Communication Efficient Primal-Dual Algorithm for Nonconvex Nonsmooth Distributed Optimization
1 PROOF OF THEOREM 1

Notation 1. Define
\[ g_i(x_i, z_i) = f_i(x_i) + h_i(x_i) + \frac{Np}{2} \|x_i - z_i\|^2, \]
\[ g(x, z) = \frac{1}{N} \sum_{i=1}^{N} g_i(x_i, z_i), \]
\[ d(y, z) = \min_{x \in \mathbb{R}^n} K(x, y, z) + h(x), \]
\[ x(y, z) = \arg \min_{x \in \mathbb{R}^n} K(x, y, z) + h(x), \]
\[ M(z) = \min_{x \in \mathbb{R}^n, Ax=0} \left( f(x) + h(x) + \frac{p}{2} \|x - z\|^2 \right), \]
\[ x^*(z) = \arg \min_{x \in \mathbb{R}^n, Ax=0} \left( f(x) + h(x) + \frac{p}{2} \|x - z\|^2 \right), \]
\[ e^t = x^t - \hat{x}^t. \]

First, we bound the compression error \( e^t \) in the following lemma.

Lemma 1. The following equality always holds:
\[ \sum_{t=1}^{T} \|e^t\|^2 \leq \frac{(1-\delta)^2}{\delta^2} \sum_{t=1}^{T} \|x^t - x^{t-1}\|^2. \]

Proof. Using the definition of \( e^t \), we can obtain:
\[ \|e^t\| = \|x^t - \hat{x}^t\| = \|x^t - \hat{x}^{t-1} - C(x^t - \hat{x}^{t-1})\| \]
\[ \leq (1-\delta) \|x^t - \hat{x}^{t-1}\| = (1-\delta) \|x^t - x^{t-1} + (x^{t-1} - \hat{x}^{t-1})\| \]
\[ \leq (1-\delta) \|x^t - x^{t-1}\| + (1-\delta) \|e^{t-1}\|, \]
where the first inequality is due to the assumption on the compression function.

By the induction, we can get
\[ \|e^t\| \leq \sum_{i=1}^{t} (1-\delta)^{t-i+1} \|x^i - x^{i-1}\|. \]

Then, using the convexity of square function and rearranging the summation terms, we can obtain
\[ \sum_{t=1}^{T} \|e^t\|^2 \leq \sum_{t=1}^{T} \left( \sum_{i=1}^{t} (1-\delta)^{t-i+1} \|x^i - x^{i-1}\|^2 \right)^2 \]
\[ \leq \sum_{t=1}^{T} \left( \sum_{i=1}^{t} (1-\delta)^{t-i+1} \right)^2 \sum_{i=1}^{t} (1-\delta)^{t-i+1} \|x^i - x^{i-1}\|^2 \]
\[ \leq \frac{1-\delta}{\delta} \sum_{t=1}^{T} \left( \sum_{i=1}^{t} (1-\delta)^{t-i+1} \right) \|x^i - x^{i-1}\|^2 \]
\[ \leq \frac{1-\delta}{\delta} \sum_{t=1}^{T} \left( \sum_{i=1}^{t} (1-\delta)^{t-i+1} \right) \|x^i - x^{i-1}\|^2 \]
\[ \leq \frac{(1-\delta)^2}{\delta^2} \sum_{i=1}^{T} \|x^i - x^{i-1}\|^2. \]

The proof is finished.

Then, we give the lower bound on the change of primal function when updating iterates.
Lemma 2. For any $t \geq 0$, if $c \leq \frac{1}{L'}$, the following inequality always holds:

$$h(x^t) - \langle \nabla_x K(x^t, y^t, z^t), x^{t+1} - x^t \rangle - h(x^{t+1}) - \frac{1}{2c} \|x^{t+1} - x^t\|^2 \geq \frac{1}{2c} \|x^{t+1} - x^t\|^2.$$

Proof. Recall the update iteration of $x^{t+1}$:

$$x^{t+1} = \arg\min_x \left(\langle \nabla_x K(x^t, y^t, z^t), x - x^t \rangle + h(x) + \frac{1}{2c} \|x - x^t\|^2\right).$$

By the optimality condition of strongly convex function, we obtain

$$h(x^t) - \langle \nabla_x K(x^t, y^t, z^t), x^{t+1} - x^t \rangle - h(x^{t+1}) - \frac{1}{2c} \|x^{t+1} - x^t\|^2 \geq \frac{1}{2c} \|x^{t+1} - x^t\|^2.$$

Besides, because $K(x, z, y)$ has Lipschitz gradient with respect to $x$, we can obtain

$$K(x^t, y^t, z^t) + h(x^t) - K(x^{t+1}, y^t, z^t) - h(x^{t+1}) \geq h(x^t) - h(x^{t+1}) - \langle \nabla_x K(x^t, y^t, z^t), x^{t+1} - x^t \rangle - \frac{L'K}{2} \|x^{t+1} - x^t\|^2 \geq \frac{1}{2c} \|x^{t+1} - x^t\|^2. \quad (1)$$

Next, according to the update of $y^{t+1}$, i.e. $y^{t+1} = y^t + \alpha Ax^{t+1}$, we can obtain

$$K(x^{t+1}, y^{t+1}, z^t) - K(x^{t+1}, y^{t+1}, z^{t+1}) = -\alpha (Ax^{t+1})^T Ax^{t+1}. \quad (2)$$

In addition, by using the update of $z^{t+1}$, i.e., $z^{t+1} = z^t + \beta (z^{t+1} - z^t)$, we can obtain

$$x^{t+1} - z^t = \frac{1}{\beta} (z^{t+1} - z^t), \quad x^{t+1} - z^{t+1} = \frac{1 - \beta}{\beta} (z^{t+1} - z^t).$$

By using above two equalities, we have

$$K(x^{t+1}, y^{t+1}, z^t) - K(x^{t+1}, y^{t+1}, z^{t+1}) = \frac{p}{2} \left(\|z^{t+1} - z^t\|^2 - \|x^{t+1} - z^{t+1}\|^2\right) = \frac{p}{2} \left((z^{t+1} - z^t)^T (x^{t+1} - z^t) + (x^{t+1} - z^{t+1})\right) = \frac{p}{2} \left(2/\beta - 1\right) \|z^t - z^{t+1}\|^2 \geq \frac{p}{2\beta} \|z^t - z^{t+1}\|^2. \quad (3)$$

Then, by combining inequalities (1), (2) and (3), we obtain the desired result.

To bound the dual function and proximal function, we first give the bound on difference of dual function and proximal function during update in the following Lemma.

Lemma 3. Suppose $p > \frac{L}{2}$, then for any $y, y' \in \mathbb{R}^n$, the following inequality holds:

$$\|y - y'\| \geq \sigma_4 \|x(y, z) - x(y', z)\|, \quad (4)$$

where $\sigma_4 = \frac{(Np - L)}{N \sqrt{\lambda_1}}$. 
Proof. First, we define \( \hat{K}(x, y, z) = K(x, y, z) + h(x) \). Note that \( \hat{K}(x, y, z) \) is a strongly convex function with respect to \( x \). Then, it holds that

\[
\begin{align*}
\hat{K}(x(y, z), y', y) - \hat{K}(x(y', z), y, y') \\
= \hat{K}(x(y, z), y, y) - \hat{K}(x(y', z), y, y) - \left( \hat{K}(x(y', z), y, y') - \hat{K}(x(y', z), y, y) \right)
\end{align*}
\]

\[
\leq -\frac{(N_p - L)}{2N} \| x(y', z) - x(y, z) \|^2 + \| y'(y') - y\|^2 \\
\leq -\frac{(N_p - L)}{2N} \| x(y', z) - x(y, z) \|^2 + \sqrt{\lambda_1} \| y'(y') - y \| \| x(y, z) - x(y', z) \|.
\]

Then, using the strongly convexity of \( \hat{K}(x, y, z) \) on \( x \) with modular \( \frac{N_p - L}{N} \), we obtain

\[
\hat{K}(x(y, z), y', y) - \hat{K}(x(y', z), y, y') \geq \frac{N_p - L}{2N} \| x(y, z) - x(y', z) \|^2.
\]

Then, combining the above two inequalities, we can obtain

\[
\| y - y' \| \geq \frac{(N_p - L)}{N \sqrt{\lambda_1}} \| x(y, z) - x(y', z) \|.
\]

The proof is finished. \( \Box \)

Lemma 4. Suppose \( p > -\frac{L}{N} \), then for any \( z, z' \in \mathbb{R}^N \), the following inequalities hold:

\[
\| z - z' \| \geq \sigma_5 \| x^*(z) - x^*(z') \|, \quad (5)
\]

\[
\| z - z' \| \geq \sigma_5 \| x(y, z) - x(y, z') \|, \quad (6)
\]

where \( \sigma_5 = \frac{N_p - L}{N_p} \).

Proof. According to the strongly convexity of function \( g \), we can obtain:

\[
g(x^*(z), z') \geq g(x^*(z'), z')
\]

\[
= g(x^*(z), z') - g(x^*(z'), z) + g(x^*(z'), z') - g(x^*(z'), z') + g_i(x^*(z'), z)' + \frac{1}{2} (z' - z)^T x^*(z') + \| z' \|^2 - 2 \| z \|^2
\]

\[
\leq -\frac{N_p - L}{2N} \| x^*(z) - x^*(z') \|^2 + p(z' - z)^T (x^*(z') - x^*(z)).
\]

On the other hand, using the strongly convexity of \( g \), it holds that

\[
g(x^*(z), z') \geq \frac{N_p - L}{2N} \| x^*(z) - x^*(z') \|^2.
\]

Hence, we have

\[
p(z' - z)^T (x^*(z') - x^*(z)) \geq \frac{N_p - L}{N} \| x^*(z) - x^*(z') \|^2.
\]

Further, according to Cauchy-Schwarz inequality, it implies that

\[
\| x^*(z) - x^*(z') \| \leq \frac{N_p}{N_p - L} \| z - z' \|.
\]
With the above two lemmas, we give the bound on the difference of the dual function and the proximal function during the algorithm, as follows.

**Lemma 5.** For any \( t \geq 0 \), it holds that

\[
\begin{align*}
\| x (y, z) - x (y, z^t) \| & \leq \frac{N p}{N p - L} \| z - z^t \|.
\end{align*}
\]

**Proof.** First, with the definition of \( x (y, z^t) \), we have

\[
\begin{align*}
d \left( y^t, z^t \right) - d \left( y^t, z^t+1 \right) &= K \left( x \left( y^t, z^t+1 \right), y^t, z^t \right) + h \left( x \left( y^t, z^t+1 \right) \right) - K \left( x \left( y^t, z^t \right), y^t, z^t \right) + h \left( x \left( y^t, z^t \right) \right) \\
& \geq K \left( x \left( y^t, z^t+1 \right), y^t, z^t \right) - K \left( x \left( y^t, z^t \right), y^t, z^t \right) \\
& \geq \frac{p}{2} \left( \| x \left( y^t, z^t+1 \right) - z^t+1 \| - \| x \left( y^t, z^t \right) - z^t \| \right)^2 \\
& \geq \frac{p}{2} \left( z^t+1 - z^t \right)^T \left( z^t+1 - z^t - 2x \left( y^t, z^t+1 \right) \right).
\end{align*}
\]

Besides, we can compute the gradient of \( d (y, z) \) as

\[
\nabla_y d (y, z) = Ax (y, z).
\]

Then, for any \( y, y^t \), we have

\[
\| \nabla_y d (y, z) - \nabla_y d (y^t, z) \| = \| Ax (y, z) - Ax \left( y^t, z \right) \| \leq \left( \sqrt{\lambda_1 / \sigma_4} \right) \| y - y^t \|,
\]

which is equivalent to say that \( d (y, z) \) has Lipschitz gradient with respect to \( y \) with Lipschitz constant \( \sqrt{\lambda_1 / \sigma_4} \).

According to the gradient Lipschitz continuity of \( d (y, z) \), it holds

\[
\begin{align*}
d \left( y^t+1, z^t+1 \right) - d \left( y^t, z^t+1 \right) & \geq \left( y^t+1 - y^t, Ax \left( y^t, z^t+1 \right) \right) - \frac{\sqrt{\lambda_1}}{2 \sigma_4} \| y^t+1 - y^t \|^2 \\
& \geq \alpha \left( A \hat{x}^t+1 \right)^T Ax \left( y^t, z^t+1 \right) - \frac{\alpha^2 \sqrt{\lambda_1}}{2 \sigma_4} ||A \hat{x}^t+1||^2.
\end{align*}
\]

Combining (7) and (8), we get the desired result.

**Lemma 6.** For any \( t \geq 0 \), it holds that

\[
M \left( z^t+1 \right) - M \left( z^t \right) \leq p \left( z^t+1 - z^t \right)^T \left( z^t - x^* \left( z^t \right) \right) + \frac{p \tilde{L}}{2} \| z^t - z^t+1 \|^2,
\]

where \( \tilde{L} = \frac{N p}{N p - L} + 1 \).

**Proof.** Recall the definition of \( M \left( z \right) \), we can compute the gradient of \( M \left( z \right) \) as follows:

\[
\nabla M \left( z \right) = p \left( z - x^* \left( z \right) \right).
\]
Then using Lemma 3, we can obtain
\[
\|\nabla M(z) - \nabla M(z')\| = \|p(z - x^*(z')) - p(z - x^*(z'))\| \\
\leq p\|z - z'\| + \|x^*(z) - x^*(z')\| \\
\leq p(1 + 1/\sigma_5)\|z - z'\|.
\]

Therefore, \(M(z)\) is a gradient Lipschitz continuous function with Lipschitz constant \(p(1 + 1/\sigma_5)\). Then, the result directly holds.

The following lemmas give the dual error bound and primal approximation error bound.

**Lemma 7.** For any \(y \in \mathbb{R}^n\), if \(Ax(y,z) = r\), then \(x(y,z) = \arg\min_{x;Ax=r} g(x,z)\).

**Proof.** Recall the definition of \(x(y,z)\):
\[
x(y,z) = \arg\min_x \{g(x,z) + y^T A x\}.
\]
Together with \(Ax(y,z) = r\), \(x(y,z)\) satisfies the optimality condition of the optimization problem.

Therefore, we define \(x^*(r,z) = x(y,z)\), if \(Ax(y,z) = r\).

**Lemma 8.** For any \(r \in \text{Range}(A)\) there exists unique \(v_i(r), i = 2, 3, \cdots, N\), such that for any \(x\) that satisfies \(Ax = r\). In addition, it holds that
\[
x_i = x_1 + v_i(r).
\]
Moreover, defining \(v(r) = (v_1(r)^T, v_2(r)^T, \cdots, v_N(r)^T)^T\), it holds that
\[
\|v(r)\| \leq \frac{1}{\lambda_3}\|r\|,
\]
for some \(\lambda_3 > 0\).

**Proof.** Based on the construction of matrix \(A\), it can be easily verify that when \(G\) is connected, \(\text{rank}(\tilde{A}) = n(N - 1)\).

Besides, for equation \(Ax = r\), we can give the solution set as follows:
\[
x_i = v_i(r) + b,
\]
where \(b\) is an arbitrary vector in \(\mathbb{R}^n\), and \(v\) is a solution with \(v_1 = 0\).

Therefore, for any vector \(x\) that satisfies \(Ax = r\), it can be written as \(x_i = x_1 + v_i(r), i = 1, 2, 3, \cdots, N\).

In addition, we can solve \(v\) by solving equation \(\tilde{A}v(r) = (r^T, 0_n^T)^T\), where
\[
\tilde{A} = \begin{bmatrix} I_n & A \\ 0_{n \times (N-1)} \end{bmatrix}.
\]

When \(G\) is connected, it can be easily verified that \(\text{rank}(\tilde{A}) = nN\), then \(v(r)\) is unique.

Let \(\tilde{W}\) be the matrix generated by removing the first column of \(W\) and \(v(\tilde{r}) = (v_2(\tilde{r})^T, v_3(\tilde{r})^T, \cdots, v_N(\tilde{r})^T)^T\).
Then, because of the connectivity of the graph, \(\text{rank}(\tilde{W}) = N - 1\), which is a full column rank matrix.

Because \(v_1 = 0\), it holds that \((W \otimes I_n)v = (\tilde{W} \otimes I_n)\tilde{v} = r\) and \(\|v\| = \|\tilde{v}\|\).

In addition, we define \(\hat{A} = W^T \tilde{W}\), which is equivalent to remove the first column and first row of \(W^T W\), and we define \(\lambda_3\) as the smallest eigenvalue of \(\hat{A}\).
We can obtain
\[ \sqrt{\lambda_3} \|v(r)\| = \sqrt{\lambda_3} \|\tilde{v}(r)\| \leq \|(\tilde{W} \otimes I_n)\tilde{v}(r)\| = \|r\|. \]

Hence, we get
\[ \|v(r)\| \leq \frac{1}{\sqrt{\lambda_3}} \|r\|, \]
for \( \lambda_3 > 0. \)

**Lemma 9.** Suppose \( p > -L/N, \) then for any \( y \in \mathbb{R}^{nM} \) and \( z \in \mathbb{R}^{nN} \), it holds that
\[ \|x(y, z) - x^*(z)\| \leq \sigma_1 \|Ax(y, z)\|, \]
where \( \sigma_1 = \frac{(3Np - L + \sqrt{N^2 p^2 - L^2})}{2\sqrt{\lambda_3 (Np - L)}}. \)

**Proof.** Let \( \Psi(u, z) = \frac{1}{N} \sum_{i=1}^{N} g_i(u, z_i) \), \( \Psi_r(u) = \frac{1}{N} \sum_{i=1}^{N} g_i(u + v_i(r), z_i). \)

Then, because of the uniqueness of \( v_i(r) \), it is straightforward to obtain
\[ x_1^*(r, z) = \arg \min_u \Psi_r(u) \]
\[ x_1^*(z) = \arg \min_u \Psi(u). \]

Because \( p > -L/N, \) \( g_i(u, z_i) \) is a strongly convex function with modular \(-L + Np\). Thus \( \Psi \) and \( \Psi_r \) are strongly convex. Using the strong convexity of \( \Psi \) and \( \Psi_r \), we obtain
\[ \Psi(x_1^*(r, z)) - \Psi(x_1^*(z)) \geq (-L + Np)\|x_1^*(r, z) - x_1^*(z)\|^2 \]
\[ \Psi_r(x_1^*(z)) - \Psi_r(x_1^*(r, z)) \geq (-L + Np)\|x_1^*(r, z) - x_1^*(z)\|^2. \]

Combining the above two inequality we can obtain
\[ 2(-L + Np)\|x_1^*(r, z) - x_1^*(z)\|^2 \]
\[ \leq \Psi(x_1^*(r, z)) - \Psi(x_1^*(z)) + \Psi_r(x_1^*(z)) - \Psi_r(x_1^*(r, z)) \]
\[ = \Psi(x_1^*(r, z)) - \Psi_r(x_1^*(r, z)) - \Psi_r(x_1^*(z)). \]

Then, using the smoothness and strongly convexity of \( g_i \) for \( i = 2, 3, ..., N \), we obtain
\[ \Psi(x_1^*(r, z)) - \Psi_r(x_1^*(r, z)) - (\Psi(x_1^*(z)) - \Psi_r(x_1^*(z))) \]
\[ = \frac{1}{N} \sum_{i=2}^{N} \left( g_i(x_1^*(r, z), z_i) - g_i(x_1^*(r, z) + v_i(r), z_i) \right) - \frac{1}{N} \sum_{i=2}^{N} \left( g_i(x_1^*(z), z_i) - g_i(x_1^*(z) + v_i(r), z_i) \right) \]
\[ \leq \frac{1}{N} \sum_{i=2}^{N} \left( \langle \nabla g_i(x_1^*(r, z) + v_i(r)), -v_i(r) \rangle + \frac{L + Np}{2} \|v_i(r)\|^2 \right) - \frac{1}{N} \sum_{i=2}^{N} \langle \nabla g_i(x_1^*(z) + v_i(r)), -v_i(r) \rangle \]
\[ \leq \frac{1}{N} \sum_{i=2}^{N} (L + Np)\|x_1^*(r, z) - x_1^*(z)\| \|v_i(r)\| + \frac{(L + Np)}{2} \|v_i(r)\|^2. \]

Then, using the convexity of square function we obtain
\[ \sum_{i=1}^{N} \|v_i(r)\| \leq \sqrt{N}\|v(r)\|. \]

Together with the above inequalities, we obtain
\[ 2(-L + Np)\|x_1^*(r, z) - x_1^*(z)\|^2 \]
\[ \leq \frac{L + Np}{\sqrt{N}} \|x_1^*(r, z) - x_1^*(z)\| \|v(r)\| + \frac{(L + Np)}{2N} \|v(r)\|^2 \]
\[ \leq \frac{L + Np}{\sqrt{N}\lambda_3} \|x_1^*(r, z) - x_1^*(z)\| \|r\| + \frac{(L + Np)}{2N\lambda_3} \|r\|^2. \]
By solving the above quadratic inequality, we can obtain
\[
\|x^*_1 (r, z) - x^*_1 (z)\| \leq \frac{(L + Np + \sqrt{N^2 p^2 - L^2})}{2\sqrt{N\lambda_3} (Np - L)} \|r\|.
\]

By the definition of \(v\), we have \(x(y, z) = x^* (r, z) = \left( x^*_1 (r, z)^T, x^*_1 (r, z)^T, \ldots, x^*_1 (r, z)^T \right)^T + v\) and \(x^* (z) = \left( x^*_1 (z)^T, x^*_1 (z)^T, \ldots, x^*_1 (z)^T \right)^T\).

Therefore, with triangular inequality, we have
\[
\|x (y, z) - x^* (z)\| \leq \sqrt{N}\|x^*_1 (r, z) - x^*_1 (z)\| + \|v\| \leq \sigma_1 \|r\| = \sigma_1 \|Ax(y, z)\|,
\]

where \(\sigma_1 = \frac{(3Np - L + \sqrt{N^2 p^2 - L^2})}{2\sqrt{N\lambda_3} (Np - L)}\). Then, the proof is finished.

**Lemma 10.** (Smooth version of Lemma 9) Suppose \(p > -L/N\) and \(h(\cdot) = 0\), then for any \(y \in \mathbb{R}^{nM}\) and \(z \in \mathbb{R}^{nN}\), it holds that
\[
\|x (y, z) - x^* (z)\| \leq \sigma_1 \|Ax(y, z)\|,
\]

where \(\sigma_1 = \frac{(3Np - L + \sqrt{N^2 p^2 - L^2})}{2\sqrt{N\lambda_3} (Np - L)}\).

**Proof.** First, we define \(\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i\), similar to Lemma 8 for any \(r \in \text{Range}(A)\) there exists a unique \(v_i (r), i = 1, 2, 3, \ldots, N\), such that for any \(x\) that satisfies \(Ax = r\), it holds that
\[
x_i = \bar{x} + v_i (r).
\]

For any vector \(u\) with \(Au = 0\), we have \(u^T(x - (\bar{x}^T, \bar{x}^T, \ldots, \bar{x}^T))^T = 0\). Then, it holds \(u^Tv = 0\).

Therefore, it holds that
\[
\|v\| \leq \frac{1}{\lambda_2} \|Av\| = \frac{1}{\lambda_2} \|r\|,
\]

where \(\lambda_2\) is the smallest nonzero eigenvalue of \(A^TA\).

Similar to the proof of Lemma 9, we define \(\Psi(u, z) = \frac{1}{N} \sum_{i=1}^{N} g_i (u, z_i), \Psi_r (u) = \frac{1}{N} \sum_{i=1}^{N} g_i (u + v_i (r), z_i), \bar{x}^*(r, z) = \arg \min_u \Psi_r (u)\) and \(\bar{x}^* (z) = \arg \min_u \Psi(u)\).

With the same decomposition in Lemma 9 it holds that
\[
\|\bar{x}^* (r, z) - \bar{x}^* (z)\| \leq \frac{(L + Np + \sqrt{N^2 p^2 - L^2})}{2\sqrt{N\lambda_3} (Np - L)} \|r\|.
\]

Together with the definition of \(x^* (r, z) = (\bar{x}^* (r, z)^T, \bar{x}^* (r, z)^T, \ldots, \bar{x}^* (r, z)^T)^T + v\) and \(x^* (z) = (\bar{x}^* (z)^T, \bar{x}^* (z)^T, \ldots, \bar{x}^* (z)^T)^T\), we can get the result.

**Lemma 11.** For a differentiable convex function \(f\) (defined on the \(\mathbb{R}^n\)) with \(L\)-Lipschitz gradient, the following inequality always holds
\[
\frac{1}{L} \|\nabla f(x) - \nabla f(y)\|^2 \leq \langle \nabla f(x) - \nabla f(y), x - y \rangle.
\]

**Proof.** For fix \(x\), we define function \(\Gamma(z) = f(z) - \langle \nabla f(x), z \rangle\).

Then \(\Gamma(z)\) is a convex function with \(L\)-Lipschitz gradient. Beside, the minimum value of function \(\Gamma(\cdot)\) will be \(\Gamma(x) = f(x) - \langle \nabla f(x), x \rangle\).

Meanwhile, using the Lipschitz smoothness of \(\Gamma(\cdot)\), for fix \(y\), we have
\[
\Gamma(z) \leq \Gamma(y) + \langle \nabla \Gamma(y), z - y \rangle + \frac{L}{2} \|z - y\|^2.
\]
Therefore, we can obtain
\[
f(x) - \langle \nabla f(x), x \rangle = \min_z \Gamma(z) \leq \min_z \{\Gamma(y) + \langle \nabla \Gamma(y), z - y \rangle + \frac{L}{2} \|z - y\|^2\}
\]
\[
\leq \Gamma(y) - \frac{1}{2L} \|\nabla \Gamma(y)\|^2 = f(y) - \langle \nabla f(x), y \rangle + \frac{1}{2L} \|\nabla f(y) - \nabla f(x)\|^2.
\]

By rearranging the terms, we can obtain
\[
\frac{1}{2L} \|\nabla f(x) - \nabla f(y)\|^2 \leq f(y) - f(x) + \langle \nabla f(x), x - y \rangle. \tag{9}
\]

By swapping \(x\) and \(y\), we have
\[
\frac{1}{2L} \|\nabla f(x) - \nabla f(y)\|^2 \leq f(x) - f(y) + \langle \nabla f(y), y - x \rangle. \tag{10}
\]

Thus, adding up (9) and (10), we can attain the result. \(\square\)

**Lemma 12.** Suppose \(p > -\frac{L}{N}\) and \(c \leq \frac{1}{\mathcal{L}K} = \frac{N}{L + Np}\), then the following inequalities hold:
\[
\begin{align*}
\|x^{t+1} - x^t\| & \geq \sigma_2 \|x^t - x(y^t, z^t)\|, \tag{11} \\
\|x^{t+1} - x^t\| & \geq \sigma_3 \|x^{t+1} - x(y^t, z^t)\|, \tag{12}
\end{align*}
\]

where
\[
\sigma_2 = \frac{c(\frac{Np - L}{2N})}{\mathcal{L}K}, \\
\sigma_3 = \sigma_2 / (1 - \sigma_2).
\]

**Proof.** For (11), we define \(\hat{g}(x; v) = \|x\|^2 - 2x^Tv + h(x)\), \(v_1 = x^t - c\nabla_x K(x^t, y^t, z^t)\) and \(v_2 = x(y^t, z^t) - c\nabla_x K(x(y^t, z^t), y^t, z^t)\).

Then, with the update iteration of \(x^t\) we have
\[
x^{t+1} = \arg \min_x \hat{g}(x, v_1).
\]

Beside, with the definition of \(x(y^t, z^t)\) we have
\[
x(y^t, z^t) = \arg \min_x \hat{g}(x, v_2).
\]

It is obvious that \(\hat{g}(\cdot, v)\) is a strongly convex function with modular 1. Then, using the strong convexity of \(\hat{g}(\cdot, v)\), we have
\[
\begin{align*}
\hat{g}(x^{t+1}; v_2) - \hat{g}(x(y^t, z^t); v_2) & \geq \|x^{t+1} - x(y^t, z^t)\|^2, \tag{13} \\
\hat{g}(x(y^t, z^t); v_1) - \hat{g}(x^{t+1}; v_1) & \geq \|x^{t+1} - x(y^t, z^t)\|^2. \tag{14}
\end{align*}
\]

On the other hand, by the definition of \(\hat{g}\), we have
\[
\begin{align*}
\hat{g}(x^{t+1}; v_1) - \hat{g}(x^{t+1}; v_2) = -(x^{t+1}, v_1 - v_2), \tag{15} \\
\hat{g}(x(y^t, z^t); v_1) - \hat{g}(x(y^t, z^t); v_2) = -(x(y^t, z^t), v_1 - v_2). \tag{16}
\end{align*}
\]

Combining (13), (14), (15) and (16), we have
\[
\|x^{t+1} - x(y^t, z^t)\|^2 \leq \langle x^{t+1} - x(y^t, z^t), v_1 - v_2 \rangle. \tag{17}
\]

Then, using Cauchy–Schwartz inequality, we have
\[
\|x^{t+1} - x(y^t, z^t)\| \leq \|v_1 - v_2\|. \tag{18}
\]
According to the Lipschitz continuity of $\nabla_x K(\cdot, y, z)$ and convexity of $K(\cdot, y, z)$, and Lemma 11 we have

$$\frac{N}{L+Np} \| \nabla_x K(x^t, y^t, z^t) - \nabla_x K(x(y^t, z^t), y^t, z^t) \|^2 \leq \langle x^t - x(y^t, z^t), \nabla_x K(x^t, y^t, z^t) - \nabla_x K(x(y^t, z^t), y^t, z^t) \rangle.$$  

Substituting the above inequality into (19), we have

$$\|v_1 - v_2\|^2 \leq \|x^t - x(y^t, z^t)\|^2 - c\langle x^t - x(y^t, z^t), \nabla_x K(x^t, y^t, z^t) - \nabla_x K(x(y^t, z^t), y^t, z^t) \rangle$$

$$\leq \|x^t - x(y^t, z^t)\|^2 - c\langle x^t - x(y^t, z^t), \nabla_x K(x^t, y^t, z^t) - \nabla_x K(x(y^t, z^t), y^t, z^t) \rangle,$$

where the last inequality holds because $c \leq \frac{N}{L+Np}$.

Then, because $K(\cdot, y, z)$ is a convex function with $(p - \frac{L}{N})$-Lipschitz gradient, according to Lemma 11 we have

$$\langle x^t - x(y^t, z^t), \nabla_x K(x^t, y^t, z^t) - \nabla_x K(x(y^t, z^t), y^t, z^t) \rangle \geq (p - \frac{L}{N})\|x^t - x(y^t, z^t)\|^2.$$

Therefore, we have

$$\|v_1 - v_2\|^2 \leq (1 - c(p - \frac{L}{N}))\|x^t - x(y^t, z^t)\|^2.$$

Note that $1 - c(p - \frac{L}{N}) < (1 - c(p - \frac{L}{N})/2)^2$. Then we have

$$\|v_1 - v_2\|^2 \leq (1 - c(p - \frac{L}{N}))^2\|x^t - x(y^t, z^t)\|^2.$$

Hence,

$$\|x^{t+1} - x(y^t, z^t)\| \leq \|v_1 - v_2\| \leq (1 - c(p - \frac{L}{N})/2)\|x^t - x(y^t, z^t)\|.$$

Besides, according to triangular inequality, we have

$$\|x^t - x^{t+1}\| \geq \|x^t - x(y^t, z^t)\| - \|x^{t+1} - x(y^t, z^t)\|$$

$$\geq \frac{c(p - \frac{L}{N})}{2}\|x^t - x(y^t, z^t)\|,$$

which yields Eq. (11).

In addition, by (20), we have

$$\|x^t - x^{t+1}\| \geq \frac{c(p - \frac{L}{N})}{2}\|x^t - x(y^t, z^t)\|$$

$$\geq \frac{c(p - \frac{L}{N})/2}{1 - c(p - \frac{L}{N})/2}\|x^{t+1} - x(y^t, z^t)\|$$

$$= \frac{\sigma_2}{1 - \sigma_2}\|x^{t+1} - x(y^t, z^t)\|,$$

which gives the result (12).
Proof of Theorem 1. Let $\Phi(x, y, z) = K(x, y, z) + h(x) - 2d(y, z) + 2M(z)$.

Recall the definition of $d(y, z)$ and $M(z)$

$$d(y, z) = \min_{x \in \mathbb{R}^{m \times n}} K(x, y, z) + h(x),$$

$$M(z) = \min_{x \in \mathbb{R}^{m \times n}, Ax = 0} \left( f(x) + h(x) + \frac{p}{2} \|x - z\|^2 \right),$$

it holds that

$$M(z) \geq d(y, z),$$

$$K(x, y, z) + h(x) \geq d(y, z).$$

Then,

$$\Phi(x, y, z) = K(x, y, z) + h(x) - 2d(y, z) + 2M(z)$$

$$\geq K(x, y, z) + h(x) - d(y, z) + M(z) - d(y, z) + M(z) \geq M(z)$$

$$\geq \min_{x} \{ f(x) + h(x) \} \geq f_0.$$

Using Lemmas 2, 5 and 6 it holds that

$$\Phi(x', y', z') - \Phi(x^{t+1}, y^{t+1}, z^{t+1})$$

$$\geq \frac{1}{2c} \|x^{t+1} - x'\|^2 - \alpha (Ax^{t+1})^T Ax^{t+1} + \frac{p}{2\beta} \|z' - z^{t+1}\|^2 + 2\alpha (Ax^{t+1})^T Ax(y', z^{t+1})$$

$$- \frac{\alpha^2 \sqrt{\lambda_1}}{\sigma_4} \|Ax^{t+1}\|^2 + p \left( z^{t+1} - z' \right)^T \left( z^{t+1} + z' - 2x(y', z^{t+1}) \right)$$

$$- 2p \left( z^{t+1} - z' \right)^T (z' - x^*(z')) - p\tilde{L}\|z' - z^{t+1}\|^2$$

$$\geq \frac{1}{2c} \|x^{t+1} - x'\|^2 + \frac{p}{2\beta} \|z' - z^{t+1}\|^2 + \alpha (Ax^{t+1})^T A(x^{t+1} - x(y', z^{t+1})) - \alpha \|Ax^{t+1}\|^2$$

$$- \alpha (Ax^{t+1})^T Ax(y', z^{t+1}) - \frac{3\alpha^2 \sqrt{\lambda_1}}{\sigma_4} \left( \|A(x^{t+1} - x(y', z^{t+1}))\|^2 + \|Ax^{t+1}\|^2 + \|Ax(y', z^{t+1})\|^2 \right)$$

$$+ 2\alpha (Ax^{t+1})^T Ax(y', z^{t+1}) + p \left( z^{t+1} - z' \right)^T \left( z^{t+1} - z' - 2(x(y', z^{t+1}) - x^*(z')) \right)$$

$$- p\tilde{L}\|z' - z^{t+1}\|^2$$

$$\geq \frac{1}{2c} \|x^{t+1} - x'\|^2 + \frac{p}{2\beta} \|z' - z^{t+1}\|^2 - \alpha \sqrt{\lambda_1} \|e^{t+1}\||A(x^{t+1} - x(y', z^{t+1}))\|^2$$

$$- \frac{3\alpha^2 \sqrt{\lambda_1}}{\sigma_4} \left( \|A(x^{t+1} - x(y', z^{t+1}))\|^2 + \lambda_1 \|e^{t+1}\|^2 + \|Ax(y', z^{t+1})\|^2 \right)$$

$$+ 2\alpha (Ax^{t+1})^T Ax(y', z^{t+1}) + p \left( z^{t+1} - z' \right)^T \left( z^{t+1} - z' - 2(x(y', z^{t+1}) - x^*(z')) \right)$$

$$- p\tilde{L}\|z' - z^{t+1}\|^2.$$

First, we can bound the terms related to the $e^{t+1}$:

$$- \alpha \sqrt{\lambda_1} \|e^{t+1}\||A(x^{t+1} - x(y', z^{t+1}))\|^2$$

$$\geq - \frac{\alpha}{2} \|A(x^{t+1} - x(y', z^{t+1}))\|^2$$

$$- \frac{\alpha}{2} \|Ax(y', z^{t+1})\|^2$$

$$- \left( \frac{3\alpha^2 \lambda^{3/2}}{\sigma_4} + \alpha \lambda_1 \right) \|e^{t+1}\|^2.$$
Then, for the terms related to \( A \), we can obtain

\[
\begin{align*}
- \alpha \sqrt{\lambda_1} \| \epsilon^{t+1} \| \| A (x^{t+1} - x(y^t, z^{t+1})) \| - \alpha \| Ax^{t+1} \|^2 + 2 \alpha (Ax^{t+1})^T Ax(y^t, z^{t+1}) \\
- \alpha \sqrt{\lambda_1} \| \epsilon^{t+1} \| \| Ax(y^t, z^{t+1}) \| - \frac{3a^2 \sqrt{\lambda_1}}{\sigma_4} \| (A (x^{t+1} - x(y^t, z^{t+1})) \|^2 + \lambda_1 \| \epsilon^{t+1} \|^2 + \| Ax(y^t, z^{t+1}) \|^2 \\
\geq - \frac{\alpha}{2} \| A (x^{t+1} - x(y^t, z^{t+1})) \|^2 - \frac{\alpha}{2} \| Ax(y^t, z^{t+1}) \|^2 - \left( \frac{3a^2 \lambda_1^{3/2}}{\sigma_4} + \lambda_1 \right) \| \epsilon^{t+1} \|^2 \\
- \left( \alpha \| Ax^{t+1} \|^2 - 2 \alpha (Ax^{t+1})^T Ax(y^t, z^{t+1}) + \alpha \| Ax(y^t, z^{t+1}) \|^2 \right) + \alpha \| Ax(y^t, z^{t+1}) \|^2 \\
- \frac{3a^2 \sqrt{\lambda_1}}{\sigma_4} \| (A (x^{t+1} - x(y^t, z^{t+1})) \|^2 + \| Ax(y^t, z^{t+1}) \|^2 \\
= \left( \frac{\alpha}{2} - \frac{3a^2 \sqrt{\lambda_1}}{\sigma_4} \right) \| Ax(y^t, z^{t+1}) \|^2 - \left( \frac{3a^2 \lambda_1^{3/2}}{\sigma_4} + \lambda_1 \right) \| \epsilon^{t+1} \|^2 \\
\geq \left( \frac{\alpha}{2} - \frac{3a^2 \sqrt{\lambda_1}}{\sigma_4} \right) \| Ax(y^t, z^{t+1}) \|^2 - \left( \frac{3a^2 \lambda_1^{3/2}}{\sigma_4} + \lambda_1 \right) \| \epsilon^{t+1} \|^2 \\
= \left( \frac{\alpha}{2} - \frac{3a^2 \sqrt{\lambda_1}}{\sigma_4} \right) \| Ax(y^t, z^{t+1}) \|^2 - \left( \frac{3a^2 \lambda_1^{3/2}}{\sigma_4} + \lambda_1 \right) \| \epsilon^{t+1} \|^2 \\
\geq \left( \frac{\alpha}{2} - \frac{3a^2 \sqrt{\lambda_1}}{\sigma_4} \right) \| Ax(y^t, z^{t+1}) \|^2 - \left( \frac{3a^2 \lambda_1^{3/2}}{\sigma_4} + \lambda_1 \right) \| \epsilon^{t+1} \|^2 \\
\geq \left( \frac{\alpha}{2} - \frac{3a^2 \sqrt{\lambda_1}}{\sigma_4} \right) \| Ax(y^t, z^{t+1}) \|^2 - \left( \frac{3a^2 \lambda_1^{3/2}}{\sigma_4} + \lambda_1 \right) \| \epsilon^{t+1} \|^2,
\end{align*}
\]

where the last inequality is due to Lemma 12 and Lemma 4.

Further, for the terms related to \( z \), we have

\[
\begin{align*}
p (z^{t+1} - z^t)^T (z^{t+1} - z^t - 2 (x(y^t, z^{t+1}) x^*(z^t))) - p \tilde{L} \| z^t - z^{t+1} \|^2 + \frac{p}{2 \beta} \| z^t - z^{t+1} \|^2 \\
= p (z^{t+1} - z^t)^T (z^{t+1} - z^t - 2 (x(y^t, z^{t+1}) - x^*(z^{t+1})) - 2 (x^*(z^{t+1}) - x^*(z^t))) \\
- p \tilde{L} \| z^t - z^{t+1} \|^2 + \frac{p}{2 \beta} \| z^t - z^{t+1} \|^2 \\
\geq \left( \frac{p}{2 \beta} + p - p \tilde{L} \right) \| z^t - z^{t+1} \|^2 - 2p (z^{t+1} - z^t)^T (x(y^t, z^{t+1}) - x^*(z^{t+1})) \\
- 2p (z^{t+1} - z^t)^T (x^*(z^{t+1}) - x^*(z^t)) \\
\geq \left( \frac{p}{2 \beta} + p - p \tilde{L} \right) \| z^t - z^{t+1} \|^2 - 2p \| z^{t+1} - z^t \| \| x(y^t, z^{t+1}) - x^*(z^{t+1}) \| \\
- 2p \| z^{t+1} - z^t \| \| x^*(z^{t+1}) - x^*(z^t) \|. \tag{22}
\end{align*}
\]
By using Lemma 1 and Lemma 2, it further holds that
\[
p (z^{t+1} - z^t)^T (z^{t+1} - z^t - 2 (x (y^t, z^{t+1}) x^* (z^t))) - pL \| z^t - z^{t+1} \|^2 + \frac{p}{2\beta} \| z^t - z^{t+1} \|^2 \\
\geq \left( \frac{p}{2\beta} + p - p\tilde{L} \right) \| z^t - z^{t+1} \|^2 - \frac{4p^2\sigma_1^2}{\alpha} \| z^{t+1} - z^t \|^2 + \frac{\alpha}{4\sigma_1^2} \| x (y^t, z^{t+1}) - x^* (z^{t+1}) \|^2 \\
- \frac{2p}{\sigma_5} \| z^{t+1} - z^t \|^2
\] (23)

Besides, by taking \( \alpha \) and \( \beta \) sufficient small so that we can obtain
\[
\frac{\alpha}{8} \geq 3\alpha^2\sqrt{\lambda_1} / \sigma_4, \\
\frac{1}{2c} \frac{-\alpha \lambda_1}{\sigma_3} - \frac{6\alpha^2\lambda_1^{3/2}}{\sigma_4 \sigma_3} - \frac{(1 - \delta)^2}{\delta^2} \left( \frac{3\alpha^2\lambda_1^{3/2}}{\sigma_4} + \alpha \lambda_1 \right) \geq \frac{1}{4c}, \\
\frac{p}{2\beta} + p - p\tilde{L} - \frac{4p^2\sigma_1^2}{\alpha} - \frac{2p}{\sigma_5} \left( \frac{\alpha \lambda_1}{\sigma_5} + \frac{6\alpha^2\lambda_1^{3/2}}{\sigma_4 \sigma_5} \right) \geq \frac{p}{4\beta}, \\
\beta \leq 1.
\] (24)

Then combining (21), (22), (23), (24), and Lemma 1, we can obtain
\[
\Phi (x^0, y^0, z^0) - f \\
\geq \sum_{t=0}^{T-1} \Phi (x^t, y^t, z^t) - \Phi (x^{t+1}, y^{t+1}, z^{t+1}) \\
\geq \sum_{t=0}^{T-1} \left( \frac{1}{2c} \frac{-\alpha \lambda_1}{\sigma_3} - \frac{6\alpha^2\lambda_1^{3/2}}{\sigma_4 \sigma_3} - \frac{(1 - \delta)^2}{\delta^2} \left( \frac{3\alpha^2\lambda_1^{3/2}}{\sigma_4} + \alpha \lambda_1 \right) \| x^{t+1} - x^t \|^2 + \frac{\alpha}{4} \frac{3\alpha^2\sqrt{\lambda_1}}{\sigma_4} \| Ax (y^t, z^{t+1}) \|^2 \\
- \left( \frac{3\alpha^2\lambda_1^{3/2}}{\sigma_4} + \alpha \lambda_1 \right) \| \epsilon_{t+1} \|^2 + \frac{p}{4\beta} \| z^t - z^{t+1} \|^2 \\
+ \left( \frac{1}{2c} \frac{-\alpha \lambda_1}{\sigma_3} - \frac{6\alpha^2\lambda_1^{3/2}}{\sigma_4 \sigma_3} - \frac{(1 - \delta)^2}{\delta^2} \left( \frac{3\alpha^2\lambda_1^{3/2}}{\sigma_4} + \alpha \lambda_1 \right) \right) \| x^{t+1} - x^t \|^2 \\
+ \frac{\alpha}{8} \| Ax (y^t, z^{t+1}) \|^2 + \frac{p}{4\beta} \| z^t - z^{t+1} \|^2 \\
\right) \\
\geq \sum_{t=0}^{T-1} \frac{1}{4c} \| x^{t+1} - x^t \|^2 + \frac{\alpha}{8} \| Ax (y^t, z^{t+1}) \|^2 + \frac{p}{4\beta} \| z^t - z^{t+1} \|^2.
\]

According to the above inequality, we define \( C = \Phi (x^0, y^0, z^0) - f \), then it holds that for any \( T > 0 \), there exists an \( s \in \{0, 1, \cdots, T - 1\} \) such that
\[
\| x^s - x^{s+1} \|^2 \leq 4cC/T, \\
\| Ax (y^s, z^{s+1}) \|^2 \leq \frac{8}{\alpha} C/T, \\
\| x^{s+1} - z^s \|^2 = \frac{1}{\beta^2} \| z^{s+1} - z^s \|^2 \leq \frac{4}{\beta p} C/T, \\
\| z^{s+1} - z^s \|^2 \leq \frac{4\beta}{p} C/T.
\]
Besides, recall the update of $x^{s+1}$, i.e. $x^{s+1} = \arg\min_{x} \left( \langle \nabla x, K(x^i, y^i, z^i), x_i - x^i \rangle + h_i(x_i) + \frac{1}{2\nu}\|x_i - x^i\|^2 \right)$, with the optimally condition we can obtain

$$0 \in \nabla x K(x^s, y^s, z^s) + \frac{1}{c}(x^{s+1} - x^s) + \partial h(x^{s+1}).$$

Therefore, let

$$\nu = \nabla x K(x^{s+1}, y^s, z^s) - \nabla x K(x^s, y^s, z^s) - \frac{1}{c}(x^{s+1} - x^s) - p(x^{s+1} - z^s).$$

we can obtain that

$$\nu \in \nabla f(x) + A^T y^s + \partial h(x^{s+1}).$$

Moreover, we have

$$\|\nu\| \leq \left( \frac{L}{N} + p \right) \|x^{s+1} - x^s\| + \frac{1}{c} \|x^{s+1} - x^s\| + p\|x^{s+1} - z^s\|$$

$$\leq \left( \frac{L}{N} + p + \frac{1}{c} \right) \sqrt{4c + \frac{2\sqrt{p}}{\sqrt{\beta}}} \sqrt{\frac{C}{T}}.$$

On the other hand, it holds that

$$\|Ax^{s+1}\| \leq \|Ax(y^s, z^{s+1})\| + \|A(x^{s+1} - x (y^s, z^s))\| + \|A(x(y^s, z^s) - x(y^s, z^{s+1}))\|$$

$$\leq \sqrt{C}/\sqrt{T}(\frac{8}{\sqrt{\alpha}} + \frac{\sqrt{\Lambda_1}4c}{\sigma_3} + \frac{\sqrt{\beta}}{\sqrt{\beta_5}}).$$

Hence, letting

$$B = \left( \frac{L}{N} + p + \frac{1}{c} \right) \sqrt{4c + \frac{2\sqrt{p}}{\sqrt{\beta}}} \sqrt{\frac{N}{\sqrt{\alpha}}} + \frac{L}{\sqrt{\Lambda_1}2} \left( \frac{8}{\sqrt{\alpha}} + \frac{\sqrt{\Lambda_1}4c}{\sigma_3} + \frac{\sqrt{\beta_5}}{\sqrt{\beta_5}} \right) \sqrt{C}. \quad (25)$$

Then, $(x^{s+1}, y^s)$ is a $B/\sqrt{T}$-solution.

**Proof of Corollary 1.** In the smooth case, we define $\gamma = \frac{\lambda_4}{\lambda_2}$ and in non-smooth case we define $\gamma = \frac{\lambda_4}{\lambda_2} \geq \frac{\lambda_4}{\lambda_2}$.

It is easy to check $\alpha$ and $\beta$ in corollary 1 that satisfy inequalities (24).

By plugging $c$, $p$, $\alpha$ and $\beta$ in to equation (25), we can get the result in smooth case.

With $\frac{\lambda_4}{\lambda_2} \geq \frac{\lambda_4}{\lambda_2}$ we can get the result in the nonsmooth case.

\qed

## 2 Proof of Remark 1

**Proof of Remark 1.** The proof will only consider the smooth case, because the definition of $2\epsilon^2$-stationary points in [Hong et al. (2017)] and $4\epsilon^2$-stationary points in [Tang et al. (2019)] are in the setting of smooth objective function.

First we show that our definition is the sufficient condition for $2\epsilon^2$-stationary points in [Hong et al. (2017)], i.e. $\|\frac{1}{N} \sum_{i=1}^{N} \nabla f_i(x_i)\|^2 + \frac{L}{\sqrt{\Lambda_2}} \sum_{(i,j) \in E} \|x_i - x_j\|^2 \leq 2\epsilon^2$.

From Definition 1, we have

$$N\|\nabla f(x) + A^T y\|^2 \leq \epsilon^2.$$ 

Besides, recall that $\mu = (\mu_1, \mu_2, \cdots, \mu_N) = A^T y$ in the algorithm and $A1 = 0$.

Thus, $\sum_{i=1}^{N} \mu_i = 0$. 
Then, it holds that

\[
\left\| \frac{1}{N} \sum_{i=1}^{N} \nabla f_i(x_i) \right\|^2 = \left\| \frac{1}{N} \sum_{i=1}^{N} \nabla f_i(x_i) + \mu_i \right\|^2 \\
\leq N \left( \frac{1}{N^2} \sum_{i=1}^{N} \| \nabla f_i(x_i) + \mu_i \|^2 \right) \\
= N \| \nabla f(x) \|^2 \leq \epsilon^2.
\]

On the other hand, we have

\[
\frac{L^2}{N \lambda_2} \| Ax \|^2 \leq \epsilon^2.
\]

Hence, it holds that

\[
\frac{L}{N \lambda_2} \sum_{(i,j) \in E} \| x_i - x_j \|^2 = \frac{L}{N \lambda_2} \| Ax \|^2 \leq \frac{\epsilon^2}{L}.
\]

Because \( L \) can take arbitrary large value, we can assume \( L > 1 \) and the following statement holds:

\[
\left\| \frac{1}{N} \sum_{i=1}^{N} \nabla f_i(x_i) \right\|^2 + \frac{L}{N \lambda_2} \sum_{(i,j) \in E} \| x_i - x_j \|^2 \leq 2 \epsilon^2.
\]

Then, for the \( 4 \epsilon^2 \)-stationary point in [Tang et al. (2019)], it is defined as

\[
\left\| \frac{1}{N} \sum_{i=1}^{N} \nabla f_i(\bar{x}) \right\|^2 \leq 4 \epsilon^2,
\]

where \( \bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i \).

Let \( \bar{x} = (\bar{x}^T, \bar{x}^T, \cdots, \bar{x}^T)^T \in \mathbb{R}^{nN} \).

Recall the definition of \( x = (x_1^T, x_2^T, \cdots, x_N^T)^T \) and \( A \), then we obtain if \( Av = 0 \) then \( v^T (\bar{x} - x) = 0 \).

Besides, we have \( A\bar{x} = 0 \).

Therefore, with definition of \( \lambda_2 \) (the smallest nonzero eigenvalue of \( A^T A \)), we have

\[
\sum_{i=1}^{N} \| \bar{x} - x_i \|^2 = \| \bar{x} - x \|^2 \leq \frac{1}{\lambda_2} \| A(\bar{x} - x) \|^2 = \frac{1}{\lambda_2} \| Ax \|^2.
\]

Combining the above inequality and the Lipschitz gradient of \( f_i \), it holds that

\[
\left\| \frac{1}{N} \sum_{i=1}^{N} \nabla f_i(\bar{x}) \right\|^2 = \left\| \frac{1}{N} \sum_{i=1}^{N} \nabla f_i(\bar{x}) - \nabla f_i(x_i) + \nabla f_i(x_i) \right\|^2 \\
\leq 2N \left( \frac{1}{N^2} \sum_{i=1}^{N} \| \nabla f_i(x_i) \|^2 \right) + \frac{2L^2}{N} \sum_{i=1}^{N} \| \bar{x} - x_i \|^2 \\
\leq 2N \| \nabla f(x) + A^T y \|^2 + \frac{2L^2}{N} \sum_{i=1}^{N} \| \bar{x} - x_i \|^2 \\
\leq 2N \| \nabla f(x) + A^T y \|^2 + \frac{2L^2}{N \lambda_2} \| Ax \|^2 \\
\leq 4 \epsilon^2.
\]

Hence, our definition is the sufficient condition for the \( 2 \epsilon^2 \)-solution in [Hong et al. (2017)] and the \( 4 \epsilon^2 \)-solution in [Tang et al. (2019)].
References
