Supplementary material:
Global Convergence of Online Optimization for Nonlinear Model Predictive Control

A  Expression of Newton System

For future references, we explicitly write out each component of \( \mathbf{F} \). For stage \( k \), we let \( H_k(z_k, \lambda_k) = \nabla z_k g_k(z_k) - \lambda_k^T f_k(z_k) \), \( A_k(z_k) = \nabla z_k^T f_k(z_k) \) and \( B_k(z_k) = \nabla u_k f_k(z_k) \). Then, we have
\[
H^t(\tilde{z}_t, \tilde{\lambda}_t) = \text{diag} \left( H_{t,1}, \ldots, H_{M_t-1}, \nabla z_{M_t} g_{M_t}(x_{M_t}, 0) + \mu I \right)
\] (12)
with \( H_k = H_k(z_k, \lambda_k) \) for \( k \in [t, M_t - 1] \), and have
\[
G^t(\tilde{z}_t) = \begin{pmatrix}
-I - B_t - I \\
-A_{t+1} - B_{t+1} - I \\
\vdots \\
-A_{M_t-1} - B_{M_t-1} - I 
\end{pmatrix}
\] (13)
with \( A_k = A_k(z_k) \) and \( B_k = B_k(z_k) \). The gradient of Lagrangian \( L^t(\cdot) \) on the right side of \( \mathbf{F} \) can be expressed as
\[
\nabla_{\tilde{z}_t} L^t(\tilde{z}_t, \tilde{\lambda}_t; \tilde{x}_t) = \begin{pmatrix}
\nabla_{\tilde{z}_t} g_t(z_t) + \lambda_{t-1}^T (z_t) \lambda_t \\
\nabla_{\tilde{z}_t} g_t(z_t) + B_t^T (z_t) \lambda_t \\
\vdots \\
\nabla_{\tilde{z}_{M_t-1}} g_{M_t-1}(z_{M_t-1}) + \lambda_{M_t-2}^T (z_{M_t-1}) \lambda_{M_t-1} \\
\nabla_{\tilde{z}_{M_t-1}} g_{M_t-1}(z_{M_t-1}) + B_{M_t-1}^T (z_{M_t-1}) \lambda_{M_t-1} \\
\vdots \\
\nabla_{\tilde{z}_t} g_t(z_t) + \lambda_{M_t-1}^T (z_t) \lambda_{M_t-1}
\end{pmatrix}
\]
\[
\nabla_{\tilde{\lambda}_t} L^t(\tilde{z}_t, \tilde{\lambda}_t; \tilde{x}_t) = \begin{pmatrix}
x_t - \tilde{x}_t \\
x_{t+1} - f_t(z_t) \\
\vdots \\
x_{M_t} - \tilde{f}_{M_t-1}(z_t)
\end{pmatrix}
\] (14)
We also explicitly write out the gradient of the augmented Lagrangian \( \mathbf{F} \) by
\[
\begin{pmatrix}
\nabla_{\tilde{z}_t} L^t_{\eta_t} \\
\nabla_{\tilde{\lambda}_t} L^t_{\eta_t}
\end{pmatrix} = \begin{pmatrix}
I + \eta_t H^t & \eta_t (G^t)^T \\
\eta_t G^t & I
\end{pmatrix} \begin{pmatrix}
\nabla_{\tilde{z}_t} L^t \\
\nabla_{\tilde{\lambda}_t} L^t
\end{pmatrix}.
\] (15)

B  Proof of Theorem 4.4

We first have a simple observation: by Assumptions 4.1, 4.2, for any \((\tilde{z}_t, \tilde{\lambda}_t) \in Z \otimes \Lambda \) by \((\tilde{z}_t, \tilde{\lambda}_t) \in Z \times \Lambda \) for all stages \( k \) of the \( t \)-th subproblem, \( \|G^t(\tilde{z}_t)\| \leq 1 + 2 \bar{Y}, \|H^t(\tilde{z}_t, \tilde{\lambda}_t)\| \leq \bar{Y}' + \mu \), and
\[
\|\nabla ((G^t)^T \nabla_{\tilde{\lambda}_t} L^t)(\tilde{z}_t, \tilde{\lambda}_t; \tilde{x}_t)\| \leq \bar{Y}', \quad \|\nabla (H^t \nabla_{\tilde{z}_t} L^t)(\tilde{z}_t, \tilde{\lambda}_t; \tilde{x}_t)\| \leq \bar{Y}' + \mu^2
\] (16)
for some constant \( \bar{Y}' \) not depending on \( \mu \). This is from the definitions (12), (14) and noting that only the last block of \( H' \) and the last row of \( \nabla_{\tilde{z}_t} L^t \) contain \( \mu \). We can also replace \( \bar{Y} \) in Assumption 4.2 by \( \bar{Y} \leftarrow (1 + 2 \bar{Y}) \vee \bar{Y}' \vee \delta \) and require \( \mu \geq \bar{Y} \). Then we have \( \|G^t\| \leq \bar{Y}, \|B^t\| \vee \|H^t\| \leq 2 \mu, \|\nabla ((G^t)^T \nabla_{\tilde{\lambda}_t} L^t)\| \leq \bar{Y}, \) and \( \|\nabla (H^t \nabla_{\tilde{z}_t} L^t)\| \leq 2 \mu^2 \). By the definition of \( H^t \) in (12), without loss of generality we let the last block of \( B^t = \mu I \).

We then provide a formula for the KKT matrix inverse. We suppress the index \( t \) since the results hold for any \( t \geq 0 \).

**Lemma B.1.** Let \( G^T = YK \) where \( Y \) has orthonormal columns that span \( \text{Im}(G^T) \) and \( K \) is a nonsingular square matrix (since \( G^T \) has full column rank), and let \( Z \) have orthonormal columns that span \( \text{Ker}(G) \). If \( Z^T BZ \) is invertible, then
\[
S := \begin{pmatrix}
B & G^T \\
G & 0
\end{pmatrix}^{-1} = \begin{pmatrix}
S_1 & S_2^T \\
S_2 & S_3
\end{pmatrix}
\]
where
\[ S_1 = Z(Z^T BZ)^{-1}Z^T, \]
\[ S_2 = K^{-1}Y^T(I - BZ(Z^T BZ)^{-1}Z^T), \]
\[ S_3 = K^{-1}Y^T(BZ(Z^T BZ)^{-1}Z^T(B - B))YK^{-1}. \]

Under Assumption 4.2, we have \( ||S|| \leq 5Y^2\mu^2/\gamma_{RH}. \)

Given Lemma B.1, we apply (3) and (15) and have
\[
\begin{pmatrix}
\nabla \tilde{z} \\
\nabla \tilde{\lambda}
\end{pmatrix}
^T
\begin{pmatrix}
\Delta \tilde{z} \\
\Delta \tilde{\lambda}
\end{pmatrix} = -\begin{pmatrix}
\nabla \tilde{z} \nabla \tilde{L}^0 \\
\nabla \tilde{\lambda} \nabla \tilde{L}^0
\end{pmatrix}
^{T}
\begin{pmatrix}
B & G \\
G & 0
\end{pmatrix}
^{-1}
\begin{pmatrix}
I + \eta_2 H & \eta_1 G^T \\
\eta_2 G & I
\end{pmatrix}
\begin{pmatrix}
\nabla \tilde{z} \nabla \tilde{L}^0 \\
\nabla \tilde{\lambda} \nabla \tilde{L}^0
\end{pmatrix}.
\]

By Lemma B.1 we define \( W = I - Z(Z^T BZ)^{-1}Z^T B \) and have
\[
\begin{pmatrix}
B & G \\
G & 0
\end{pmatrix}
^{-1}
\begin{pmatrix}
I + \eta_2 H & \eta_1 G^T \\
\eta_2 G & I
\end{pmatrix}
= \begin{pmatrix}
\eta_1 I + Z(Z^T BZ)^{-1}Z^T \{ I + \eta_2 (H - B) \} & WY(K^{-1})^T \\
K^{-1}Y^TW^T \{ I + \eta_2 (H - B) \} & \eta_1 I - K^{-1}Y^T B W Y (K^{-1})^T
\end{pmatrix}
=: W_1 + W_2 + W_3,
\]
where
\[
W_1 = \begin{pmatrix}
\frac{\eta_2}{2} I & 0 \\
0 & \frac{\eta_2}{2} I
\end{pmatrix},
\]
\[
W_2 = \begin{pmatrix}
\frac{\eta_2}{2} I & \frac{\eta_2}{2} I \\
K^{-1}Y^TW^T & -\frac{\eta_2}{2} I - K^{-1}Y^T B W Y (K^{-1})^T
\end{pmatrix},
\]
\[
W_3 = \begin{pmatrix}
Z(Z^T BZ)^{-1}Z^T \{ I + \eta_2 (H - B) \} & 0 \\
\eta_2 K^{-1}Y^TW^T (H - B) & 0
\end{pmatrix}.
\]

We deal with each term separately. First, we have
\[
\begin{pmatrix}
\nabla \tilde{z} \nabla \tilde{L}^0 \\
\nabla \tilde{\lambda} \nabla \tilde{L}^0
\end{pmatrix}
W_3 \begin{pmatrix}
\nabla \tilde{z} \nabla \tilde{L}^0 \\
\nabla \tilde{\lambda} \nabla \tilde{L}^0
\end{pmatrix}
\]
\[
= \nabla \tilde{z} \nabla \tilde{L}^0 Z(Z^T BZ)^{-1}Z^T \nabla \tilde{z} \nabla \tilde{L}^0 + \eta_2 \nabla \tilde{\lambda} \nabla \tilde{L}^0 Z(Z^T BZ)^{-1}Z^T (H - B) \nabla \tilde{z} \nabla \tilde{L}^0
\]
\[
+ \eta_2 \nabla \tilde{\lambda} \nabla \tilde{L}^0 K^{-1}Y^TW^T (H - B) \nabla \tilde{z} \nabla \tilde{L}^0
\]
\[
= (\Delta \tilde{z})^T Z B Z \Delta \tilde{z} + \eta_2 (\Delta \tilde{z})^T B Z (Z^T BZ)^{-1}Z^T (H - B) \nabla \tilde{z} \nabla \tilde{L}^0
\]
\[
- \eta_2 (\Delta \tilde{z})^T Y Y^TW^T (H - B) \nabla \tilde{z} \nabla \tilde{L}^0
\]
\[
= (\Delta \tilde{z})^T B Z (Z^T BZ)^{-1}Z^T B \Delta \tilde{z} - \eta_2 (\Delta \tilde{z})^T (I - W^T)(H - B) \nabla \tilde{z} \nabla \tilde{L}^0
\]
\[
- \eta_2 (\Delta \tilde{z})^T Y Y^TW^T (H - B) \nabla \tilde{z} \nabla \nabla \tilde{L}^0
\]
\[
= (\Delta \tilde{z})^T B Z (Z^T BZ)^{-1}Z^T B \Delta \tilde{z} - \eta_2 (\Delta \tilde{z})^T (H - B) \nabla \tilde{z} \nabla \tilde{L}^0.
\]

Here, the second equality is due to the KKT system \( \text{[3]} \) and the fact that \( GZ = 0 \); the third equality is due to the definition of \( W \); and the fourth equality is due to \( Y Y^TW^T = W \). Let us decompose \( \Delta \tilde{z} = \Delta \tilde{v} + \Delta \tilde{u} \), where \( \Delta \tilde{v} = Z \Delta u \) is a vector in \( \text{Im}(Z) \), and \( \Delta \tilde{u} = G^T \Delta u \) is a vector in \( \text{Im}(G^T) \). Since \( G \Delta u = -\nabla \tilde{\lambda} \nabla \tilde{L}^0 \) from \( \text{[3]} \), we know \( \Delta u = -(GG^T)^{-1} \nabla \tilde{\lambda} \nabla \tilde{L}^0 \) and hence \( \Delta \tilde{u} = -G^T (GG^T)^{-1} \nabla \tilde{\lambda} \nabla \tilde{L}^0 = -Y(K^{-1})^T \nabla \tilde{\lambda} \nabla \tilde{L}^0 \). Plugging the decomposition into \( \text{[18]} \), we have
\[
\begin{pmatrix}
\nabla \tilde{z} \nabla \tilde{L}^0 \\
\nabla \tilde{\lambda} \nabla \tilde{L}^0
\end{pmatrix}
W_3 \begin{pmatrix}
\nabla \tilde{z} \nabla \tilde{L}^0 \\
\nabla \tilde{\lambda} \nabla \tilde{L}^0
\end{pmatrix}
\]
\[
= (\Delta \tilde{v})^T Z B Z \Delta \tilde{v} - 2(\Delta \tilde{v})^T Z B Y (K^{-1})^T \nabla \tilde{\lambda} \nabla \tilde{L}^0 - \eta_2 (\Delta \tilde{z})^T (H - B) \nabla \tilde{z} \nabla \tilde{L}^0
\]
\[
+ \nabla \tilde{\lambda} \nabla \tilde{L}^0 K^{-1}Y^TW^T (Z^T BZ)^{-1}Z^T B Y (K^{-1})^T \nabla \tilde{\lambda} \nabla \tilde{L}^0
\]
\[
\geq \gamma_{RH} ||\Delta \tilde{v}||^2 - 4\mu Y ||\Delta \tilde{v}|| ||\nabla \tilde{\lambda} \nabla \tilde{L}^0|| - \eta_2 \delta ||\Delta \tilde{z}|| ||\nabla \tilde{z} \nabla \tilde{L}^0||.
\]
where the second and fifth inequalities are due to Assumption 4.2, which implies $\|K^{-1}\| \leq T, \|B\| \vee \|H\| \leq 2\mu$; the third inequality is due to Young’s inequality; and the fourth equality is due to $\|\Delta T\|^2 = \|\Delta \hat{v}\|^2 + \|\Delta \hat{u}\|^2 = \|\Delta v\|^2 + \|\Delta \hat{u}\|^2$. Using the above display and supposing

$$\frac{\gamma_{RH}}{2} - \eta_2 \delta^2 \geq 0 \iff \eta_2 \leq \frac{\gamma_{RH}}{2\delta^2},$$

we further have

$$\left( \begin{array}{c} \nabla \hat{L}^0 \nabla \hat{L}^0 \end{array} \right) W_3 \left( \begin{array}{c} \nabla \hat{L}^0 \nabla \hat{L}^0 \end{array} \right)^T \geq - \left( \frac{8\mu^2 Y^2 \gamma_{RH} + \eta_2}{\eta_2} \right) \|\nabla \hat{L}^0\|^2 - \eta_2 \delta^2 \|\nabla \hat{L}^0\|^2$$

$$\geq - \frac{4\mu^2 Y^2 \gamma_{RH}}{\eta_2} \|\nabla \hat{L}^0\|^2 - \frac{\eta_2}{4} \|\nabla \hat{L}^0\|^2.$$

Let us now deal with $W_2$. By Schur complement, in order to show $W_2 \succeq 0$, we only have to let

$$\frac{\eta_1}{2} I - K^{-1}Y^T BWYW(K^{-1})^T - \frac{2}{\eta_2} K^{-1}Y^T W^T Y(K^{-1})^T \succeq 0.$$

Note that $-K^{-1}Y^T BWYW(K^{-1})^T \succeq -K^{-1}Y^T BY(K^{-1})^T$ and

$$\|K^{-1}Y^T BWYW(K^{-1})^T + \frac{2}{\eta_2} K^{-1}Y^T W^T Y(K^{-1})^T\| \leq 2\mu Y^2 + \frac{2\mu Y^2}{\eta_2} \|W\|^2$$

$$\leq 2\mu Y^2 + \frac{2\mu Y^2}{\eta_2} \left( 1 + \frac{2\mu}{\gamma_{RH}} \right)^2 = 2\mu Y^2 + \frac{2\mu Y^2}{\eta_2} + \frac{8\mu Y^2}{\eta_2 \gamma_{RH}} + \frac{8\mu^2 Y^2}{\eta_2 \gamma_{RH}^2}$$

$$\leq 12\mu Y^2 + \frac{8\mu^2 Y^2}{\eta_2 \gamma_{RH}} \leq \frac{10\mu Y^2}{\eta_2 \gamma_{RH}}.$$

where the fifth inequality supposes $\gamma_{RH} \leq \sqrt{2}\delta$ (without loss of generality, since $\delta$ is upper bound and $\gamma_{RH}$ is lower bound in Assumption 4.2) so that $\eta_2 \gamma_{RH} \leq 1$; and the last inequality uses $\mu \geq 2\gamma_{RH}$. Thus, we only have to let

$$\frac{\eta_1}{2} \geq \frac{10\mu Y^2}{\eta_2 \gamma_{RH}} \iff \eta_1 \eta_2 \geq \frac{32\mu^2 Y^2}{\gamma_{RH}^2},$$

then (21) is satisfied and $W_2 \succeq 0$. Combining (17), (20), and noting that $W_1$ is a diagonal matrix, we obtain that under (19) and (22),

$$\left( \begin{array}{c} \nabla \hat{L}^0 \nabla \hat{L}^0 \end{array} \right) \eta \left( \begin{array}{c} \Delta \hat{z} \Delta \lambda \end{array} \right) \leq - \left( \frac{\eta_2}{2} \right) \left( \begin{array}{c} \nabla \hat{L}^0 \nabla \hat{L}^0 \end{array} \right)^T \left( \begin{array}{c} \nabla \hat{L}^0 \nabla \hat{L}^0 \end{array} \right)^T \frac{32\mu^2 Y^2}{\gamma_{RH}^2}.$$

Using $\gamma_{RH} \leq 6\mu \delta Y$, we can easily check that, as long as $\eta = (\eta_1, \eta_2)$ satisfies

$$\eta_1 \geq \frac{25\mu^2 Y^2}{\gamma_{RH}} =: \tau_1, \quad \eta_2 \leq \frac{\gamma_{RH}}{2\delta^2} =: \tau_2, \quad \eta_1 \eta_2 \geq \frac{32\mu^2 Y^2}{\gamma_{RH}} =: \tau_3,$$

we have

$$\left( \begin{array}{c} \nabla \hat{L}^0 \nabla \hat{L}^0 \end{array} \right) \eta \left( \begin{array}{c} \Delta \hat{z} \Delta \lambda \end{array} \right) \leq - \frac{\eta_2}{4} \left\| \left( \begin{array}{c} \nabla \hat{L}^0 \nabla \hat{L}^0 \end{array} \right) \right\|^2.$$

This completes the proof of the first part of the statement. For the second part of the statement, we note that $\eta_2^2 = 1$ and each While loop decreases $\eta_2$ by $\rho$. Thus, to satisfy $\eta_2 \leq \tau_2$, the number of the
required While loop iterations $T$ only need satisfy $\rho^T \geq 1/\tau_2$. For the similar reason, we require $\rho^T \geq \tau_3/\mu^2$ and $\rho^T \geq \sqrt{\tau_1/\mu^2}$. Combining them together, we know if $T$ satisfies

$$\rho^T \geq \left( \frac{1}{\tau_2} \vee \frac{\tau_3}{\mu^2} \vee \sqrt{\frac{\tau_1}{\mu^2}} \right) = \frac{32 \Upsilon^2}{\gamma_{RH}^2},$$

then no other iterations will go into the While loop again. Thus, we know $\rho^T \leq \frac{32 \Upsilon^5 \gamma_{RH}}{\mu^2}$. Moreover,

$$\bar{\eta}_2 = 1/\rho^T \geq \frac{\gamma_{RH}}{32 \Upsilon^2 \rho}, \quad \text{and} \quad \bar{\eta}_1 = \mu^2 (\rho^T)^2 \leq \frac{32 \rho^2 \mu^2 \Upsilon^4}{\gamma_{RH}^4}.$$

This completes the second part of the statement.

## C Proof of Lemma 4.1

We note that $YY^T + ZZ^T = I$. Thus, $YY^T (I - BZ(Z^T BZ)^{-1} Z^T) = I - BZ(Z^T BZ)^{-1} Z^T$. Using this observation, the formula of $S$ can be verified directly by checking $SS^{-1} = I$. Moreover, under Assumption 4.2, we know

$$\| (Z^T BZ)^{-1} \| \leq 1/\gamma_{RH}, \quad \| K^{-1} \| \leq \Upsilon, \quad \text{and} \quad \| B \| \leq 2\mu.$$

Therefore,

$$\| S \| \leq \| S_1 \| + 2\| S_2 \| + \| S_3 \| \leq \frac{1}{\gamma_{RH}} + 2\Upsilon (1 + \frac{2\mu}{\gamma_{RH}}) + \Upsilon^2 \left( \frac{4\mu^2}{\gamma_{RH}} + 2\mu \right).$$

Without loss of generality, we suppose $\Upsilon \geq 4$ and $\mu \geq 2(\gamma_{RH} + 1)$. Then

$$\| S \| \leq \frac{6\Upsilon\mu}{\gamma_{RH}} + 2\Upsilon^2 + \frac{4\Upsilon^2\mu^2}{\gamma_{RH}} \leq \frac{\Upsilon^2 \mu^2}{\gamma_{RH}} + \frac{4\Upsilon^2 \mu^2}{\gamma_{RH}} + \frac{5\Upsilon^2 \mu^2}{\gamma_{RH}}.$$

This completes the proof.

## D Proof of Theorem 4.5

We drop off the index $t$ for simplicity. By the definition of $\mathcal{L}_\eta (\cdot)$ in (5), we have

$$\nabla^2 \mathcal{L}_\eta (\tilde{z}, \tilde{\lambda}; \bar{x}) = \begin{pmatrix} H + \eta_2 \nabla z (H \nabla \bar{z} \mathcal{L}) + \eta_1 \nabla \bar{z} (G^T \nabla \bar{z} \mathcal{L}) & \eta_2 \nabla \bar{z} (H \nabla \bar{z} \mathcal{L}) + G \\ \eta_2 \nabla \bar{z} (H \nabla \bar{z} \mathcal{L}) + G & \eta_2 \nabla \bar{z} (H \nabla \bar{z} \mathcal{L}) + G \end{pmatrix}.$$ 

Using Assumption 4.2 [16], and Theorem 4.4, we know

$$\| \nabla^2 \mathcal{L}_\eta (\tilde{z}, \tilde{\lambda}; \bar{x}) \| \leq \frac{32 \rho^2 \mu^2 \Upsilon^5}{\gamma_{RH}^4} =: \mu^2 \Upsilon'.$$

Therefore, by Taylor expansion

$$\mathcal{L}_{\eta}^1 \leq \mathcal{L}_{\eta}^0 + \alpha \left( \frac{\nabla z \mathcal{L}_0^0}{\nabla \bar{z} \mathcal{L}_0^0} \right)^T \left( \Delta \tilde{z} \right) + \frac{\mu^2 \Upsilon'^2}{2} \left( \Delta \tilde{\lambda} \right)^2.$$ 

Moreover, by Lemma B.1 and the condition (7), we further have

$$\left| \left( \frac{\nabla \bar{z} \mathcal{L}_0^0}{\nabla \bar{z} \mathcal{L}_0^0} \right)^T \left( \nabla \bar{z} \Delta \tilde{z} \right) \right| \leq \frac{-100 \rho^4 \mu^4}{\tilde{\eta}_2 \gamma_{RH}} \left( \frac{\nabla \bar{z} \mathcal{L}_0^0}{\nabla \bar{z} \mathcal{L}_0^0} \right)^T \left( \Delta \tilde{\lambda} \right).$$

Plugging the above display into (23),

$$\mathcal{L}_{\eta}^1 \leq \mathcal{L}_{\eta}^0 + \alpha \left( 1 - \frac{50 \rho^6 \Upsilon'' \Upsilon^4}{\tilde{\eta}_2 \gamma_{RH}} \alpha \right) \left( \frac{\nabla \bar{z} \mathcal{L}_0^0}{\nabla \bar{z} \mathcal{L}_0^0} \right)^T \left( \Delta \tilde{z} \right).$$

Thus, as long as

$$1 - \frac{50 \rho^6 \Upsilon'' \Upsilon^4}{\tilde{\eta}_2 \gamma_{RH}} \alpha \geq \beta \iff \alpha \leq \frac{(1 - \beta) \gamma_{RH}}{50 \rho^6 \Upsilon'' \Upsilon^4} \iff \alpha \leq \frac{(1 - \beta) \gamma_{RH}^2}{32 \cdot 50 \rho^6 \Upsilon'' \Upsilon^4} =: \bar{\alpha},$$

then Armijo condition (6) is satisfied. Thus, if we use backtracking line search, the selected stepsize $\alpha \geq \nu \bar{\alpha} =: \bar{\alpha}$ for some $\nu \in (0, 1)$. Moreover, by Armijo condition,

$$\mathcal{L}_{\eta}^1 \leq \mathcal{L}_{\eta}^0 + \alpha \beta \left( \frac{\nabla \bar{z} \mathcal{L}_0^0}{\nabla \bar{z} \mathcal{L}_0^0} \right)^T \left( \Delta \tilde{z} \right) \leq \mathcal{L}_{\eta}^0 - \frac{\tilde{\eta}_2 \beta}{4} \left( \frac{\nabla \bar{z} \mathcal{L}_0^0}{\nabla \bar{z} \mathcal{L}_0^0} \right)^2.$$ 

This completes the proof.
E Proof of Lemma 4.6

By the definition (5), we know
\[
L_t^{t+1} - L_{t+1}^{t+1} = \sum_{k=t+1}^{M_t} \{ g_k(z_{k,t+1}^0) + (\lambda_{k-1,t+1}^0)^T x_{k,t+1}^0 - (\lambda_{k,t+1}^0)^T f_k(z_{k,t+1}^0) \} + g_{M_t}(x_{M,t,t+1}^0, 0) \\
+ \mu \| x_{M_t,t+1,t+1}^0 + (\lambda_{M_t,t+1,t+1}^0)^T x_{M_t,t+1,t+1} - (\lambda_{M_t,t+1,t+1})^T \overline{x}_{t+1} \\
= \sum_{k=t+1}^{M_t-1} \{ g_k(z_{k,t}^1) + (\lambda_{k-1,t}^1)^T x_{k,t}^1 - (\lambda_{k,t}^1)^T f_k(z_{k,t}^1) \} + g_{M_t}(x_{M,t,t}^1, 0) + (\lambda_{M_t-1,t}^1)^T x_{M,t}^1 \\
- (\lambda_{M_t-1,t}^1)^T f_{t}(z_{t}^1). 
\]

Using the above display, we further have
\[
Term_1 = L_t^{t+1} - L_{t+1}^{t+1} \\
= \sum_{k=t+1}^{M_t-1} \{ g_k(z_{k,t}^1) + (\lambda_{k-1,t}^1)^T x_{k,t}^1 - (\lambda_{k,t}^1)^T f_k(z_{k,t}^1) \} + g_{M_t}(x_{M,t,t}^1, 0) + (\lambda_{M_t-1,t}^1)^T x_{M,t}^1 \\
- \mu \| x_{M,t}^1 - \overline{x}_{t} \|^2 + \frac{\mu}{2} \| x_{M,t}^1 \|^2. 
\]

where the last inequality is due to Assumption 4.3(ii). For Term_2, we apply the formula (14) and the transition (8). We have
\[
\| \nabla \lambda_{t+1} L_{t+1}^{t+1} \|^2 = \sum_{k=t+1}^{M_t} \| x_{k+1,t+1}^0 - f_k(z_{k,t+1}^0) \|^2 + \| x_{t+1,t+1}^0 - \overline{x}_{t+1} \|^2 \\
= \sum_{k=t+1}^{M_t-1} \| x_{k+1,t}^1 - f_k(z_{k,t}^1) \|^2 + \| f_{M_t}(x_{M,t,t}^1, 0) \|^2 + \| x_{t+1,t+1}^1 - f_{t}(z_{t}^1) \|^2 \\
= \sum_{k=t}^{M_t-1} \| x_{k+1,t}^1 - f_k(z_{k,t}^1) \|^2 + \| f_{M_t}(x_{M,t,t}^1, 0) \|^2. 
\]

Using the above display, we further have
\[
Term_2 = \frac{\bar{\eta}}{2} \left( \| \nabla \lambda_{t} L_{t}^{t+1} \|^2 - \| \nabla \lambda_{t+1} L_{t+1}^{t+1} \|^2 \right) \\
= \frac{\bar{\eta}}{2} \| x_{t,t}^1 - \overline{x}_{t} \|^2 - \frac{\bar{\eta}}{2} \| f_{M_t}(x_{M,t,t}^1, 0) \|^2 \\
\geq \frac{\bar{\eta}}{2} \| x_{t,t}^1 - \overline{x}_{t} \|^2 - \frac{\bar{\eta}}{2} \| x_{M,t}^1 \|^2, 
\]

(24)

Let us deal with each term separately. For Term_1, we apply the definition of Lagrangian function, the transition (5) and the fact that \( g(0, 0) = 0 \). Then
\[
L_{t+1}^{t+1} = \sum_{k=t}^{M_t-1} \{ g_k(z_{k,t}^1) + (\lambda_{k-1,t}^1)^T x_{k,t}^1 - (\lambda_{k,t}^1)^T f_k(z_{k,t}^1) \} + g_{M_t}(x_{M,t,t}^1, 0) + (\lambda_{M_t-1,t}^1)^T x_{M,t}^1 \\
- \mu \| x_{M,t}^1 - \overline{x}_{t} \|^2 + \frac{\mu}{2} \| x_{M,t}^1 \|^2, 
\]

(25)

(26)
where the last inequality is due to Assumption 4.2. Last, for Term 3, we apply the formula (14) and the transition (8). We have

\begin{align*}
\| \nabla z_{t+1} L^{t+1,0} \|^2 &= \sum_{k=t+1}^{M_t} \left( \left\| \nabla x_k g_k (z_{k,t+1}^0, \lambda_{k-1,t+1}^0) - A_k^T z_{k,t+1}^0 \lambda_{k,t+1}^0 \right\| + \| \nabla x_{M_k} g_{M_k} (z_{M_k,t+1}^0, \lambda_{M_k,t+1}^0) \| \right)^2 \\
&= \sum_{k=t+1}^{M_t-1} \left( \left\| \nabla x_k g_k (z_{k,t}^1) + \lambda_{k-1,t}^1 - A_k^T (z_{k,t}^1) \lambda_{k,t}^1 \right\| + \left\| \nabla x_{M_k} g_{M_k} (x_{M_k,t}^1, 0) + \lambda_{M_k,t-1}^1 \right\| \right)^2 \\
&= \sum_{k=t+1}^{M_t-1} \left( \left\| \nabla x_k g_k (z_{k,t}^1) + \lambda_{k-1,t}^1 - A_k^T (z_{k,t}^1) \lambda_{k,t}^1 \right\| \right)^2 \\
&= \sum_{k=t+1}^{M_t-1} \left( \left\| \nabla x_k g_k (z_{k,t}^1) + \lambda_{k-1,t}^1 - A_k^T (z_{k,t}^1) \lambda_{k,t}^1 \right\| \right)^2.
\end{align*}

Using the above display, we further have

\begin{align*}
\text{Term 3} &= \frac{n_2}{2} \left( \| \nabla z_t L^{t,1} \|^2 - \| \nabla z_{t+1} L^{t+1,0} \|^2 \right) \\
&\geq \frac{n_2}{2} \left( \| \nabla x_{M_t} g_{M_t} (x_{M_t,t}^1, 0) + \lambda_{M_t-1,t}^1 + \mu x_{M_t,t}^1 \|^2 - \left\| \left( \nabla x_{M_t} g_{M_t} (x_{M_t,t}^1, 0) + \lambda_{M_t-1,t}^1 \right) \right\|^2 \right) \\
&\geq \frac{n_2}{2} \left( \| \nabla x_{M_t} g_{M_t} (x_{M_t,t}^1, 0) + \lambda_{M_t-1,t}^1 + \mu x_{M_t,t}^1 \|^2 \right) \\
&\geq \frac{n_2}{2} \left( \frac{\mu^2 - \gamma^2}{2} \| x_{M_t,t}^1 \|^2 + \frac{n_2}{2} \mu \| x_{M_t,t}^1 \|^2 \right) \\
&\geq \frac{n_2}{2} \left( \frac{\mu^2 - \gamma^2}{2} \right) \| x_{M_t,t}^1 \|^2 + \frac{n_2}{2} \mu \| x_{M_t,t}^1 \|^2,
\end{align*}

where the second inequality is due to the definition of $\nabla z_t L^{t,1}$; and the third and the fourth inequalities are due to Assumption 4.2 which implies $\| \nabla x_{M_t} g_{M_t} (x_{M_t,t}^1, 0) \| \leq \gamma \| x_{M_t,t}^1 \|$. Noting that $\lambda_{M_t-1,t}^1 = 0$ and, by (3),

\[
\mu \Delta x_{M_t,t} + \Delta \lambda_{M_t-1,t} = - \left( \nabla x_{M_t} g_{M_t} (x_{M_t,t}^0, 0) + \lambda_{M_t-1,t}^1 + \mu x_{M_t,t}^0 \right) = 0,
\]

we then have

\[
(x_{M_t,t}^1)^T \lambda_{M_t-1,t} = - \alpha t \mu (x_{M_t,t}^1)^T \Delta x_{M_t,t} = - \mu \| x_{M_t,t}^1 \|^2.
\]

Suppose $\mu \geq 4 \gamma$, then $\mu^2 - \gamma^2 - 2 \mu \gamma \| x_{M_t,t}^1 \|^2 \geq \mu^2 / 2$. Together with the above three displays,

\[
\text{Term 3} \geq \frac{n_2}{2} \mu \| x_{M_t,t}^1 \|^2.
\]

Combining (24), (25), (26), and (27), and noting that $\tilde{n}_2 \mu^2 \leq \mu^2 \leq \tilde{n}_1 \gamma^2 / 2$, we have

\begin{align*}
\mathcal{L}_{t}^{t-1} - \mathcal{L}_{t-1}^{t+1,0} &\geq \left( \frac{n_1}{2} - C \right) \| x_{t,t} - \tilde{x}_t \|^2 + \left( \frac{\mu}{2} - \frac{n_1}{2} \gamma^2 - \tilde{n}_2 \mu \right) \| x_{M_t,t}^1 \|^2 \\
&\geq \left( \frac{\mu}{2} - C \right) \| x_{t,t} - \tilde{x}_t \|^2 - \frac{n_1}{2} \gamma^2 \| x_{M_t,t}^1 \|^2 \geq - \frac{n_1}{2} \gamma^2 \| x_{M_t,t}^1 \|^2,
\end{align*}

where the last inequality holds if $C \leq \mu^2 / 2$. By Lemma B.1 [Theorem 4.4 and Assumption 4.3](1),

\[
\tilde{n}_1 \gamma^2 \| x_{M_t,t}^1 \|^2 \leq \frac{32 \rho^2 \gamma^6}{\gamma_R H} \alpha^2 \| \Delta x_{M_t,t} \|^2 = \frac{32 \rho^2 \gamma^6}{\gamma_R H} \alpha^2 \| \Delta \lambda_{M_t-1,t} \|^2 \\
\leq \frac{32 \rho^2 \gamma^6}{\gamma_R H} c^2 \| (\Delta \tilde{x}_t, \Delta \lambda_t) \|^2 \leq \frac{32 \rho^2 \gamma^6 c^2}{\gamma_R H} \| \Delta \tilde{x}_t, \Delta \lambda_t \|^2.
\]

We require

\[
\frac{32 \rho^2 \gamma^6 c^2}{\gamma_R H} \leq \frac{\tilde{n}_2 \alpha \beta}{8} \leq \frac{32 \rho^2 \gamma^6 c^2}{\gamma_R H} \leq \frac{\beta \gamma_R H \tilde{\alpha}}{8 \times 32 \rho^2 \gamma^2} \leq \frac{\beta \gamma_R H \tilde{\alpha}}{240} \leq \frac{\beta (1 - \beta) \gamma_R^3}{20^2 \times 32^2 \rho \mu \gamma \gamma^8} \leq \frac{\gamma_R H}{n_0}.
\]
where the first implication is due to Theorem 4.4 and the second implication is due to Theorem 4.5. Then, we have

\[ L^{t,1}_{\bar{\eta}} - L^{t+1,0}_{\bar{\eta}} \geq -\frac{\bar{\eta}_2 \bar{\alpha} \beta}{8} \| \nabla L^{t,0} \|^2. \]

This completes the proof.

**F Proof of Theorem 4.7**

Summing over \( t \) from \( \tau \) to \( \infty \) on both sides of (11), we have

\[ \frac{\bar{\eta}_2 \bar{\alpha} \beta}{8} \sum_{t=\tau}^{\infty} \| \nabla L^{t,0} \|^2 \leq L^{0,\tau}_{\bar{\eta}} - \min_{z \otimes \Lambda} L_{\bar{\eta}}(\tilde{z}, \tilde{\lambda}; \bar{x}) < \infty. \]

Thus, \( \| \nabla L^{t,0} \|^2 \to 0 \) as \( t \to \infty \). We complete the proof.