MS-E2122 - Nonlinear optimization — **Thu 22.11.2018**
Exercise 9 — **U149 (U6)**

Project Work 2 is online https://mycourses.aalto.fi/mod/assign/view.php?id=389823 and is due no later than **Sunday 16.12.2018 23:55**. Homework will be published ASAP.

Exercise 9.1 ADMM and Scaled Form ADMM

In this exercise, we derive a scaled form for the Alternating Direction Method of Multipliers (ADMM). Let $f : \mathbf{R}^n \to \mathbf{R}$ and $g : \mathbf{R}^m \to \mathbf{R}$ be convex functions. Consider the following optimization problem

$$\begin{aligned} \underset{x,z}{\text{minimize}} \quad & f(x) + g(z) && (1) \\ \text{subject to} \quad & Ax + Bz = c && (2) \end{aligned}$$

with variables $x \in \mathbf{R}^n$ and $z \in \mathbf{R}^m$. Assume that the problem data is $A \in \mathbf{R}^{p\times n}$, $B \in \mathbf{R}^{p\times m}$, and $c \in \mathbf{R}^p$. Notice that the objective function has two independent sets of variables x and z. Let us define the augmented Lagrangian of (1) – (2) as

$$L_\rho(x, z, y) = f(x) + g(z) + y^\top(Ax + Bz - c) + \frac{\rho}{2}||Ax + Bz - x||_2^2. \quad (3)$$

with dual variables $y \in \mathbf{R}^p$ and penalty parameter $\rho > 0$. The augmented Lagrangian (3) can be seen as the (unaugmented) Lagrangian of the problem

$$\begin{aligned} \underset{x,z}{\text{minimize}} \quad & f(x) + g(z) + \frac{\rho}{2}||Ax + Bz - c||_2^2 && (4) \\ \text{subject to} \quad & Ax + Bz = c && (5) \end{aligned}$$

The problem (4) – (5) is equivalent to the problem (1) – (2): for any feasible solution (x, z), the additional term in the objective (4) evaluates to zero. Solving the augmented Lagrangian (3) by ADMM consists of the following iterations

$$\begin{aligned} x^{k+1} &= \underset{x}{\text{argmin}}\, L_\rho(x, z^k, y^k) && (6) \\ z^{k+1} &= \underset{z}{\text{argmin}}\, L_\rho(x^{k+1}, z, y^k) && (7) \\ y^{k+1} &= y^k + \rho(Ax^{k+1} + Bz^{k+1} - c) && (8) \end{aligned}$$

(a) Motivate a suitable stopping criterion for the ADMM iterations (6) – (8).

(b) Derive the *scaled form* for the ADMM iterations (6) – (8) by defining the *primal residual* r and the *scaled dual variables* u as

$$r = Ax + Bz - c \quad \text{and} \quad u = \frac{1}{\rho}y \quad (9)$$

Hint: Apply the definitions or r and u to (3) and rewrite the ADMM iterations (6) – (8) by replacing the original dual variables y by their scaled counterparts u.

(c) Verify that the stopping criterion derived in part (a) is equivalent for the scaled version of the ADMM derived in part (b).

Solution.

(a) Let us motivate a stopping criteria for the unscaled form of the ADMM iterations (6) – (8). The optimality conditions for the problem (1) – (2) are

$$\begin{aligned} & Ax^* + Bz^* - c = 0 && \text{(primal feasibility)} && (10) \\ & \nabla_x f(x^*) + A^\top y^* = 0 && \text{(dual feasibility 1)} && (11) \\ & \nabla_z g(z^*) + B^\top y^* = 0 && \text{(dual feasibility 2)} && (12) \end{aligned}$$

Since z^{k+1} minimizes $L_\rho(x^{k+1}, z, y^k)$, taking the gradient of (7) with respect to z, we have

$$\begin{aligned}0 &= \nabla g_z(z^{k+1}) + B^\top y^k + \rho B^\top (Ax^{k+1} + Bz^{k+1} - c)\\ &= \nabla g_z(z^{k+1}) + B^\top (y^k + \rho(Ax^{k+1} + Bz^{k+1} - c))\\ &= \nabla g_z(z^{k+1}) + B^\top y^{k+1}\end{aligned}$$

Thus, the $z-$step (7) always satisfies the dual feasibility condition (12). Similarly, since x^{k+1} minimizes $L_\rho(x, z^k, y^k)$, taking the gradient of (6) with respect to x, we have

$$\begin{aligned}0 &= \nabla_x f(x^{k+1}) + A^\top y^k + \rho A^\top (Ax^{k+1} + Bz^k - c)\\ &= \nabla_x f(x^{k+1}) + A^\top (y^k + \rho(Ax^{k+1} + Bz^k - c))\\ &= \nabla_x f(x^{k+1}) + A^\top (y^k + \rho(Ax^{k+1} + Bz^{k+1} - c + Bz^k - Bz^{k+1}))\\ &= \nabla_x f(x^{k+1}) + A^\top (y^k + \rho(Ax^{k+1} + Bz^{k+1} - c) + \rho B(z^k - z^{k+1}))\\ &= \nabla_x f(x^{k+1}) + A^\top y^{k+1} + \rho A^\top B(z^k - z^{k+1}),\end{aligned}$$

or equivalently

$$\rho A^\top B(z^{k+1} - z^k) = \nabla f(x^{k+1}) + A^\top y^{k+1} \tag{13}$$

Comparing (13) to (11), we have an additional term called *dual residual* which is defined as

$$s^{k+1} = \rho A^\top B(z^{k+1} - z^k)$$

To cover the primal feasibility condition (10), we define the primal residual as

$$r^{k+1} = Ax^{k+1} + By^{k+1} - c$$

A reasonable stopping condition for the ADMM (6) – (8) can be defined as the sum of the primal and dual residual norms:

$$||r^{k+1}||_2 + ||s^{k+1}||_2 < \epsilon$$

for some tolerance $\epsilon > 0$.

(b) Using the definition of the primal residual $r = Ax + Bz - c$, we can first rewrite (3) as

$$L_\rho(x, z, y) = f(x) + g(z) + y^\top r + \frac{\rho}{2}||r||_2^2 \tag{14}$$

and proceed by writing the last two terms of (14) as

$$\begin{aligned}y^\top r + \frac{\rho}{2}||r||_2^2 &= \frac{\rho}{2}||r||_2^2 + y^\top r + \frac{1}{2\rho}||y||_2^2 - \frac{1}{2\rho}||y||_2^2\\ &= \frac{\rho}{2}||r + \frac{1}{\rho}y||_2^2 - \frac{1}{2\rho}||y||_2^2 \qquad (15)\end{aligned}$$

Now, by replacing the dual variables y with the scaled dual variables $u = (1/\rho)y$ (or $y = \rho u$), we can rewrite (15) as

$$\begin{aligned}y^\top r + \frac{\rho}{2}||r||_2^2 &= \frac{\rho}{2}||r + u||_2^2 - \frac{1}{2\rho}||\rho u||_2^2\\ &= \frac{\rho}{2}||r + u||_2^2 - \underbrace{\frac{\rho}{2}||u||_2^2}_{\text{constant}}\\ &= \frac{\rho}{2}||r + u||_2^2 + K \qquad (16)\end{aligned}$$

where $K = -(\rho/2)||u||_2^2$ is a constant. We can thus rewrite (14) as

$$\begin{aligned} L_\rho(x,z,u) &= f(x) + g(z) + y^\top r + \frac{\rho}{2}||r||_2^2 \\ &= f(x) + g(z) + \frac{\rho}{2}||r+u||_2^2 + K \end{aligned} \tag{17}$$

Now, by using (17), we can finally rewrite the ADMM iterations (6) – (8) as

$$x^{k+1} = \underset{x}{\operatorname{argmin}} \left\{ f(x) + \frac{\rho}{2}||Ax + Bz^k - c + u^k||_2^2 \right\} \tag{18}$$

$$z^{k+1} = \underset{z}{\operatorname{argmin}} \left\{ g(z) + \frac{\rho}{2}||Ax^{k+1} + Bz - c + u^k||_2^2 \right\} \tag{19}$$

$$u^{k+1} = u^k + Ax^{k+1} + Bz^{k+1} - c \tag{20}$$

Notice that we do not have the penalty parameter ρ in (20) as in (8) since we defined $y = \rho u$. We also do not have the constant K in the formulas, since it does not affect the solutions of the iterations. We only need to add it to the final objective value. The motivation for using the scaled form of ADMM is that formulas related to the scaled version of ADMM are typically shorter and easier to interpret.

(c) First, let us verify that the $z-$step is always dual feasible as in part (a). To this end, since z^{k+1} minimizes $L_\rho(x^{k+1}, z, u^k)$, taking the gradient of (19) with respect to z, we have

$$\begin{aligned} 0 &= \nabla g_z(z^{k+1}) + \rho B^\top(Ax^{k+1} + Bz^{k+1} - c + u^k) \\ &= \nabla g_z(z^{k+1}) + \rho B^\top u^{k+1} \\ &= \nabla g_z(z^{k+1}) + B^\top y^{k+1} \end{aligned}$$

Thus, the $z-$step (19) always satisfies the dual feasibility condition (12). Similarly, since x^{k+1} minimizes $L_\rho(x, z^k, u^k)$, taking the gradient of (18) with respect to x, we have

$$\begin{aligned} 0 &= \nabla_x f(x^{k+1}) + \rho A^\top(Ax^{k+1} + Bz^k - c + u^k) \\ &= \nabla_x f(x^{k+1}) + \rho A^\top(Ax^{k+1} + Bz^{k+1} - c + Bz^k - Bz^{k+1})) \\ &= \nabla_x f(x^{k+1}) + \rho A^\top u^k + \rho A^\top B(z^k - z^{k+1})), \end{aligned}$$

or equivalently

$$\begin{aligned} \rho A^\top B(z^{k+1} - z^k) &= \nabla_x f(x^{k+1}) + \rho A^\top u^{k+1} \\ &= \nabla_x f(x^{k+1}) + A^\top y^{k+1} \end{aligned} \tag{21}$$

Comparing (21) to (13), we get the same dual residual term as in the unscaled version, defined as

$$s^{k+1} = \rho A^\top B(z^{k+1} - z^k)$$

Finally, the primal residual does not depend on the u variable and remains the same

$$r^{k+1} = Ax^{k+1} + By^{k+1} - c$$

Exercise 9.2 ADMM for Quadratic Optimization Problems

Consider the following standard form quadratic optimization problem

$$\underset{x}{\text{minimize}} \quad \frac{1}{2}x^\top Px + q^\top x \tag{22}$$

$$\text{subject to} \quad Ax = b \tag{23}$$

$$x \geq 0 \tag{24}$$

with variables $x \in \mathbf{R}^n$. Assume that $P \in \mathbf{R}^{n\times n}$ is a symmetric positive semidefinite matrix, $q \in \mathbf{R}^n$, $A \in \mathbf{R}^{p\times n}$, and $b \in \mathbf{R}^p$. We can express the problem (22) – (24) in ADMM form as

$$\underset{x,z}{\text{minimize}} \quad f(x) + g(z) \tag{25}$$

$$\text{subject to} \quad x = z \tag{26}$$

where

$$f(x) = \frac{1}{2}x^\top P x + q^\top x \qquad \text{with} \qquad \mathbf{dom}\, f = \{x \in \mathbf{R}^n : Ax = b\}$$

is the original objective with a restricted domain, and $g : \mathbf{R}^n \to \{0, \infty\}$ is the indicator function of the nonnegative orthant $\mathbf{R}^n_+$ corresponding to the constraint $x \geq 0$. Write the augmented Lagrangian for (25) – (26) using the scaled dual variables and the corresponding scaled form of ADMM iterations using the results of Exercise 9.1.

Solution.

According to (17) of Exercise 9.1, the augmented Lagrangian for (25) – (26) using the scaled dual variables is of the form

$$\begin{aligned} L_\rho(x, z, u) &= f(x) + g(z) + \frac{\rho}{2}||r + u||_2^2 + K \\ &= f(x) + g(z) + \frac{\rho}{2}||x - z + u||_2^2 + K \end{aligned}$$

where $r = x - z$ is the *primal residual*, $u = (1/\rho)y$ are the scaled dual variables, and $K = -(\rho/2)||u||_2^2$ is a constant. Now we can write the scaled form of ADMM, which consists of the following iterations:

$$x^{k+1} = \underset{x:Ax=b}{\text{argmin}} \left\{ f(x) + \frac{\rho}{2}||x - z^k + u^k||_2^2 \right\} \tag{27}$$

$$z^{k+1} = \underset{z}{\text{argmin}} \left\{ g(z) + \frac{\rho}{2}||x^{k+1} - z + u^k||_2^2 \right\} \tag{28}$$

$$u^{k+1} = u^k + x^{k+1} - z^{k+1} \tag{29}$$

As $g(z)$ is the indicator function which takes value $g(z) = 0$ if $z = x \geq 0$ and $g(z) = \infty$ if $z = x < 0$, we can simplify the $z-$update (28). Taking the gradient of (28) and setting it to zero, we get

$$\nabla g(z) - \rho(x^{k+1} - z + u^k) = 0$$

Since $\nabla g(z) = 0$, we get $x^{k+1} - z + u^k = 0$ and the optimal next iterate z^{k+1} becomes

$$z^{k+1} = (x^{k+1} + u^k)_+ \tag{30}$$

where $z^{k+1} = (x^{k+1} + u^k)_+$ is the euclidean projection of $(x^{k+1} + u^k)$ to $\mathbf{R}^n_+$:

$$z_i^{k+1} = \begin{cases} x_i^{k+1} + u_i^k, & \text{if } x_i^{k+1} + u_i^k > 0 \\ 0, & \text{otherwise} \end{cases}$$

for all $i = 1, \ldots, n$. Using the notation (30), we can simplify (27) – (29) as

$$x^{k+1} = \underset{x:Ax=b}{\text{argmin}} \left\{ f(x) + \frac{\rho}{2}||x - z^k + u^k||_2^2 \right\} \tag{31}$$

$$z^{k+1} = (x^{k+1} + u^k)_+ \tag{32}$$

$$u^{k+1} = u^k + x^{k+1} - z^{k+1} \tag{33}$$

Notice that the $x-$update in (31) corresponds to computing the following equality constrained least squares problem

$$\begin{aligned} x^{k+1} \; = \; & \underset{x}{\operatorname{argmin}} \quad f(x) + \frac{\rho}{2}||x - z^k + u^k||_2^2 \\ & \text{subject to} \quad Ax = b \end{aligned}$$

and the value of x^{k+1} at each iteration can be obtained by solving the following system corresponding to the KKT optimality conditions

$$\begin{bmatrix} P + \rho I & A^\top \\ A & 0 \end{bmatrix} \begin{bmatrix} x^{k+1} \\ v \end{bmatrix} + \begin{bmatrix} q - \rho(z^k - u^k) \\ -b \end{bmatrix} = 0$$

where $v \in \mathbf{R}^p$ are the dual variables for relaxing $Ax = b$ in Lagrangian fashion.

The optimality conditions for (25) – (26) with the unscaled dual variables $y \in \mathbf{R}^n$ with are

$$x^* - z^* = 0 \qquad \text{(primal feasibility)} \tag{34}$$
$$\nabla_x f(x^*) + y^* = 0 \qquad \text{(dual feasibility 1)} \tag{35}$$
$$\nabla_z g(z^*) - y^* = 0 \qquad \text{(dual feasibility 2)} \tag{36}$$

Since z^{k+1} minimizes $L_\rho(x^{k+1}, z, u^k)$, by taking the gradient of (32) with respect to z, we have

$$\begin{aligned} 0 &= \nabla_z g(z^{k+1}) - \rho(x^{k+1} - z^{k+1} + u^k) \\ &= \nabla_z g(z^{k+1}) - \rho u^{k+1} \\ &= \nabla_z g(z^{k+1}) - y^{k+1} \end{aligned}$$

Thus, the $z-$step (32) always satisfies the dual feasibility condition (36). Similarly, since x^{k+1} minimizes $L_\rho(x, z^k, u^k)$, by taking the gradient of (31), we have

$$\begin{aligned} 0 &= \nabla_x f(x^{k+1}) + \rho(x^{k+1} - z^k + u^k) \\ &= \nabla_x f(x^{k+1}) + \rho(x^{k+1} - z^{k+1} + u^k - z^k + z^{k+1}) \\ &= \nabla_x f(x^{k+1}) + \rho u^{k+1} + \rho(z^{k+1} - z^k) \\ &= \nabla_x f(x^{k+1}) + y^{k+1} + \rho(z^{k+1} - z^k) \end{aligned} \tag{37}$$

Comparing (37) to (35), we have an additional term corresponding to the *dual residual* which is defined as

$$s^{k+1} = \rho(z^{k+1} - z^k)$$

In this problem, the primal residual is defined as

$$r^{k+1} = x^{k+1} - z^{k+1}$$

And we can again use

$$||r^{k+1}||_2 + ||s^{k+1}||_2 < \epsilon$$

as a stopping condition, for example.

Exercise 9.3 ADMM for Stochastic Linear Optimization Problems

Consider the following two-stage stochastic linear optimization problem

$$\zeta = \underset{x}{\text{minimize}} \; \{c^\top x + \mathcal{Q}(x) : x \in X\}, \tag{38}$$

with variables $x \in \mathbf{R}^{n_x}$ and known first-stage costs $c \in \mathbf{R}^{n_x}$. The set X consists of linear constraints on the variables x and the function $\mathcal{Q} : \mathbf{R}^n \to \mathbf{R}$ is the *expected recourse value*

$$\mathcal{Q}(x) = \mathbf{E}_\xi \left[\underset{y}{\text{minimize}} \left\{ q(\xi)^\top y : W(\xi)y = h(\xi) - T(\xi)x,\ y \in Y(\xi) \right\} \right]. \tag{39}$$

Suppose that the random variable ξ is part of a discrete distribution indexed by a finite set $\mathcal{S}$, consisting of realizations $\xi_1, \dots, \xi_{|S|}$, corresponding to realization probabilities $p_1, \dots, p_{|S|}$. Each realization ξ_s of ξ is called a *scenario* and encodes realizations observed by the random elements

$$(q(\xi_s),\, h(\xi_s),\, W(\xi_s),\, T(\xi_s),\, Y(\xi_s))$$

To simplify notation, we refer to this collection of random elements respectively as

$$(q_s,\, h_s,\, W_s,\, T_s,\, Y_s)$$

For each scenario $s \in S$, the set Y_s consists of linear constraints on the variables $y_s \in \mathbf{R}^{n_y}$. We can reformulate problem (38) as an *equivalent deterministic problem*

$$\zeta = \underset{x,y}{\text{minimize}} \left\{ c^\top x + \sum_{s \in \mathcal{S}} p_s q_s^\top y_s : (x, y_s) \in K_s\ \forall s \in S \right\}, \tag{40}$$

where

$$K_s = \left\{ (x, y_s) : W_s y_s = h_s - T_s x,\ x \in X,\ y_s \in Y_s \right\}.$$

Problem (40) has a decomposable structure that can be exploited. To induce this structure, let us introduce scenario-dependent copy variables x_s of the first-stage variable x for each $s \in \mathcal{S}$. Using these copy variables, we can reformulate (40) as

$$\zeta = \underset{x,y,z}{\text{minimize}} \left\{ \sum_{s \in \mathcal{S}} p_s (c^\top x_s + q_s^\top y_s) : (x_s, y_s) \in K_s,\ x_s = z\ \forall s \in \mathcal{S},\ z \in \mathbf{R}^{n_x} \right\}, \tag{41}$$

The variable $z \in \mathbf{R}^{n_x}$ is a common global variable, and the constraints $x_s = z$ for all $s \in \mathcal{S}$ enforce *nonanticipativity* for the first-stage decisions: all first-stage decisions x_s must be the same (z) for each scenario $s \in \mathcal{S}$.

Relaxing the nonanticipativity constraints $x_s = z$ for all $s \in \mathcal{S}$ in (41) in Lagrangian fashion yields the following *augmented* Lagrangian dual function

$$\phi(\mu) = \underset{x,y,z}{\text{minimize}} \quad \sum_{s \in \mathcal{S}} \left[p_s (c^\top x_s + q_s^\top y_s) + \mu_s^\top (x_s - z) + \frac{\rho}{2} ||x_s - z||_2^2 \right] \tag{42}$$

$$\text{subject to} \quad (x_s, y_s) \in K_s, \quad \forall s \in \mathcal{S} \tag{43}$$

By defining $v_s = \mu_s / p_s$ and letting the penalty parameter ρ be multiplied by the scenario probabilities p_s for all $s \in \mathcal{S}$, we can rewrite (42) – (43) as

$$\phi(v) = \underset{x,y,z}{\text{minimize}} \quad \sum_{s \in \mathcal{S}} p_s L_s^\rho (x_s, y_s, z, v_s) \tag{44}$$

$$\text{subject to} \quad (x_s, y_s) \in K_s, \quad \forall s \in \mathcal{S} \tag{45}$$

where

$$L_s^\rho (x_s, y_s, z, v_s) = c^\top x_s + q_s^\top y_s + v_s^\top (x_s - z) + \frac{\rho}{2} ||x_s - z||_2^2 \tag{46}$$

Since z is unconstrained in (46), the value of $\phi(v)$ in (44) can be made arbitrarily small unless

$$v_s^\top z = 0, \quad \text{for all } s \in \mathcal{S} \tag{47}$$

Therefore, this term vanishes from (46) for all $s \in S$. Derive the ADMM iterations for solving the problem (44) – (45) in a distributed fashion for each scenario $s \in \mathcal{S}$ separately.

Solution.

Since the term $v_s^\top z$ vanishes from $L_s^\rho(x_s, y_s, z, v_s)$ according to (47), we can rewrite (46) as

$$L_s^\rho(x_s, y_s, z, v_s) = (c + v_s)^\top x_s + q_s^\top y_s + \frac{\rho}{2}||x_s - z||_2^2 \tag{48}$$

In this case, the ADMM update step of (x_s, y_s) for all $s \in \mathcal{S}$ is of the form

$$\begin{aligned}(x_s^{k+1}, y_s^{k+1}) &= \underset{(x_s,y_s)\in K_s}{\operatorname{argmin}} \; L_s^\rho(x_s, y_s, z^k, v_s^k) \\ &= \underset{(x_s,y_s)\in K_s}{\operatorname{argmin}} \left\{(c + v_s^k)^\top x_s + q_s^\top y_s + \frac{\rho}{2}||x_s - z^k||_2^2\right\}\end{aligned} \tag{49}$$

which can be done in parallel for each scenario $s \in S$. Thus, computing (x_s^{k+1}, y_s^{k+1}) for each $s \in S$ amounts to solving a quadratic problem with linear constraints defined in $(x_s, y_s) \in K_s$. After updating x_s^{k+1} and y_s^{k+1} for each scenario $s \in S$, the $z-$update is of the form

$$\begin{aligned}z^{k+1} &= \underset{z}{\operatorname{argmin}} \sum_{s\in S} p_s L_s^\rho(x_s^{k+1}, y_s^{k+1}, z, v_s^k) \\ &= \underset{z}{\operatorname{argmin}} \sum_{s\in S} p_s \left[(c + v_s^k)^\top x_s^{k+1} + q_s^\top y_s^{k+1} + \frac{\rho}{2}||x_s^{k+1} - z||_2^2\right]\end{aligned} \tag{50}$$

Taking the gradient of (50) with regard to z and setting it to zero, we get

$$\begin{aligned}\sum_{s\in S} p_s \rho(x_s^{k+1} - z) = 0 \\ \sum_{s\in S} p_s x_s^{k+1} - z \sum_{s\in S} p_s = 0\end{aligned} \tag{51}$$

since $\sum_{s\in S} p_s = 1$, we get the following $z-$update from (51):

$$z^{k+1} = \sum_{s\in S} p_s x_s^{k+1} \tag{52}$$

Finally, the update for the dual variables v is computed as

$$v_s^{k+1} = v_s^k + \rho(x_s^{k+1} - z^{k+1}) \tag{53}$$

which can be done in parallel for each scenario $s \in S$. To recap, we first update (x_s^{k+1}, y_s^{k+1}) for each scenario $s \in S$ separately (we can do this in parallel) by using (49) which corresponds to solving a quadratic problem with linear constraints. Then, we update z^{k+1} by simply using (52). Finally, we update v_s^{k+1} for each scenario $s \in S$ in a regular way by using the formula (53).

The squared primal residual norm in this case is $||r_s^{k+1}||_2^2 = p_s||x^{k+1} - z^{k+1}||_2^2$ for all $s \in S$ and the squared dual residual norm becomes $p_s||s^{k+1}||_2^2 = ||z^{k+1} - z^k||_2^2$. Summing these two yields

$$\sum_{s\in S} p_s \left[||x_s^{k+1} - z^{k+1}||_2^2 + ||z^{k+1} - z^k||_2^2\right] = \sum_{x\in S} p_s||x_s^{k+1} - z^k||_2^2 \tag{54}$$

which holds since for every $s \in S$ the cross term resulting from expanding the squared norm

$$||(x_s^{k+1} - z^{k+1}) + (z^{k+1} - z^k)||_2^2$$

vanishes. This is seen in the equality $\sum_{s\in S} p_s(x_s^{k+1} - z^{k+1}) = 0$ due to how z^k is constructed.

However, in Project Work 2, we will use as stopping criterion the following modified sum of primal and dual residuals term which makes convergence smoother:

$$\sum_{s \in S} p_s \rho ||x_s^{k+1} - z^k||_2 \tag{55}$$

The stopping criterion (55) can be computed in parallel for each $s \in S$ after updating x^{k+1}. You can see this part in the skeleton code in Project Work 2. Basically, the algorithm stops when

$$\sum_{s \in S} p_s \rho ||x_s^{k+1} - z^k||_2 < \epsilon$$

for some tolerance $\epsilon > 0$.