SUPPLEMENTAL MATERIAL: GLOBAL CONVERGENCE OF ALTERNATING DIRECTION METHOD OF MULTIPLIERS FOR INVEX OBJECTIVE LOSSES

Samuel Pinilla, Siu-Lun Yeung, and Jeyan Thiyagalingam

Scientific Computing Department, Science and Technology Facilities Council, Harwell, UK
{samuel.pinilla,siu-lun.yeung,t.jeyan}@stfc.ac.uk

1. PROOF OF THEOREM 1

In order to establish a versatile framework for invex functions, we provide proof of invexity for a specific class of summation functions. The key result is encapsulated in the following lemma, which reveals a fundamental connection: the invexity of the summation function, obtained by applying a unidimensional real function to distinct entries of a vector, hinges upon the invexity of this underlying function.

Lemma 1 (Invexity of Function Sum). Let $g : \mathbb{R}^n \to \mathbb{R}$ be a function defined as

$$g(\boldsymbol{x}) = \sum_{i=1}^{n} r(\boldsymbol{x}[i]), \tag{1}$$

where $r : \mathbb{R} \to \mathbb{R}$. If $r(w)$ is an invex function then $g(\boldsymbol{x})$ is also invex.

Proof. Let $g : \mathbb{R}^n \to \mathbb{R}$ be a function defined as $g(\boldsymbol{x}) = \sum_{i=1}^{n} r(\boldsymbol{x}[i])$. Assume $r : \mathbb{R} \to \mathbb{R}$ is invex. Then, from the invexity of $r(w)$ we have that there exists $\eta_r : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ such that

$$r(w_1) - r(w_2) \geq \zeta_{w_2} \cdot \eta_r(w_1, w_2), \tag{2}$$

for all $w_1, w_2 \in \mathbb{R}$, and any $\zeta_{w_2} \in \partial r(w_2)$. Take $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$. Then, we have

$$r(\boldsymbol{x}[i]) - r(\boldsymbol{y}[i]) \geq \zeta_{\boldsymbol{y}[i]} \cdot \eta_r(\boldsymbol{x}[i], \boldsymbol{y}[i]), \tag{3}$$

for any $i = 1, \ldots, n$, and $\forall \zeta_{\boldsymbol{y}[i]} \in \partial r(\boldsymbol{y}[i])$. From the above inequality we conclude that for any $\boldsymbol{\zeta} \in \partial g(\boldsymbol{y})$

$$\begin{aligned}
\sum_{i=1}^{n} r(\boldsymbol{x}[i]) - r(\boldsymbol{y}[i]) &\geq \sum_{i=1}^{n} \zeta_{\boldsymbol{y}[i]} \cdot \eta_r(\boldsymbol{x}[i], \boldsymbol{y}[i]) \\
g(\boldsymbol{x}) - g(\boldsymbol{y}) &\geq \sum_{i=1}^{n} \zeta_{\boldsymbol{y}[i]} \cdot \eta_r(\boldsymbol{x}[i], \boldsymbol{y}[i]) \\
g(\boldsymbol{x}) - g(\boldsymbol{y}) &\geq \boldsymbol{\zeta}^T \eta(\boldsymbol{x}, \boldsymbol{y}),
\end{aligned} \tag{4}$$

such that $\eta(\boldsymbol{x}, \boldsymbol{y}) = [\eta_r(\boldsymbol{x}[1], \boldsymbol{y}[1]), \ldots, \eta_r(\boldsymbol{x}[n], \boldsymbol{y}[n])]^T$, and $\boldsymbol{\zeta} = [\zeta_{\boldsymbol{y}[1]}, \ldots, \zeta_{\boldsymbol{y}[n]}]^T$. Thus, from Eq. (4) the results holds. □

Now we prove Theorem 1 using the result in Lemma 1

Proof. We proceed by cases.

1. Let f be an admissible function. Then according to Definition 4 we know that $f(\boldsymbol{x}) = \sum_{i=1}^{n} s(|\boldsymbol{x}[i]|)$, for some $s : [0, \infty) \to [0, \infty)$ with $s'(w) > 0$ where $w \in (0, \infty)$. Since the structure of f fits assumption in Lemma 1 then if $s(w)$ is invex then $f(\boldsymbol{x})$ is invex. Take $w_1, w_2 \in (0, \infty)$, and define $\eta : (0, \infty)^2 \to \mathbb{R}$ as

$$\eta(w_1, w_2) = \begin{cases} 0 & \text{if } s(w_1) > s(w_2) \\ \frac{s(w_1) - s(w_2)}{(\zeta_*)^2} \zeta_* & \text{otherwise} \end{cases}, \tag{5}$$

   where ζ_* is an element in $\partial s(w_2)$ of minimal absolute value which satisfies $\frac{\zeta_* \cdot \zeta}{(\zeta_*)^2} \geq 1$ for all $\zeta \in \partial s(w_2)$. The existence of ζ_* is guaranteed because $s'(w) > 0$ and therefore $0 \notin \partial s(w_2)$ [1, Page 64]. From the above equation it is clear that for all w_1, w_2 we have

$$s(w_1) - s(w_2) \geq \eta(w_1, w_2) \cdot \zeta_{w_2}, \quad \forall \zeta_{w_2} \in \partial s(w_2) \tag{6}$$

   which means s is invex. Therefore $f(\boldsymbol{x})$ is invex.

2. Let f, g be an admissible functions. Considering the result in previous statement, it is enough to show that $h = \beta f + \alpha g$ is an admissible function for any $\beta, \alpha \geq 0$. By definition we have that $h(0) = 0$, and $h'(w) > 0$. In addition, since both f, g are positive functions it implies that $h(w)/w^2$ is non-increasing. Thus the result holds.

3. Let $f, g : \mathbb{R}^n \to \mathbb{R}$ be two admissible functions as in Definition 4, such that $f(\boldsymbol{x}) = \sum_{i=1}^n s_f(|\boldsymbol{x}[i]|)$, and $g(\boldsymbol{x}) = \sum_{i=1}^n s_g(|\boldsymbol{x}[i]|)$. Define $h_c(\boldsymbol{x}) = \sum_{i=1}^n (s_f \circ s_g)(|\boldsymbol{x}[i]|)$. Then, observe that $(s_f \circ s_g)(0) = s_f(s_g(0)) = s_f(0) = 0$. Additionally, since $s_f'(w), s_g'(w) > 0$, then $(s_f \circ s_g)'(w) > 0$ (by the chain rule) for all $w \in (0, \infty)$. Finally, we know that $s_f'(w), s_g'(w) > 0$ implies $s_f(w), s_g(w) > 0$ to be strictly increasing. Therefore, for any $w_1 < w_2$ we know $(s_f \circ s_g)(w_1) < (s_f \circ s_g)(w_2)$, which implies $\frac{(s_f \circ s_g)(w_1)}{(w_1)^2} > \frac{(s_f \circ s_g)(w_2)}{(w_2)^2}$. Thus the result holds.

4. This holds trivially from definition of admissible function.

5. This holds trivially from definition of admissible function.

Thus we proved the first part of this theorem. □

2. PROOF OF THEOREM 2

In this proof we seek to guarantee that the list of functions in Theorem 2 are admissible functions, and we proceed by cases.

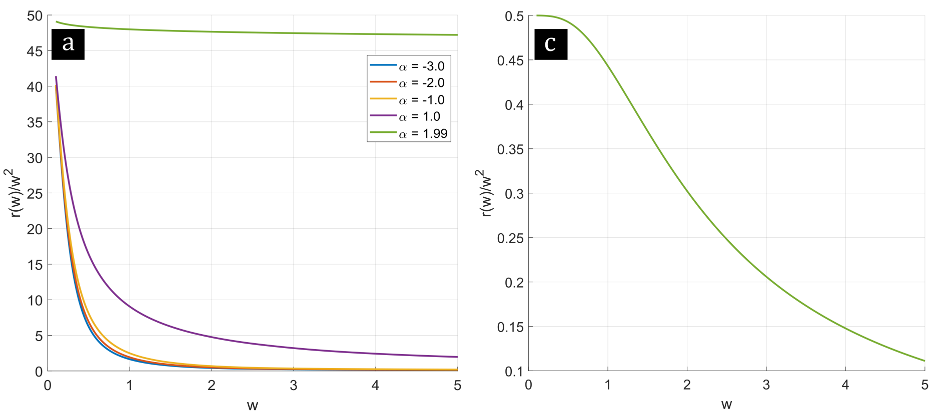

Fig. 1: Plot of $r(w)/w$ for $r(w)$ being (a) Eq. (7) for $c = 0.1$, and (c) Eq. (8) and $w > 0$ to check that $r(w)/w^2$ is non-increasing on $(0, \infty)$

Eq. (5)

Proof. Take $r(w) = \log(1 + \frac{w^2}{\delta^2})$ for any $w \neq 0$, and fixed $\delta \in \mathbb{R}$. It is trivial to see that $r(0) = 0$, that $r(w)$ it is not identically zero, and non-decreasing on $(0, \infty)$. Then, we just need to show that $r(w)/w^2$ is non-increasing on $(0, \infty)$. Observe that the first derivative of $h(w) = r(w)/w^2$ is given by $h'(w) = \frac{2\left(\frac{w^2}{\delta^2+w^2} - \log(1+\frac{w^2}{\delta^2})\right)}{w^3}$. Since $\frac{w^2}{\delta^2+w^2} - \log(1 + \frac{w^2}{\delta^2}) < 0$, then we have that $h'(w) < 0$, which leads to conclude that $r(w)/w^2$ is non-increasing on $(0, \infty)$. Then it is clear that $r(w)/w^2$ is non-increasing on $(0, \infty)$. □

Eq. (6)

Proof. Take $r(w) = \frac{2w^2}{w^2+4\delta^2}$ for any $w \neq 0$, and fixed $\delta \in \mathbb{R}$. It is trivial to see that $r(0) = 0$, that $r(w)$ it is not identically zero, and non-decreasing on $(0, \infty)$. Then, we just need to show that $r(w)/w^2$ is non-increasing on $(0, \infty)$. Observe that $h(w) = r(w)/w^2$ is given by $h(w) = \frac{2}{w^2+4\delta^2}$, which leads to conclude that $r(w)/w^2$ is non-increasing on $(0, \infty)$. Then it is clear that $r(w)/w^2$ is non-increasing on $(0, \infty)$. □

Eq. (7)

Take $r(w) = \frac{|\alpha-2|}{\alpha}\left(\left(\frac{(w/c)^2}{|\alpha-2|} + 1\right)^{\alpha/2} - 1\right)$ for any $w \neq 0$, and fixed $\alpha \in \mathbb{R}$, $c > 0$. It is trivial to see that $r(0) = 0$, that $r(w)$ it is not identically zero, and non-decreasing on $(0, \infty)$. Then, we just need to show that $r(w)/w^2$ is non-increasing on $(0, \infty)$. For easy of exposition we present in Figure 1(b) the plot of $r(w)/w^2$. Then it is clear that $r(w)/w^2$ is non-increasing on $(0, \infty)$.

Eq. (8)

Take $r(w) = \log\left(1 + w^2\right) - \frac{w^2}{2w^2+2}$ for any $w \neq 0$. It is trivial to see that $r(0) = 0$, that $r(w)$ it is not identically zero, and non-decreasing on $(0, \infty)$. Then, we just need to show that $r(w)/w^2$ is non-increasing on $(0, \infty)$. For easy of exposition we present in Figure 1(c) the plot of $r(w)/w^2$. Then it is clear that $r(w)/w^2$ is non-increasing on $(0, \infty)$.

3. PROOF OF THEOREM 3

Before proving this theorem, we present an auxiliary theoretical result needed to ensure the update steps of $\boldsymbol{x}^{(t+1)}$ and $\boldsymbol{z}^{(t+1)}$ are unique when assuming f, and g are admissible functions.

Lemma 2. Let $f : \mathbb{R}^n \to \mathbb{R}$ be an admissible function. Take $\boldsymbol{x} \in \mathbb{R}^n$ such that $\|\boldsymbol{x}\|_\infty < \infty$. If $f(\boldsymbol{x}) < \infty$, then $\|\boldsymbol{x}\|_2^2 < \frac{1}{c(\boldsymbol{x})} f(\boldsymbol{x})$ for a constant $c(\boldsymbol{x}) > 0$.

Proof. Since f is an admissible function, then $f(\boldsymbol{x}) = \sum_{i=1}^{n} s(\boldsymbol{x}[i])$ and $s(w)/w^2$ is nonincreasing. Therefore, we have for all $i = 1, \ldots, n$

$$\frac{s(\boldsymbol{x}[i])}{\boldsymbol{x}^2[i]} \geq \frac{s(\|\boldsymbol{x}\|_\infty)}{\|\boldsymbol{x}\|_\infty^2} = c(\boldsymbol{x}) > 0. \tag{7}$$

Thus, we obtain

$$\|\boldsymbol{x}\|_2^2 \leq \frac{1}{c(\boldsymbol{x})} f(\boldsymbol{x}). \tag{8}$$

Thus the result holds. □

Now, we proceed to prove Theorem 3 following a similar strategy as presented in [2].

Proof. Since $(\boldsymbol{x}^*, \boldsymbol{z}^*, \boldsymbol{v}^*)$ is a saddle point for $\mathcal{L}_0$, we have

$$\mathcal{L}_0(\boldsymbol{x}^*, \boldsymbol{z}^*, \boldsymbol{v}^*) \leq \mathcal{L}_0(\boldsymbol{x}^{(t+1)}, \boldsymbol{z}^{(t+1)}, \boldsymbol{v}^*). \tag{9}$$

Using $\boldsymbol{A}\boldsymbol{x}^* + \boldsymbol{B}\boldsymbol{z}^* = \boldsymbol{y}$ the left hand side is $h^* = \inf\{f(\boldsymbol{x}) + g(\boldsymbol{z}) \mid \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{z} = \boldsymbol{y}\}$. With $h^{(t+1)} = f(\boldsymbol{x}^{(t+1)}) + g(\boldsymbol{z}^{(t+1)})$, this can be written as

$$h^* \leq h^{(t+1)} + (\boldsymbol{v}^*)^T \boldsymbol{q}^{(t+1)}, \tag{10}$$

for $\boldsymbol{q}^{(t+1)} = \boldsymbol{A}\boldsymbol{x}^{(t+1)} + \boldsymbol{B}\boldsymbol{z}^{(t+1)} - \boldsymbol{y}$. Now, by definition, $\boldsymbol{x}^{(t+1)}$ minimizes $\mathcal{L}_\rho(\boldsymbol{x}, \boldsymbol{z}^{(t)}, \boldsymbol{v}^{(t)})$. From Lemma 2 and the fact that $\rho\sigma_n(\boldsymbol{A}) \geq 1$, $\rho\sigma_p(\boldsymbol{B}) \geq 1$ we can appeal to [3, Proposition 5.2.13] that ensures the necessary and sufficient optimality condition for $\mathcal{L}_\rho(\boldsymbol{x}, \boldsymbol{z}^{(t)}, \boldsymbol{v}^{(t)})$ is given by

$$\boldsymbol{0} \in \partial\mathcal{L}_\rho(\boldsymbol{x}^{(t+1)}, \boldsymbol{z}^{(t)}, \boldsymbol{z}^{(t)}) = \partial f(\boldsymbol{x}^{(t+1)}) + \boldsymbol{A}^T \boldsymbol{v}^{(t)} + \rho \boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{x}^{(t+1)} + \boldsymbol{B}\boldsymbol{z}^{(t)} - \boldsymbol{y}). \tag{11}$$

Since $\boldsymbol{v}^{(t+1)} = \boldsymbol{v}^{(t)} + \rho\boldsymbol{q}^{(t+1)}$, we can plug in $\boldsymbol{v}^{(t)} = \boldsymbol{v}^{(t+1)} - \rho\boldsymbol{q}^{(t+1)}$ and rearrange to obtain

$$\boldsymbol{0} \in \partial f(\boldsymbol{x}^{(t+1)}) + \boldsymbol{A}^T(\boldsymbol{v}^{(t+1)} - \rho\boldsymbol{B}(\boldsymbol{z}^{(t+1)} - \boldsymbol{z}^{(t)})). \tag{12}$$

This implies that $\boldsymbol{x}^{(t+1)}$ minimizes

$$f(\boldsymbol{x}) + (\boldsymbol{v}^{(t+1)} - \rho\boldsymbol{B}(\boldsymbol{z}^{(t+1)} - \boldsymbol{z}^{(t)}))^T \boldsymbol{A}\boldsymbol{x}. \tag{13}$$

A similar argument shows that $\boldsymbol{z}^{(t+1)}$ minimizes $g(\boldsymbol{z}) + (\boldsymbol{v}^{(t+1)})^T \boldsymbol{B}\boldsymbol{z}$. It follows that

$$\begin{aligned} &f(\boldsymbol{x}^{(t+1)}) + (\boldsymbol{v}^{(t+1)} - \rho\boldsymbol{B}(\boldsymbol{z}^{(t+1)} - \boldsymbol{z}^{(t)}))^T \boldsymbol{A}\boldsymbol{x}^{(t+1)} \\ &\qquad \leq f(\boldsymbol{x}^*) + (\boldsymbol{v}^{(t+1)} - \rho\boldsymbol{B}(\boldsymbol{z}^{(t+1)} - \boldsymbol{z}^{(t)}))^T \boldsymbol{A}\boldsymbol{x}^*, \end{aligned} \tag{14}$$

and that

$$g(\boldsymbol{z}^{(t+1)}) + (\boldsymbol{v}^{(t+1)})^T \boldsymbol{B}\boldsymbol{z}^{(t+1)} \leq g(\boldsymbol{z}^*) + (\boldsymbol{v}^{(t+1)})^T \boldsymbol{B}\boldsymbol{z}^*. \tag{15}$$

Adding the two inequalities above, using $\boldsymbol{A}\boldsymbol{x}^* + \boldsymbol{B}\boldsymbol{z}^* = \boldsymbol{y}$, and rearranging, we obtain

$$h^{(t+1)} - h^* \leq -(\boldsymbol{v}^{(t+1)})^T \boldsymbol{q}^{(t+1)} - \rho(\boldsymbol{B}(\boldsymbol{z}^{(t+1)} - \boldsymbol{z}^{(t)}))^T(-\boldsymbol{q}^{(t+1)} + \boldsymbol{B}(\boldsymbol{z}^{(t+1)} - \boldsymbol{z}^*)). \tag{16}$$

On the other hand, adding Eqs. (10), and (16), regrouping terms, and multiplying through by 2 gives

$$\begin{aligned}2(\boldsymbol{v}^{(t+1)}-\boldsymbol{v}^*)^T\boldsymbol{q}^{(t+1)} &- 2\rho(\boldsymbol{B}(\boldsymbol{z}^{(t+1)}-\boldsymbol{z}^{(t)}))^T\boldsymbol{q}^{(t+1)} \\ &+ 2\rho(\boldsymbol{B}(\boldsymbol{z}^{(t+1)}-\boldsymbol{z}^{(t)}))^T(\boldsymbol{B}(\boldsymbol{z}^{(t+1)}-\boldsymbol{z}^*)) \le 0.\end{aligned} \tag{17}$$

Now by rewriting the first term in Eq. (17), and substituting $\boldsymbol{v}^{(t+1)} = \boldsymbol{v}^{(t)} + \rho\boldsymbol{q}^{(t+1)}$ it gives

$$2(\boldsymbol{v}^{(t+1)}-\boldsymbol{v}^*)^T\boldsymbol{q}^{(t+1)} + \rho\|\boldsymbol{q}^{(t+1)}\|_2^2 + \rho\|\boldsymbol{q}^{(t+1)}\|_2^2, \tag{18}$$

and substituting $\boldsymbol{q}^{(t+1)} = (1/\rho)(\boldsymbol{v}^{(t+1)} - \boldsymbol{v}^{(t)})$ in the first two terms gives

$$(2/\rho)(\boldsymbol{v}^{(t)}-\boldsymbol{v}^*)^T(\boldsymbol{v}^{(t+1)}-\boldsymbol{v}^{(t)}) + (1/\rho)\|\boldsymbol{v}^{(t+1)}-\boldsymbol{v}^{(t)}\|_2^2 + \rho\|\boldsymbol{q}^{(t+1)}\|_2^2. \tag{19}$$

Since $\boldsymbol{q}^{(t+1)} - \boldsymbol{q}^{(t)} = (\boldsymbol{q}^{(t+1)} - \boldsymbol{q}^*) - (\boldsymbol{q}^{(t)} - \boldsymbol{q}^*)$, this can be written as

$$(1/\rho)(\|\boldsymbol{v}^{(t+1)}-\boldsymbol{v}^*\|_2^2 - \|\boldsymbol{v}^{(t)}-\boldsymbol{v}^*\|_2^2) + \rho\|\boldsymbol{q}^{(t+1)}\|_2^2. \tag{20}$$

We now rewrite the remaining terms

$$\rho\|\boldsymbol{q}^{(t+1)}\|_2^2 - 2\rho(\boldsymbol{B}(\boldsymbol{z}^{(t+1)}-\boldsymbol{z}^{(t)}))^T\boldsymbol{q}^{(t+1)} + 2\rho(\boldsymbol{B}(\boldsymbol{z}^{(t+1)}-\boldsymbol{z}^{(t)}))^T(\boldsymbol{B}(\boldsymbol{z}^{(t+1)}-\boldsymbol{z}^*)), \tag{21}$$

where $\rho\|\boldsymbol{q}^{(t+1)}\|_2^2$ is taken from Eq. (20). Substituting

$$\boldsymbol{z}^{(t+1)} - \boldsymbol{z}^* = (\boldsymbol{z}^{(t+1)} - \boldsymbol{z}^{(t)}) + (\boldsymbol{z}^{(t)} - \boldsymbol{z}^*), \tag{22}$$

in the last term gives

$$\begin{aligned}\rho\|\boldsymbol{q}^{(t+1)} - \boldsymbol{B}(\boldsymbol{z}^{(t+1)}-\boldsymbol{z}^{(t)})\|_2^2 &+ \rho\|\boldsymbol{B}(\boldsymbol{z}^{(t+1)}-\boldsymbol{z}^{(t)})\|_2^2 \\ &+ 2\rho(\boldsymbol{B}(\boldsymbol{z}^{(t+1)}-\boldsymbol{z}^{(t)}))^T(\boldsymbol{B}(\boldsymbol{z}^{(t+1)}-\boldsymbol{z}^*)),\end{aligned} \tag{23}$$

and substituting

$$\boldsymbol{z}^{(t+1)} - \boldsymbol{z}^{(t)} = (\boldsymbol{z}^{(t+1)} - \boldsymbol{z}^*) - (\boldsymbol{z}^{(t)} - \boldsymbol{z}^*), \tag{24}$$

in the last two terms, we get

$$\rho\|\boldsymbol{q}^{(t+1)} - \boldsymbol{B}(\boldsymbol{z}^{(t+1)}-\boldsymbol{z}^{(t)})\|_2^2 + \rho\left(\|\boldsymbol{B}(\boldsymbol{z}^{(t+1)}-\boldsymbol{z}^*)\|_2^2 - \|\boldsymbol{B}(\boldsymbol{z}^{(t)}-\boldsymbol{z}^*)\|_2^2\right). \tag{25}$$

With the previous step, this implies that Eq. (17) can be written as

$$V^{(t)} - V^{(t+1)} \ge \rho\|\boldsymbol{q}^{(t+1)} - \boldsymbol{B}(\boldsymbol{z}^{(t+1)}-\boldsymbol{z}^{(t)})\|_2^2, \tag{26}$$

where $V^{(t)} = (1/\rho)\|\boldsymbol{v}^{(t)} - \boldsymbol{v}^*\|_2^2 + \rho\|\boldsymbol{B}(\boldsymbol{z}^{(t)} - \boldsymbol{z}^*)\|_2^2$.

Now, we show that the middle term $-2\rho(\boldsymbol{q}^{(t+1)})^T(\boldsymbol{B}(\boldsymbol{z}^{(t+1)} - \boldsymbol{z}^{(t)}))$ of the expanded right hand side of Eq. (26) is positive. To see this, recall that $\boldsymbol{z}^{(t+1)}$ minimizes $g(\boldsymbol{z}) + (\boldsymbol{v}^{(t+1)})^T\boldsymbol{B}\boldsymbol{z}$, and $\boldsymbol{z}^{(t)}$ minimizes $g(\boldsymbol{z}) + (\boldsymbol{v}^{(t)})^T\boldsymbol{B}\boldsymbol{z}$, so we can add

$$g(\boldsymbol{z}^{(t+1)}) + (\boldsymbol{v}^{(t+1)})^T\boldsymbol{B}\boldsymbol{z}^{(t+1)} \le g(\boldsymbol{z}^{(t)}) + (\boldsymbol{v}^{(t+1)})^T\boldsymbol{B}\boldsymbol{z}^{(t)}, \tag{27}$$

and

$$g(\boldsymbol{z}^{(t)}) + (\boldsymbol{v}^{(t)})^T\boldsymbol{B}\boldsymbol{z}^{(t)} \le g(\boldsymbol{z}^{(t+1)}) + (\boldsymbol{v}^{(t)})^T\boldsymbol{B}\boldsymbol{z}^{(t+1)}, \tag{28}$$

to get that

$$(\boldsymbol{v}^{(t+1)} - \boldsymbol{v}^{(t)})^T\boldsymbol{B}(\boldsymbol{z}^{(t+1)} - \boldsymbol{z}^{(t)}) \le 0. \tag{29}$$

Substituting $\boldsymbol{v}^{(t+1)} - \boldsymbol{v}^{(t)} = \rho\boldsymbol{q}^{(t+1)}$ gives the result, since $\rho > 0$. Thus, from Eqs. (26), and (29) we obtain

$$V^{(t+1)} \le V^{(t)} - \rho\|\boldsymbol{q}^{(t+1)}\|_2^2 - \rho\|\boldsymbol{B}(\boldsymbol{z}^{(t+1)} - \boldsymbol{z}^{(t)})\|_2^2, \tag{30}$$

which states that $V^{(t)}$ decreases in each iteration by an amount that depends on the norm of the residual $\boldsymbol{q}^{(t)}$ and on the change in $\boldsymbol{z}^{(t)}$ over one iteration. Then, because $V^{(t)} \le V^{(0)}$, it follows that $\boldsymbol{v}^{(t)}$ and $\boldsymbol{B}\boldsymbol{z}^{(t)}$ are bounded. Iterating the inequality above gives that

$$\rho\sum_{t=0}^{\infty}\left(\|\boldsymbol{q}^{(t+1)}\|_2^2 + \|\boldsymbol{B}(\boldsymbol{z}^{(t+1)} - \boldsymbol{z}^{(t)})\|_2^2\right) \le V^{(0)}, \tag{31}$$

which implies that $\boldsymbol{q}^{(t)} = \boldsymbol{A}\boldsymbol{x}^{(t)} + \boldsymbol{B}\boldsymbol{z}^{(t)} - \boldsymbol{y} \to 0$, and $\boldsymbol{B}(\boldsymbol{z}^{(t+1)} - \boldsymbol{z}^{(t)}) \to 0$ as $t \to \infty$. Additionally, applying [4, Lemma 1.2] on Eq. (31) we obtain a convergence rate for $\boldsymbol{q}^{(t)}$, $\boldsymbol{B}(\boldsymbol{z}^{(t+1)} - \boldsymbol{z}^{(t)})$ to zero of $\mathcal{O}(1/t)$. Eq. (31) also implies that the right hand side in Eq. (16) goes to zero as $t \to \infty$, because $\boldsymbol{B}(\boldsymbol{z}^{(t+1)} - \boldsymbol{z}^*)$ is bounded and both $\boldsymbol{q}^{(t+1)}$ and $\boldsymbol{B}(\boldsymbol{z}^{(t+1)} - \boldsymbol{z}^{(t)})$ go to zero. The right hand side in Eq. (10) goes to zero as $t \to \infty$, since $\boldsymbol{q}^{(t)}$ goes to zero. Thus we have $\lim_{t\to\infty} h^{(t)} = h^*$, i.e., objective convergence. Therefore the result of Theorem 3 holds. $\square$

3.1. Remarks on Stability of ADMM

In this section we wish to emphasize that the stability and reliability (of the ADMM) can be established from the following: 1) ADMM decomposes the overall optimization problem into a number of simpler subproblems that have a unique solution (as rigorously demonstrated in previous section), aiding convergence and stability. 2) As shown in previous section, the sequences $\boldsymbol{x}^{(t+1)}$, $\boldsymbol{z}^{(t+1)}$, and $\boldsymbol{v}^{(t+1)}$, constructed by ADMM algorithm, always converge to global optima irrespective of the initial states $\boldsymbol{x}^{(0)}$, $\boldsymbol{z}^{(0)}$, and $\boldsymbol{v}^{(0)}$, conferring a steadfast assurance of reliable attainment of optimal solutions, and finally, 3) The effectiveness of the ADMM is well demonstrated in the literature using a range of real-world applications [5–11], which we believe can reaffirm the reliability (and potentially the stability) of ADMM.

3.2. Proof of Lemma 1

In this section we prove a prox-regular function is quasi-invex. To that end, we introduce the following definition first.

Definition 1. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a lower semi-continuous function, and $\boldsymbol{u} \in \mathbb{R}^n$. Then f is said to be *prox-regular* if $f(\boldsymbol{x}) + \frac{1}{2\lambda}\|\boldsymbol{x} - \boldsymbol{u}\|_2^2$ is convex for some $\lambda > 0$.

Now we proceed with the proof.

Proof. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a prox-regular function for some $\lambda > 0$. Then we know that for any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$

$$f(\boldsymbol{x}) - f(\boldsymbol{y}) \geq \boldsymbol{\zeta}^T(\boldsymbol{x} - \boldsymbol{y}) - \frac{1}{2\lambda}\|\boldsymbol{x} - \boldsymbol{y}\|_2^2 \tag{32}$$

for all $\boldsymbol{\zeta} \in \partial f(\boldsymbol{y})$. Define function $\eta(\boldsymbol{x}, \boldsymbol{y})$ as

$$\eta(\boldsymbol{x}, \boldsymbol{y}) = \begin{cases} \mathbf{0} & \text{if } \mathbf{0} \in \partial f(\boldsymbol{y}) \\ \boldsymbol{x} - \boldsymbol{y} - \frac{\|\boldsymbol{x}-\boldsymbol{y}\|_2^2}{2\lambda\|\boldsymbol{\zeta}^*\|_2^2}\boldsymbol{\zeta}^* & \text{otherwise} \end{cases}, \tag{33}$$

where $\boldsymbol{\zeta}^*$ is an element in $\partial f(\boldsymbol{y})$ of minimum norm. Take $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ and assume

$$f(\boldsymbol{x}) - f(\boldsymbol{y}) \leq 0. \tag{34}$$

Observe that if $\mathbf{0} \in \partial f(\boldsymbol{y})$ then we get $\boldsymbol{\zeta}^T\eta(\boldsymbol{x}, \boldsymbol{y}) = 0$, for all $\boldsymbol{\zeta} \in \partial f(\boldsymbol{y})$. Additionally, if $\mathbf{0} \notin \partial f(\boldsymbol{y})$, then from Eq. (32) we obtain

$$\begin{aligned} 0 &\geq \boldsymbol{\zeta}^T(\boldsymbol{x} - \boldsymbol{y}) - \frac{1}{2\lambda}\|\boldsymbol{x} - \boldsymbol{y}\|_2^2 \\ &\geq \boldsymbol{\zeta}^T\left(\boldsymbol{x} - \boldsymbol{y} - \frac{\|\boldsymbol{x} - \boldsymbol{y}\|_2^2}{2\lambda\|\boldsymbol{\zeta}^*\|_2^2}\boldsymbol{\zeta}^*\right) = \boldsymbol{\zeta}^T\eta(\boldsymbol{x}, \boldsymbol{y}), \end{aligned} \tag{35}$$

where the second inequality comes from the fact that $\boldsymbol{\zeta}^*$ is an element in $\partial f(\boldsymbol{y})$ of minimum norm i.e. $\frac{\boldsymbol{\zeta}^T\boldsymbol{\zeta}^*}{\|\boldsymbol{\zeta}^*\|_2^2} \geq 1$ for all $\boldsymbol{\zeta} \in \partial f(\boldsymbol{y})$ [12, Theorem 2.4.4]. From the above inequality the result holds. □

4. REFERENCES

[1] Shashi K Mishra and Giorgio Giorgi, *Invexity and optimization*, vol. 88, Springer Science & Business Media, 2008.

[2] Neal Parikh and Stephen Boyd, "Proximal algorithms," *Foundations and Trends in optimization*, vol. 1, no. 3, pp. 127–239, 2014.

[3] Diethard Ernst Pallaschke and Stefan Rolewicz, *Foundations of mathematical optimization: convex analysis without linearity*, vol. 388, Springer Science & Business Media, 2013.

[4] Wei Deng, Ming-Jun Lai, Zhimin Peng, and Wotao Yin, "Parallel multi-block admm with $\mathcal{O}(1/k)$ convergence," *Journal of Scientific Computing*, vol. 71, pp. 712–736, 2017.

[5] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, Jonathan Eckstein, et al., "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine learning*, vol. 3, no. 1, pp. 1–122, 2011.

[6] Yu Wang, Wotao Yin, and Jinshan Zeng, "Global convergence of ADMM in nonconvex nonsmooth optimization," *Journal of Scientific Computing*, vol. 78, pp. 29–63, 2019.

[7] Roland Glowinski, "On alternating direction methods of multipliers: a historical perspective," *Modeling, simulation and optimization for science and technology*, pp. 59–82, 2014.

[8] Ji Li, "Solving blind ptychography effectively via linearized alternating direction method of multipliers," *Journal of Scientific Computing*, vol. 94, no. 1, pp. 19, 2023.

[9] Junli Liang, Petre Stoica, Yang Jing, and Jian Li, "Phase retrieval via the alternating direction method of multipliers," *IEEE Signal Processing Letters*, vol. 25, no. 1, pp. 5–9, 2017.

[10] Xingyu Xie, Jianlong Wu, Guangcan Liu, Zhisheng Zhong, and Zhouchen Lin, "Differentiable linearized ADMM," in *International Conference on Machine Learning*. PMLR, 2019, pp. 6902–6911.

[11] Jian Sun, Huibin Li, Zongben Xu, et al., "Deep ADMM-Net for compressive sensing MRI," *Advances in neural information processing systems*, vol. 29, 2016.

[12] Mokhtar S Bazaraa and Chitharanjan M Shetty, *Foundations of optimization*, vol. 122, Springer Science & Business Media, 2012.