Problem 9.1: ADMM and Scaled Form ADMM

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

$$\min_{x,z}. \; f(x) + g(z) \tag{1}$$

$$\text{subject to: } Ax + Bz = c \tag{2}$$

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

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

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

$$\min_{x,z}. \; f(x) + g(z) + \frac{\rho}{2}||Ax + Bz - c||_2^2 \tag{4}$$

$$\text{subject to: } Ax + Bz = c \tag{5}$$

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

$$x^{k+1} = \operatorname*{argmin}_x L_\rho(x, z^k, y^k) \tag{6}$$

$$z^{k+1} = \operatorname*{argmin}_z L_\rho(x^{k+1}, z, y^k) \tag{7}$$

$$y^{k+1} = y^k + \rho(Ax^{k+1} + Bz^{k+1} - c) \tag{8}$$

(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

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

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.

Solution.

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

$$Ax^* + Bz^* - c = 0 \qquad \text{(primal feasibility)} \tag{10}$$

$$\nabla_x f(x^*) + A^\top y^* = 0 \qquad \text{(dual feasibility 1)} \tag{11}$$

$$\nabla_z g(z^*) + B^\top y^* = 0 \qquad \text{(dual feasibility 2)} \tag{12}$$

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

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

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 get

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

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, for instance, 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

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

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

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

where we can interpret $K = -(\rho/2)||u||_2^2$ as a constant, because it does not affect the x–step or the z–step (indeed, it remains constant in both). Moreover, since the values of y (and thus u) are updated using a Gradient Descent formula, the term K does not affect the ADMM iterations. Thus, we can ignore the constant K completely and rewrite (14) as

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

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 as a step size in the Gradient Descent update (20) as in (8) since we defined $y = \rho u$. The main reason for using the scaled form ADMM is that formulas related to the scaled ADMM variant are typically shorter and easier to interpret compared to the unscaled variant.

Problem 9.2: ADMM for Quadratic Optimization Problems

Consider the following standard form quadratic optimization problem

$$\min_{x}. \ \frac{1}{2}x^\top Px + q^\top x \tag{21}$$

$$\text{subject to: } Ax = b \tag{22}$$

$$x \geq 0 \tag{23}$$

with variables $x \in \mathbb{R}^n$. Assume that $P \in \mathbb{R}^{n\times n}$ is a symmetric positive definite matrix, $q \in \mathbb{R}^n$, $A \in \mathbb{R}^{p\times n}$, and $b \in \mathbb{R}^p$. We can express the problem (21) – (23) in ADMM form as

$$\min_{x,z}. \ f(x) + g(z) \tag{24}$$

$$\text{subject to: } x = z \tag{25}$$

where

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

is the original objective with a restricted domain, and $g : \mathbb{R}^n \to \{0, \infty\}$ is the indicator function of the nonnegative orthant $\mathbb{R}^n_+$ corresponding to the constraint $x \geq 0$. Write the augmented Lagrangian for (24) – (25) using the scaled dual variables, and write the corresponding scaled form ADMM iterations using the results of Exercise 9.1.

Solution.

According to (17) of Exercise 9.1, the augmented Lagrangian for (24) – (25) using the scaled dual variables is of the form

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

where $r = x - z$ is the *primal residual* and $u = y/\rho$ are the scaled dual variables. Now we can write the scaled form ADMM, which consists of the following iterations:

$$x^{k+1} = \operatorname*{argmin}_{x:Ax=b} \left\{ f(x) + \frac{\rho}{2}||x - z^k + u^k||_2^2 \right\} \tag{26}$$

$$z^{k+1} = \operatorname*{argmin}_{z} \left\{ g(z) + \frac{\rho}{2}||x^{k+1} - z + u^k||_2^2 \right\} \tag{27}$$

$$u^{k+1} = u^k + x^{k+1} - z^{k+1} \tag{28}$$

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 (27). Taking the gradient of (27) 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

$$z^{k+1} = (x^{k+1} + u^k)_+ \tag{29}$$

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

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

for all $i = 1, \dots, n$. Using the notation (29), we can simplify (26) – (28) as

$$x^{k+1} = \underset{x:Ax=b}{\operatorname{argmin}} \left\{ f(x) + \frac{\rho}{2}||x - z^k + u^k||_2^2 \right\} \tag{30}$$

$$z^{k+1} = (x^{k+1} + u^k)_+ \tag{31}$$

$$u^{k+1} = u^k + x^{k+1} - z^{k+1} \tag{32}$$

Notice that the $x-$update in (30) corresponds to computing the following equality constrained least squares problem

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

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 $v \in \mathbb{R}^p$ is the dual variable vector corresponding to the constraints $Ax = b$.

The optimality conditions for (24) – (25) with the unscaled dual variables $y \in \mathbb{R}^n$ are

$$x^* - z^* = 0 \qquad \text{(primal feasibility)} \tag{33}$$

$$\nabla_x f(x^*) + y^* = 0 \qquad \text{(dual feasibility 1)} \tag{34}$$

$$\nabla_z g(z^*) - y^* = 0 \qquad \text{(dual feasibility 2)} \tag{35}$$

Since z^{k+1} minimizes $L_\rho(x^{k+1}, z, u^k)$, by taking the gradient of (27) 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 (31) always satisfies the dual feasibility condition (35). Similarly, since x^{k+1} minimizes $L_\rho(x, z^k, u^k)$, by taking the gradient of (26) with respect to x, we get

$$\begin{aligned} 0 &= \nabla_x f(x^{k+1}) + \rho(x^{k+1} - z^k + u^k) \\ &= \nabla_x f(x^{k+1}) + \rho(x^{k+1} - z^{k+1} + u^k - z^k + z^{k+1}) \\ &= \nabla_x f(x^{k+1}) + \rho u^{k+1} + \rho(z^{k+1} - z^k) \\ &= \nabla_x f(x^{k+1}) + y^{k+1} + \rho(z^{k+1} - z^k) \end{aligned} \tag{36}$$

Comparing (36) to (34), 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 like before as

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

Thus, we can again use, for example, the following stopping criterion:

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

It is worth mentioning that ADMM is best suited for large-scale problems with decomposable structures whose subproblems can be solved in parallel using several CPUs, GPUs, or better yet a combination of both. However, ADMM is not typically highly accurate, so if accuracy is of high priority, some other algorithm may be more appropriate. Also, problems that cannot be decomposed and thus where parallel computing cannot be exploited, the normal Method of Multipliers (MM) is a better choice.

Nevertheless, ADMM has become the method of choice in a growing number of applications where standard sequential algorithms fail to provide a sufficiently good solutions in reasonable times. Indeed, the accuracy of ADMM is often more than enough in practice, and decomposable structures can be found in most surprising contexts, one example being Exercise 9.2 where use of the indicator function provided a decomposable structure to the problem. Moreover, in some applications it is possible that the main steps can be further decomposed into even smaller subproblems, thus increasing the appeal of ADMM.

This technical paper provides an excellent review of ADMM and gives several potential application areas where ADMM can be applied in.