\}$ is the set of arcs that leave node v and $\mathcal{T}(v) = \{a \in \mathcal{A} : T(a) = v\}$ is the set of arcs that arrive at node v.

(b) Show that the update (9) boils down to

$$\begin{aligned} \left(y_a(t+1), z_a(t+1)\right) = {} & \arg\min_{(y_a,z_a)} g_a\left(y_a, z_a\right) + \frac{\rho}{2}\left\|y_a - \left(x_{S(a)}(t+1) + \frac{s_a(t)}{\rho}\right)\right\|^2 \\ & + \frac{\rho}{2}\left\|z_a - \left(x_{T(a)}(t+1) + \frac{t_a(t)}{\rho}\right)\right\|^2 \end{aligned} \tag{13}$$

for all $a \in \mathcal{A}$.

(c) Let agent v store the variables $x_v(t)$ and $s_a(t)$, $t_a(t)$, $y_a(t)$, $z_a(t)$ for $a \in \mathcal{S}(v) \cup \mathcal{T}(v)$. Show that (12) can be carried out at agent v and (13) can be carried out (simultaneously) by agents $S(a)$ and $T(a)$, if some communication steps are inserted between the computations: what should be communicated between neighbors at each iteration and when (relative to the updates (8) to (9))?